Visible Vowels: a Tool for the Visualization of Vowel Variation

Wilbert Heeringa
Fryske Akademy
Ljouwert, The Netherlands
wheeringa@fryske-akademy.nl

Hans Van de Velde
Fryske Akademy
Ljouwert, The Netherlands
HVandeVelde@fryske-akademy.nl

Abstract

Visible Vowels is a web app for the analysis and visualization of acoustic vowel measurements: f0, formants and duration. The app is a useful instrument for research in linguistics. The app combines user friendliness with maximum functionality and flexibility, using a live plot view.

1 Introduction

Researchers in phonetics, sociolinguistics, dialectology, forensic linguistics, speech pathology, language acquisition, pyscholinguistics and neurolinguistics study the acoustic characteristics of vowels, measuring vowel formants, fundamental frequency, duration and other acoustic variables. Next, the measurements are visualized by graphs in order to find patterns that reflect particular external (e.g., region, age, gender, language background, pathological vs. non-pathological) or internal (e.g., word stress, following segment) factors. While programs like Praat, Speech Filing System (SFS) and WaveSurfer are frequently used for making acoustic measurements, most researchers use programs like PLOTNIK, NORM and VOIS3D or the R packages `vowels` and `phonR` for the visualization of these measurements. However, these packages do not meet the demands of most researchers in the field, due to a lack of features, flexibility and user-friendliness.

PLOTNIK was developed by Labov (2011) for the study of vowel variation and change in North American English and plots the vowels in F1/F2 space. Any selection of vowels can be visualized. Subsets of vowels can be highlighted within the larger set. Procedures for speaker normalization of formant frequencies are available. Means, standard deviations and the significance of difference between means can be calculated. Vowel systems of different speakers can be compared, by overlaying their vowel plots. PLOTNIK runs only on Macintosh operating systems.

The R package `vowels` includes “procedures for the manipulation, normalization, and plotting of phonetic and sociophonetic vowel formant data.” (Kendall and Thomas, 2015). This package is the backend for the vowel normalization and plotting suite NORM. The NORM web application offers the possibility to run a number of normalization techniques on formant data and to quickly compare the results. The web application NORM runs on any platform and does not require any knowledge of the programming language R. However, NORM is less flexible than the `vowels` package, therefore, the authors of NORM encourage users to use their R package `vowels` rather than using NORM.

VOIS3D offers the possibility to normalize formant frequencies and duration. Additionally, it provides an analytic geometric solution for assessment of spectral overlap. Normalized scatter for two vowel distributions is modeled as two best-fit ellipses oriented at angles with respect to the F1 and F2 axes. The output of the metric is an overlap fraction, which represents the region shared by both best-fit ellipses. This tool runs only on Windows operating systems (Wassink, 2006). An important limitation of both PLOTNIK and NORM is that they only offer visualization of formants, not of duration and f0 of vowels. VOIS3D, however, can visualize vowel duration variation.

A package related to the `vowels` package is the `phonR` package (McCloy, 2016) that can be used to visualize trajectories with an unlimited number of measure points. Additionally, IPA glyphs, ellipses showing degree of confidence in the location of the mean of each vowel/group and convex hulls representing the outline of the vowel space can be added to the vowel plot. The degree of encroachment or overlap between vowel categories can be calculated and plotted by means of a heat map.

In Section 2 we present Visible Vowels, a web application that aims to combine user friendliness with maximum flexibility and functionality. The app is a useful instrument for research in phonetics, sociolinguistics, dialectology, forensic linguistics, and speech-language pathology. Visible Vowels includes a large deal of the functionality of the aforementioned packages, but offers also new functionalities such as the measurement of long-term formants and comparison of speakers with Huckvale's ACCDIST metric (Huckvale, 2004). Different from earlier vowel plot programs, Visible Vowels uses a live view, i.e. each time the user changes something in the settings, the plot shown in the viewer is immediately adjusted accordingly. This makes the comparison of, for example, different normalization techniques extremely easy. Being web-based, Visible Vowel runs on any platform. Future lines of development are presented in Section 3.

2 Program

The web app Visible Vowels is implemented in the programming language R (R Core Team, 2017). For drawing graphs functions from the packages `ggplot2` (Wickham, 2009), `plot3D` (Soetaert, 2017) and `ggdendro` (De Vries and Ripley, 2016) are used. The user interface has been built using the packages `shiny` (Chang et al., 2017) and `shinyBS` (Bailey, 2015). Visible Vowels is freely available at `https://visiblevowels.org`. The app is tested in Firefox, Google Chrome, Opera, Edge and Safari. A standalone version can be used by installing the package `visvow` in R. The app consists of seven tab panels: 'Load file', 'Contours', 'Formants', 'Dynamics', 'Duration', 'Explore' and 'Help'. In the panels 'f0', 'Formants', 'Duration' and 'Explore' the user can set the size and font of the axis ticks and labels and the size of the graph as a whole. Data files can be downloaded either as tab-delimited text file or as Office Open XML file. Graphs can be saved in five different formats (SVG, PDF, EPS, JPG, PNG).

2.1 Load file

In this panel, a spreadsheet file that contains the measurements is loaded. Spreadsheets are usally made in Microsoft Excel, therefore a file with Excel's default file extension ('.xlsx') can be used without the need to export this file to a text file. The spreadsheet should contain the speaker labels, vowel labels, categorical variables (for example, region, gender, etc.) and duration. Next a set of five variables –'time', 'f0', 'F1', 'F2' and 'F3'– may be repeated as *many times* as the user wishes. More explanation about the table format is provided in the 'Help' panel. In this panel an example spreadsheet is provided as well.

2.2 Contours

In this panel the contours of f0, F1, F2 or F3 can be visualized using the time points selected by the user.

Multiple line graphs can be generated, where each contour represents a category of a categorical variable. A plot can also be divided in panels, where each panel shows a graph with a contour for a value of a categorical variable. When multiple line graphs are combined with multiple panels, it is, for example, possible to compare contours of vowels [i] and [a] for older and younger speakers, having a panel for vowel [i] and a panel for vowel [a], and with two contours in each panel, for each age group one contour. Hertz (Hz) values can be converted to bark, ERB, ln, mel and ST values and exported to a data file.

2.3 Formants

In this panel vowels are plotted either in two- or three-dimensional space, using combinations of F1, F2 and F3. It is possible to average categories of the categorical variables. For example, for 10 vowels, three regions and two genders the plot will show $10 \times 3 \times 2 = 60$ plot symbols in three different colors (region in the example) and two different shapes (gender). Additionally, long-term formants can be shown: per speaker formant values are averaged over the vowels, and the points in the graph represent speakers instead of vowels. When multiple five-column sets are included in the input table, vowel trajectories can be drawn. Given *n* sets, the user can include all time points or any subset of points, with a minimum of two subsets.

Hz values can be converted to bark, ERB, log and mel. There are 14 vowel normalization methods available that are divided in four groups: formant-ratio normalization, range normalization, centroid normalization and log-mean normalization. We added a new normalization method 'Heeringa & Van de Velde 2018' that is similar to the normalization method of Watt and Fabricius (2002), but it uses all the points the convex hull is made up of. Some methods require f0 and/or F3 measurements. The normalization methods can be used on the basis of any scale (Hz, bark, etc.) except for the methods that use a log transformation, which can only be applied to Hz values.

In 2D graphs the different colors of the plot symbols represent the categories of one categorical variable, and the shapes represent the categories of another categorical variable. Shapes can be circles, triangles, squares etc, but when 'vowel' is chosen, the vowel labels are shown instead. It is also possible to show multiple panels, each graph corresponding with a value of a categorical variable. Thus vowel systems of, for example, different regions can easily be compared.

Groups as defined by categorical variables can be made more visible by drawing spokes between each vowel and the gravity center of the group to which it belongs. Ellipses showing the degree of confidence in the location of the mean of each group can be shown for any confidence level. Convex hulls can be drawn in order to delineate the outline of the vowel space per category of a categorical variable. By drawing a convex hull around the vowel space, the effects of the different normalization methods can be compared due to the live plot view.

In 3D graphs vertical spikes can be added in order to see more clearly the *x*,*y* location of the points and to enhance the three-dimensional effect. The user can change the angles with respect to respectively the *x*-axis and the *z*-axis in order to view the graph from different perspectives.

2.4 Dynamics

Vowels measurements change over time during the vowel segment. In order to measure vowel dynamics, we implemented two methods: trajectory length and spectral rate of change. Fox and Jacewicz (2009) used these methods on the basis of F1 and F2 measurements. In Visible Vowels the user can choose any subset of f0, F1, F2 and F3, including the individual variables or all of the variables.

Using either multiple line graphs or grouped bar graphs it is possible to visualize vowel dynamic measurements for different values of a categorical variable, where each line in a line graph or 'sub bar' in a grouped bar graph represents a category. In a grouped bar graph, per category of the first categorical variable, bars are shown for all categories of the second categorical variable. Additionally, multiple panels can be shown, where each panel corresponds with a value of a categorical variable.

2.5 Duration

Duration can be visualized in the same way as vowel dynamics, using multiple line graphs or grouped bar graphs and panels as well. Duration measurements can be normalized by means of Lobanov's *z*-score transformation (Wang and Van Heuven, 2006).

2.6 Explore

In this panel distances between speakers on the basis of their vowel systems can be calculated using the ACCDIST metric of Huckvale (2004) which we extended by offering the possibility to include F3 (in addition to F1 and F2), and to compare groups of speakers that are defined according to one or more categorical variables. The results are visualized by either dendrograms or multidimensional scaling plots. The user can choose from five cluster algorithms and four multidimensional scaling routines.

2.7 Help

The help panel provides an 'about' section, describes the format of the input file and provides an example spreadsheet. A document is provided that explains how methods for averaging and measuring long-term formants, scale conversion, speaker normalization of formants, speaker normalization of duration, measuring vowel dynamics and the ACCDIST metric are implemented in Visible Vowels. Additionally, in the help panel references for the R packages that are used in the program are provided.

3 Future work

Currently, Visible Vowels is available as a web app and a standalone version. In the latter version we will implement clickable maps, in which a user can click on a vowel in the F1/F2 space and hear the recording. We welcome other suggestions for extensions and further improvements. Visible Vowels will be included in the CLARIAH Virtual Research Environment (VRE). The CLARIAH VRE for linguistics will be hosted at the Meertens Institute/Humanities Cluster in Amsterdam and will integrate selected tools and services for linguistic research developed and supported in the Dutch CLARIN NL and CLARIAH projects. Finally, we plan to develop 'Visible Consonants, a similar tool for consonants.

Acknowledgements

We thank Vincent van Heuven and several other users of Visible Vowels for their valuable suggestions which we were happy to implement.

References

Eric Bailey. 2015. Package 'shinyBS': Twitter Bootstrap Components for Shiny. version 0.61.

Winston Chang, Joe Cheng, J.J. Allaire, Yihui Xie, and Jonathan McPherson. 2017. Package 'shiny': Web Application Framework for R. version 1.0.1.

A. De Vries and B. D. Ripley. 2016. ggdendro: Create Dendrograms and Tree Diagrams Using 'ggplot2'. https://CRAN.R-project.org/package=ggdendro. R package version 0.1-20.

R. A. Fox and E. Jacewicz. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, 126(5):2603–2618. https://doi.org/10.1121/1.3212921.

M. Huckvale. 2004. ACCDIST: a Metric for Comparing Speakers' Accents. In *Proceedings of the International Conference on Spoken Language Processing, Jeju, Korea, Oct 2004*, pages 29–32.

Tyler Kendall and Erik R. Thomas. 2015. Package 'vowels': Vowel Manipulation, Normalization, and Plotting. version 1.2-1.

W. Labov, 2002. *A PLOTNIK tutorial*.

W. Labov. 2011. Plotnik. https://www.ling.upenn.edu/ wlabov/Plotnik.html. version 10.

Daniel R. McCloy. 2016. phonr: Tools for Phoneticians and Phonologists. https://cran.r-project.org/web/packages/phonR/index.html. R package version 1.0-7.

R Core Team. 2017. R: A language and environment for statistical computing. URL: `http://www.R-project.org/`.

K. Soetaert. 2017. plot3d: Plotting Multi-Dimensional Data. https://CRAN.R-project.org/package=plot3D. R package version 1.1.1.

H. Wang and V. J. Van Heuven. 2006. Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In Jeroen van de Weijer and Bettelou Los, editors, *Linguistics in the Netherlands 2006*, volume 23, pages 237–248. https://doi.org/10.1075/avt.23.23wan. John Benjamins, Amsterdam.

A.B. Wassink. 2006. A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three varieties. *Journal of the Acoustical Society of America*, 119(4):2334–2350. https://doi.org/10.1121/1.2168414.

D. J. L. Watt and A. H. Fabricius. 2002. Evaluation of a technique for improving the mapping of multiple speakers. *Leeds Working Papers in Linguistics and Phonetics*, 9:159–173.

H. Wickham. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.