The PracTEX Journal, 2012, No. 1
Article revision 2012/2/17

The vocal tract LaTeX package

Dimitrios Ververidis, Daniel Schneider, and Joachim Köhler

Email jimver04@gmail.com; joachim.koehler@iais.fraunhofer.de
Website dimitriosververidis.blogspot.com;
Address Fraunhofer Institute for Intelligent Analysis & Information Systems (IAIS), St. Augustin, Germany

Abstract `VocalTract.sty` is a package to visualize the vocal tract. Vocal tract is manipulated by a vector of articulation parameters according to S. Maeda model. Animation can be achieved by providing a sequence of vectors over time, e.g. from Matlab®. An embedded sequence of vectors in the `VocalTract.sty` for certain German phonemes allows for a sequence of phonemes animation when no vector is available.

1 Introduction

The package provided is a vocal tract visualization tool to be used in speech research. The LaTeX/PSTricks engine used provides non-aliazed vocal tract images in postscript or portable document format (PDF) suitable for manuscripts. The package is also loaded through the VTCalcs software, a software for simulating vocal tract in Matlab® [4].

The vocal tract is manipulated with a vector of variables that vary over time. These are denoted as articulatory model (AM) parameters. The certain 10 AM parameters used here were proposed in [8, 9], where the number of variables related to the tongue was minimized by using factor analysis on vocal tract x-ray animations. The first simulation of the vocal tract was written in Fortran and could solve the differential equations for the voice to vocal tract and the inverse problem [8]. Faster implementations in C++ and Matlab® appeared [4]. Later, the model was augmented to provide control for the vocal chords tension [5]. In our work, the breathing is also simulated by an additional parameter. We consider that the provided package will enhance the vocal tract drawings to appear in manuscripts and help to understand the origins of speech production.

D. Ververidis work was carried out during the tenure of an ERCIM fellowship;

The outline of this manuscript is as follows. A study about the phonemes of the German language and their respective vocal tract set up using our package is shown in Section 2. Details about how to use the the vocal tract package are described in Section 3. Future work is discussed in Section 4.

2 German phonemes categorization

German language phonemes are categorized into vowels, unvoiced consonants, and voiced consonants according to the oscillation source, that is glottal, contrived or both types, respectively. Vowels are presented in Figure 1(a) and are produced with glottis oscillation. The alphabet used is the speech assessment methods phonetic alphabet (SAMPA) that is widely used in computer science [15]. Unvoiced consonants are shown in Figure 1(b) and are generated from the oscillation (due to turbulence) of a constriction formed by a part of the oral cavity with the tongue. Voiced consonants depicted in Figure 1(c) are generated by both sources of oscillation. Apart from the 26 German phonemes shown here, the phoneticians recognize 18 to 20 other phonemes [3]. In our work we assumed that the additional phonemes combine the 26 basic phonemes, i.e. the affricative 'ts' for the word 'Zahl' (tsa:l) can be considered as a sequence of the plosive 't' and the fricative 's'.

Vowels are separated into front, central and back depending on the position of the tongue; Rounded and unrounded depending on the shape of the lips; Close to open depending on the distance between the tongue and the ceiling. Unvoiced consonants are classified depending on which part of the oral cavity is coupled with the tongue to provide an oscillation source into 4 categories, namely: bilabial, labio-dental, alveolar, and velar. They can be also categorized into plosives and frivatives depending on the amount of pressure accumulated before pronunciation, i.e. high pressure for plosives and low for fricatives. The voiced consonants are classified according to the tongue position and the lips aperture into bilabial, labiodental, alveolar, palatal, velar and uvular. Special cases are the 'r' that is pronounced with a movement of the tongue from palatal to alveolar position and 'l' that is generated with air escape from the sides of the tongue when the latter is raised [3, 9].

The sequence of the certain 26 vectors of 10 parameters each is embedded in the style file, so that visualization is possible when no vectors are available. The

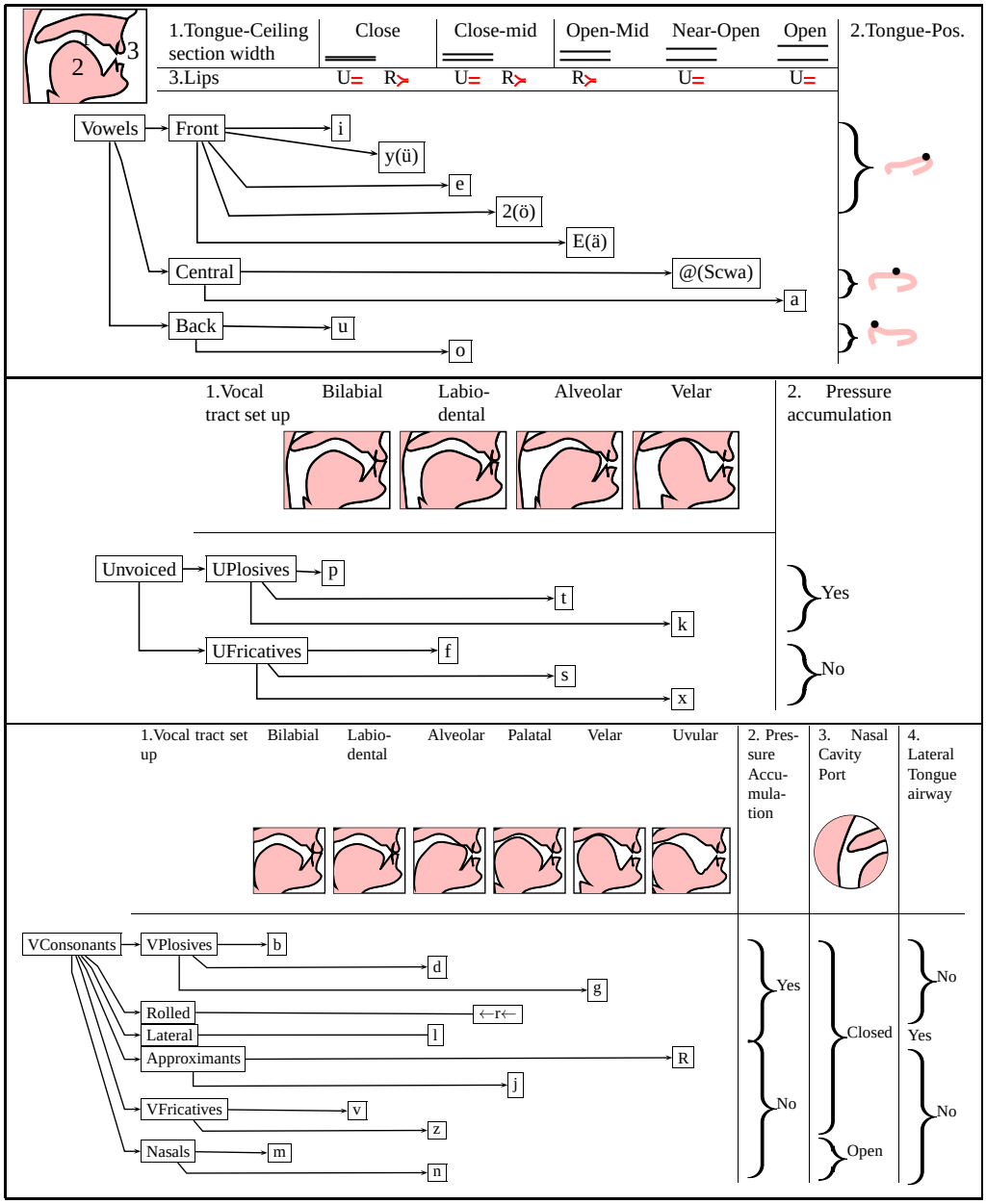

Figure 1: German phonemes categorization into (a) Vowels (b) Unvoiced consonants. (c) Voiced consonants. Legend: U=Unrounded; R=Rounded; Pos.=Position.

breathing parameter is set as 0.5 for normal breathing (0.1 for short and slow breath to 0.9 for deep and fast breath). The output is the animation of Figure 2. The red space corresponds to the pressure accumulated for generating plosive phonemes.

3 Installation and examples

`VocalTract.sty` can be downloaded from [13]. In order to be compiled, a LaTeX2e package [7], such as Miktex for Windows [10] and TexLive for Linux [12] must be already installed. The following PSTricks packages have been included in the preample of the style file: `pstricks, pst-coil` for drawing lines and patches; `arrayjob, ifthen, fp, fltpoint` for algebraic operations among variables; and `multimedia, multido, animate` for rendering animations [14].

In Figure 3, a logic diagram of scripts to call the package either from LaTeX or Matlab® is depicted. 'vtLatex_FigureDemo.tex' produces a visualization of the vocal tract in DVI, PS, or PDF format. 'vtLatex_AnimationDemo.tex' produces a

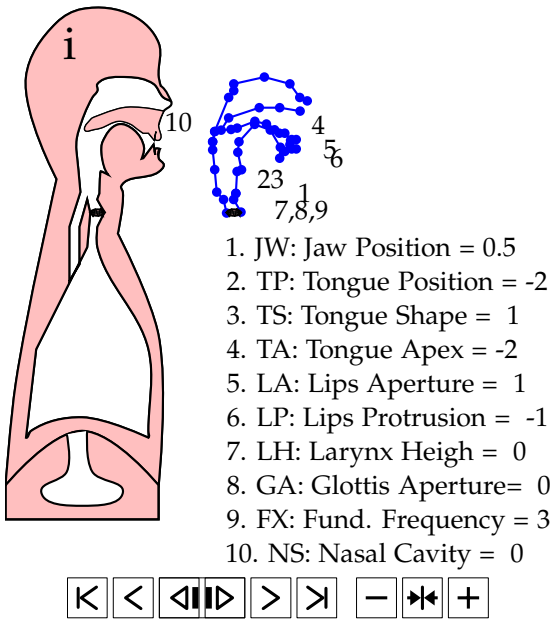

Figure 2: Vocal tract system for the vowel 'i'. Left: using sp-lines; Right: using lines only.

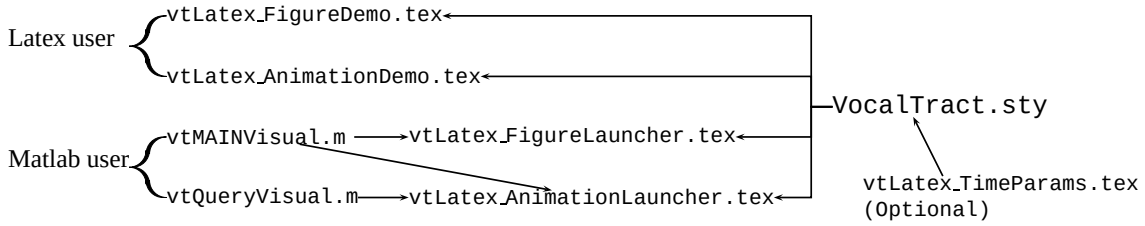

Figure 3: Functions that generate figures or animation. Two functions named as Launchers are responsible for connecting Matlab® functions with LaTeX engine.

sequence of ordered visualizations of the vocal tract in a PDF animation object. Figure 2 presents the compilation result of 'vtLatex_AnimationDemo.tex' as it follows:

```
%======= Vocal Tract Animation Demo =======
\documentclass{article}\pagestyle{empty}
\usepackage{VocalTract}
%\input{vtLatex_TimeParams} % AM Parameters (if commented, use the embedded
%                                        vectors in the Style file)
\begin{document}
\def\BreathType{0.5}  % Normal Breathing [slow:0.1 to fast:0.9]
\begin{animateinline}[poster=first,loop,controls]{1}%
\multiframe{\Nframes}{IndexStepVocalTract=0+1}{% Nframes defined in
%                                                 vtLatexTimeParams.tex
\begin{pspicture}(0,-0.5)(6,8)
\SetOscAmp{\IndexStepVocalTract}{\BreathType} % Sinusoid generator for breathing
\UpdateVocalTract{\IndexStepVocalTract}  % in the respective time
\VocalChords              % Show VCords
\StomachCompartment       % Show Stomach Compartment
\rput(1.75,7){\ShowPhonemesGerman{\IndexStepVocalTract}} % Transcriptions
\ShowPressure{\IndexStepVocalTract}  % Show Pressure as Red Space
\rput(2,0){\ShowLinearTubes{\IndexStepVocalTract}\VocalChords} % The linear
%                                                       model next to the human
\rput(2.5,0){\ShowParameters}  % Explain the parameters
\end{pspicture}}% Multiframe
\end{animateinline}
\end{document}
```

In order to produce an animation for a certain word, e.g. 'satz', the 'vtLatex_AnimationDemo.tex' should be compiled and the third line should be uncommented. In the end, the 'VocalTract_TimeParams.tex' should be:

```
%============= VocalTract_TimeParams.tex ==================
\newarray\SpeakVec
\readarray{SpeakVec}{%
%%=============================================================================================
%% 1. Jaw Position     4. Tongue Apex     7. Larynx Height       10. Nasal Cavity
%%  2. Tongue Position  5. Lip Aperture    8. Glottal Aperture    11. Phoneme Label
%%   3. Tongue Shape     6. Lip Protrusion  9. Glottal Frequency
%%=============================================================================================
%%  1       2     3       4      5      6     7    8     9    10      11
%% JW      TP    TS      TA     LA     LP    LH   GA    FX    NS
%%  3       3     3       3      3      3     3    3     3     3               Max
%%  0       0     0       0      0      0     0    0     0     0               Relax
%% -3       3    -3      -3     -3     -3    -3   -3    -3    -3              Min
%%=============================================================================================
    2.5  &  0   & 0    &  0.4  & 0  &  0    & 0  & -1 &  0  &  0 &z& % 24
 -1.5    & 2.5  & 0    &-0.5   & 0.5& -0.5  & 0  & 0  &  0  & 0  &a& % 7
    0    & -0.5 &-1    &  2.3  & 0  &  0    & 0  &-3  & -3  & -3 &t& % 11
    2.5  &  0   & 0    &  0.4  & 0  &  0    & 0  &-3  & -3  & -3 &s& % 14
\dataheight=11
}
\dataheight=11
\def\Nframes{5}
```

'vtMAINVisual.m' is a Matlab® function designed as a plug-in function for the VTCalcs package [4]. It can be fed with Maeda's articulatory parameters to visualize the vocal tract in DVI, PS or PDF. The 'vtQueryVisual.m' provides a PDF animation from Matlab® command line when a sequence of phonemes from the set of the phonemes described in Section 2 is given, e.g. » `vtQueryVisual("zats")` stands for the visualzation of word 'satz'.

4 Future work

A method to depict the vocal tract using the LaTeX and the PSTricks engine was presented. Although the functionality of the method is high, several issues should be addressed in the future, namely:

- The shape of the curves generated by sp-lines in PSTricks to connect nodes are erroneously estimated among several postscript zoom levels. That is a curve is not the same at 100% zoom level compared with 200% zoom level.

It might happen due to the different step size in sp-line estimation through zoom levels.

- Audio should be included in PDF animations. Phonemes can be synthesized with VTCalcs [4];
- The floating point operations should be faster in order to improve compilation speed which is now 1 second per frame;
- The 2D model should become 3D model as in [1, 2] and [6];
- The set of 26 German phonemes should be augmented to include phonemes from other languages;
- Functions to call the package from Python language [11] or C should be written.

Acknowledgment

We would like to thank Prof. Mark Huckvale for his offer to provide code and to explain the S. Maeda model. We would like to thank also Prof. Francisco Reinaldo for the useful comments on the paper.

References

[1] P. Birkholz, D. Jackél, and K.J. Kroger. Construction and control of a three-dimensional vocal tract model. In *Proc. IEEE Intern. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2006.

[2] J. Dang and K. Honda. Estimation of vocal tract shapes from speech sounds with a physiological articulatory model. *J. Phonetics*, 30(3):511–532, 2002.

[3] K. Duden. *The Phonetic Dictionary*. Dudenverlag: Mannheim/Leipzig/Wien/Zürich, 2005.

[4] S. Ghosh. VTCalcs for Matlab: `http://www.cns.bu.edu/~speech/VTCalcs.php`, 2000.

[5] M. Huckvale. VTDemo: Vocal Tract Acoustics Demonstrator. Web site: http://www.phon.ucl.ac.uk/resource/vtdemo/, 2005.

[6] H. Kjellström and O. Engwall. Audiovisual-to-articulatory inversion. *Speech Communication*, 51(3):195 – 209, 2009.

[7] D. E. Knuth. *Computers & Typesetting*. Addison-Wesley Reading, MA, 1986.

[8] S. Maeda. A digital simulation method of the vocal-tract system. *Speech Communication*, 1(3-4):199 – 229, 1982.

[9] S. Maeda. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W. J. Hardcastle and A. Marchal, editors, *Speech Production and Speech Modelling*, pages 131–149. Kluwer Academic Publisher, Boston, 1990.

[10] Miktex. Project page: http://miktex.org, 2010.

[11] Python Software Foundation. Python Programming Language, Official Website: http://www.python.org, 1990.

[12] TexLive. Tex users group web site: http://www.tug.org, 2010.

[13] D. Ververidis. Blog page. dimitriosververidis.blogspot.com, 2011.

[14] H. Voß. *PSTricks: Grafik für TEX und LATEX*. Lehmanns Fachbuchh., 2007.

[15] J. C. Wells. SAMPA computer readable phonetic alphabet. In D. Gibbon, R. Moore, and R. Winski, editors, *Handbook of Standards and Resources for Spoken Language Systems*, volume 4B. Berlin and New York: Mouton de Gruyter, 1989.