An improved method for vowel space depiction

Terese Anderson[1], Grandon Goertz[2]
[1]Chemistry and Chemical Biology, University of New Mexico
[2]Educational Linguistics, College of Education & Human Sciences, University of New Mexico
https://doi.org/10.36505/ExLing-2021/12/0006/000479

Abstract

All speech is composed of combined individual sound waves that gives speech its particular qualities. Vowel space measurements are usually plotted in charts in which the second formant is plotted against the first formant. For this investigation, the formant values are taken from a computer program that employs a Fourier methodology to analyse the sound signal. This paper presents information about an improved method for vowel space depiction that was created by employing the mathematical integration of formant data. This model produces a more reliable mathematical model the vowel space. Each vowel formant can be represented by an algebraic equation.

Keywords: formant measurement, formant equation, vowel space, speech signal

Introduction

Speech is a product of a vocal source which is the excitation signal, and a filter, or time-varying vocal tract. "The spectral envelope determines the relative magnitudes of the different harmonics, and it, in turn, is determined from the specific shape of the vocal tract during the phonation of that vowel" (Gold and Morgan 2000, p 25; Parker 1988).

A speech wave is a time evolution phenomenon that is modelled using partial differential equations that have a dependent variable that represents the wave value, an independent variable time, and one or more independent spatial variables (Elmore and Heald 1969; Pain 1993). Sound is the compilation of several simultaneous frequency waves that move through air. Sound waves interact as additive and subtractive sine wave components and form a unique sound, which is represented visually as a waveform in frequency space. *The essential idea is that speech is comprised of numerous sine waves, and each wave component has a distinct continuous mathematical function.*

The illustration in Figure 1, taken from page 765 of *The Handbook of Phonetic* Sciences (Ellis 2010), illustrates that a periodic function can be divided into its mathematical harmonic components. All sound waveforms can be mathematically represented by combinations of the mathematical equations that make up the waveform. Because speech is composed of multiple waves it is also possible to separate the sounds into formants.

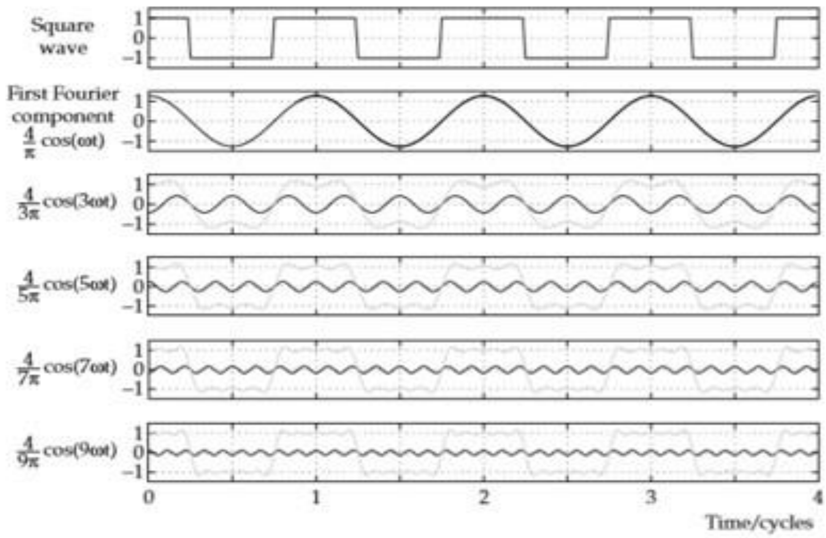

Figure 1. Each harmonic is an integer multiple of the fundamental frequency. Each of the five waves shown below the top square wave are components of the square wave.

Speech as a mathematical phenomenon

A speech wave is modelled using partial differential equations, and these equations have an independent variable, time, and one or more independent spatial variables. The actual form of the wave is strongly dependent on the system's initial conditions, the boundary conditions, and system disturbances. Waves are described by their solutions to either linear or nonlinear partial differential equations (Elmore and Heald 1985).

There are numerous convolution methods that will separate speech into its component parts, the spectral envelope and excitation values. Once the speech is separated, both the spectral envelope and spectral fine structure can be efficiently parameterized. Methods of source-filter separation include linear prediction, cepstral analysis which produces a spectrum, and formant vocoding.

The vowel space values are traditionally produced by taking the linear average of the frequency values over time to produce a single value. These values are then plotted on a F1 vs F2 plot.

Finding formant equations

Praat produces formant bands or regions, typically: F1, F2, F3, and F4. The Praat formant listing includes points of the wave equation, and not the equation, but the formant listing information is sufficient to determine the equation of the wave.

Separating the formant components of the speech wave allows the determination of a mathematical model equation which can be integrated and differentiated to produce the actual centroid of each function (see Kreyszig 1983, Stein 1973, and Thomas Calculus 2005). Laws of calculus state that the "average" or centroid for a continuous function is determined by the integral of that function. Equations were computed using MATLAB's Curve Fitting Toolbox using the F1-F4 values from Praat. It was found that each sound formant can be effectively represented by an 8-term Gaussian equation. A total of 254 individual sounds were evaluated for comparison and to check the validity of the procedure. The general equation for a formant is the Gaussian Equation, below.

$$f(x) = A1 * e^{\left(-\left(\frac{x-B1}{C1}\right)^2\right)} + A2 * e^{\left(-\left(\frac{x-B2}{C2}\right)^2\right)} + A3 * e^{\left(-\left(\frac{x-B3}{C3}\right)^2\right)} + A4 * e^{\left(-\left(\frac{x-B4}{C4}\right)^2\right)} \ldots . + A8 * e^{\left(-\left(\frac{x-B8}{C8}\right)^2\right)}$$

This equation remains the same for all input formant data. These coefficient values (A1-A8, B1-B8, and C1-C8) are unique to each sound and formant and vary from small negative values to over 1000, and may include zero.

Results

Integrating the equations of formant values consistently provides significantly better data fits than a method that uses a means procedure. Vowel spaces can be more accurately depicted. Integral values of the equations were compared to the traditional singular linear average (means) values of F1, F2, and F3 for the purposes of depicting vowel spaces.

Figure 2a (left) shows the F1 plot of a female speaker of Hmong producing a high front vowel. Linear modelling of the data does not result in very accurate information, and a R^2 of 0.06. his F1 formant represents a modified wave function and so a wave equation can be fitted to the data, producing a much better fit as seen in Figure 2b (right). This is the same formant data as in Figure 2a but 2b has a R^2 of 0.89.

Integrating the Gaussian equation for each formant provides a more precise weighting of the information, which is a centroid value of the data points as seen in figure 3.

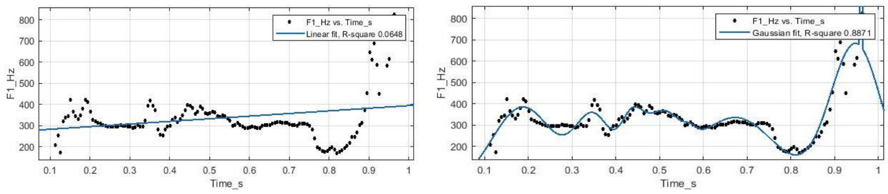

Figure 2a, 2b. 2a Left, The F1 of a high front vowel spoken by female speaker of Hmong, fitted with a linear equation (blue line). Right, the same vowel fitted with a Gaussian model that produces an equation that can be integrated (blue).

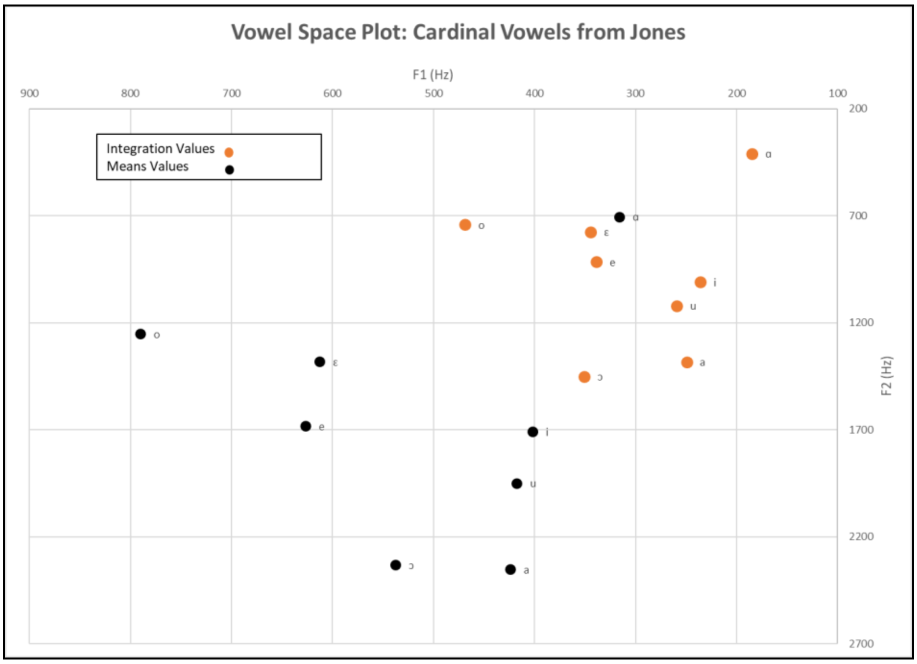

Figure 3. Vowel space plot of the cardinal vowels (Ladefoged, 2001). The red dots represent integrated vowel values, while the black dots depict average values of the formants. Vowels were not shown between slashes.

Summary

This study delivered an effective model of formants which can produce integrated vowel space values that are more robust than the values created by means calculations.

References

Ellis, D. 2010. An Introduction to Signal Processing for Speech. In William Hardcastle, et al., The Handbook of Phonetic Sciences. New York: John Wiley & Sons.

Elmore, W., Heald, M. 1985. Physics of Waves. New York: Dover.

Gold, B., Morgan, N. 2000. Speech and Audio Signal Processing. NY: John Wiley & Sons.

Kreyszig, E. 1983. Advanced Engineering Mathematics. Hoboken, NJ: John Wiley & Sons.

Ladefoged, P. 2001. A Course in Phonetics. Harcourt College Publishers: Fort Worth, TX.

Pain, H.J. 1993. The Physics of Vibrations and Waves. NY: John Wiley & Sons.

Parker, S. 1988. Acoustics Sourcebook. NY: McGraw-Hill Book Company.

Stein, Sh. 1973. Calculus and Analytic Geometry. New York: McGraw-Hill.

Thomas, G., Weir, M., Haas, J., Giordano, F. 2005. Thomas' Calculus, 11th Edition. Boston: Addison-Wesley Longman, Inc.