Sound dimensions and formants

Grandon Goertz[1], Terese Anderson[2]

[1]Department of Language, Literacy and Sociocultural Studies, University of New Mexico, US

[2]Department of Chemistry and Chemical Biology University of New Mexico, US

https://doi.org/10.36505/ExLing-2020/11/0023/000438

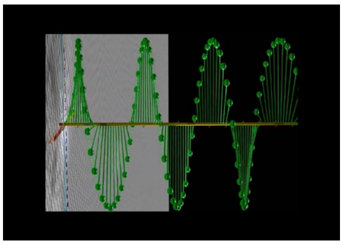
Abstract

Every physical event that can be observed can be measured and described, including sounds. This paper discusses computer algorithms that were developed to depict vowels and speech sounds in their three dimensions: frequency, energy, and time. Each vowel has a separate distinguishable shape based on its dimensions. Two-dimensional vowel plots can be more accurately represented in three-dimensional plots. Algorithms using the Chebyshev Transform were written and vowel speech signals were converted to accurate numerical data sets that were examined and then plotted. Comparisons of vowels can be made, based on their sonic shape. This algorithm also used the Singular Value Decomposition (SVD) to measure, vowel formants giving clear formant regions with the frequency regions identified on the y-axis plots.

Keywords: Three-dimensional depiction, Chebyshev Transform, formant

Introduction

Sound is a vector-defined, moving, measurable physical quality that can be depicted in three dimensions, as illustrated by Khutoryansky (2019).



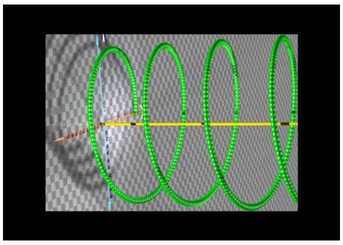


Figure 1. Side view of a pure tone, showing vectors. The right picture shows the same sound at an angle, illustrating the three dimensions.

Algorithms were written to measure voice data in order to produce accurate depictions of sounds by first converting the data to an accurate numerical base, then analyzing the data with a Chebyshev Transform (Boyd 2001), and finally using the Singular Value Decomposition (Gold, Morgan 2001). The next step was to produce three-dimensional plots of the vowels and measured plots of the formants (Johnson 2012). Empirical data can be evaluated, and the data used to create a picture for each vowel that depicts the waveform envelope and shows the sound shapes with its properties of frequency, time, and energy.

Method

Formant values produced with a 2-meter refractometer, a mechanical device which separates sound into its components, did not corelate to the calculated formants from a popular computer program used by linguists. It was also observed that formants produced by a traditional FFT-based program would not plot in a way that is consistent with the understanding of formants as being harmonic-frequency intense locations. We experimented with several sound files of various languages from the University of California phonetics website (UCLA: phonetics.ucla.edu) and vowel sounds from the first author. It was discovered that sound data was not correlating accurately when it was plotted using base 10. Base 10 is also known as the 0-9 decimal scale.

We assessed if the data was better represented in a different numeric base. Evaluations were done with a mathematical procedure, Principal Component Analysis (PCA), to determine which numeric base described the data best. The process of determining the base began with evaluating each vowel using its T-square values and histograms. The data represented the correlation of the frequency values of the formants of German, English, Swedish, French, and Japanese vowels.

Base e, which occurs very often in biological phenomenon such as growth, was examined and found to be the most accurate base to represent sound frequencies and propagation. Base e, also called the natural logarithmic space, uses a numerical scale in which the numbers ascend in a logarithmical sequence. When base 10 is used the numbers are misrepresented as being too small.

In figure 2 below, the top grey line represents the numerical values of base 2, and the red line just under it represents base e (the value of base e is 2.71828). Moving toward the x-axis the grey lines represent bases 3, 4, 5, 6,7, 8, and 9. The bottom green line represents base 10, a log scale, the scale that is most familiar.

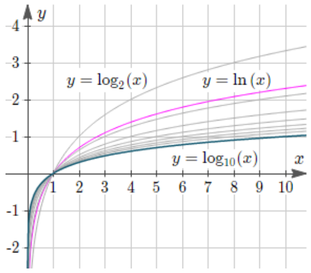


Figure 2. Bases. The base used for sound is represented by the red line.

An algorithm written in Matlab, processes the speech data by first converting the .wav or .AIFF file into frequencies. The frequency data was transformed using Chebfun (Trefethen 2000). A time vector was computed, and the frequency axis was scaled according to the data size. The transformed frequency data was then converted to base e, the natural log values. Energy was also computed for each data point.

The SVD was next computed, the values were checked for accuracy, and the result is a matrix of formants. SVD is an empirical based matrix computation that effectively reduces empirical data to the most significant data values that bests represents the original data. This allows for detection of the true structure of data for modeling.

Results and discussion

The first discovery is the ability to visualize the distinct vowel shapes as they propagate through space (without dispersion) in three dimensions. After converting the sound data to base e, it was possible to plot the data as a picture of a three-dimensional event.

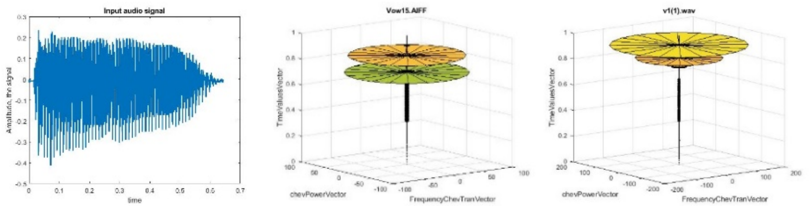


Figure 3. A two-dimension waveform of /o/ produced by a male speaker. Center is the three-dimensional depiction of the same /o/. On the right is the vowel /i/, also by a male speaker. (Black center lines in the center and right pictures are computer-generated zero values.)

Using the data in the correct natural log e space allowed us to also plot formants that were un-ambiguous. The SVD calculation method produced verifiable frequency values that, when plotted, show clear formant banding.

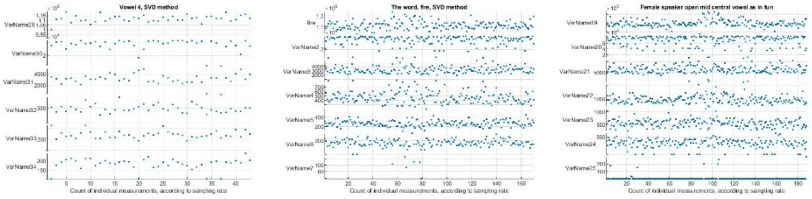


Figure 4. The vowel /ʉ/ on left, and center, the word English word 'fire' spoken by a male. Both plots show clear and calculated formant bands. On the right, /Λ/ spoken by a female, showing formants.

Formants exist in clear bands, typically not more than seven. The frequency zones are shown on the y-axis and time is indicated by sample number on the x-axis. Both formants and three-dimension depictions define the vowel, in natural log space.

References

Johnson, K. 2012. Acoustics and Auditory Phonetics. MA, Wiley-Blackwell Malden.

Boyd, J. 2001. Chebyshev and Fourier Spectral Methods. Mineola, NY, Dover Publications, Inc.

Gold, B., Morgan, N. 2000. Signal Processing and Perception of Speech and Music. NY, John *Wiley & Sons.*

https:// www.youtube.com/user/EugeneKhutoryansky. January 25, 2019.

Trefethen, L. 2019., *Spectral Methods in MATLAB*, SIAM, 2000.