TOOL: The Rosetta Stone method of Digital Signal Processing (DSP) is a graphical approach for analyzing the spectra of both analog and digitized signals in a Fourier transform framework. The following ingredients constitute the Rosetta Stone method:

1) Time-domain waveforms for successive blocks in a signal processing system are sketched are sketched on the left side of the page, starting at the top of the page. The frequency-domain spectrum for each time domain waveform is sketched on the right side of the page, to the right of the corresponding time-domain waveform.

2) Key Fourier transform pairs (time domain, frequency domain) are easy to sketch and play key roles in analyzing spectra of signals:

   - $\delta(t) < \mathcal{F} > 1$ — impulse function and constant
   - $\text{III}(t) < \mathcal{F} > \text{III}(f)$ — impulse train (or shah or comb function)
   - $e^{-\pi x^2} < \mathcal{F} > e^{-\pi f^2}$ — Gaussian function
   - $\text{rect}(t) < \mathcal{F} > \text{sinc}(f)$ — rectangle function and $\sin(\pi f)/\pi f$
   - $\cos(\pi t) < \mathcal{F} > \text{I I}(t)$ — cosine and impulses $\frac{1}{2}\left[\delta\left(t+\frac{1}{2}\right)+[\delta\left(t-\frac{1}{2}\right)\right]$

   The above transforms apply when time and frequency are exchanged

3) Convolution in the time domain corresponds to multiplication in the frequency domain, and vice versa. Convolution may be approximately sketched in many cases by smoothing and widening a waveform.

4) Convolution of a function with an impulse function replicates the function with the origin translated to the location of the impulse.

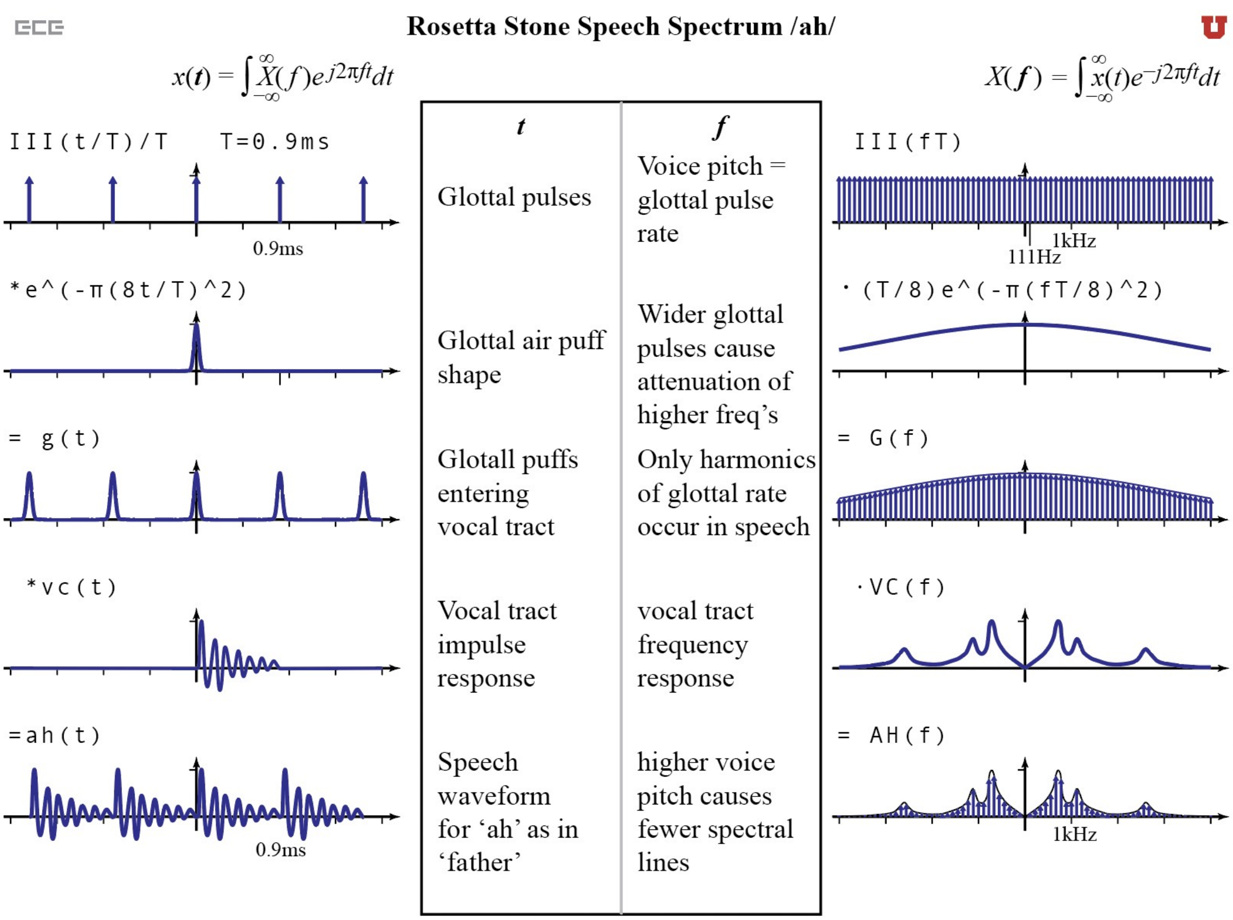
5) Convolutions of functions consisting of constant segments or sloped segments may be computed by the following steps: differentiating the waveform until impulses are obtained, performing the convolution with the impulses (which merely replicates impulses), and integrating as many times as the original waveform was differentiated.

6) Sampling (or digitizing) signals corresponds to multiplication by an impulse train (or shah function) in the time domain and replication of the spectrum of

the signal at multiples of the sampling frequency. This result may be viewed in a continuous-time and continuous-frequency framework.

7) Stretching out a waveform in time corresponds to compressing it in the frequency domain and vice versa.

An example of the Rosetta Stone analysis of speech production of the vowel sound 'ah' is shown below, followed by a text description that provides a more detailed exposition of the Rosetta Stone method and its benefits.



line 1) The vocal cords are flaps of skin that slap together at the rate of the voice pitch. During a vowel, the rate of the flaps is nearly constant, and an impulse train (or comb or shah) function captures this timing. In the

frequency domain, the voice pitch and all its harmonics are present as impulses. In the frequency domain, these spikes correspond to pure tones.

line 2) The vocal cords coming together produce glottal pulses, or sharp puffs of air, that drive the vocal tract, which acts like a filter. The puffs of air are a bit wider than impulses, which are an abstract idea requiring a signal with infinitely high frequency content. A Gaussian shape convolved with the impulse train gives a good approximation to the true waveform for the glottal pulses. In the frequency domain, the narrow Gaussian glottal pulse shape results in a wide Gaussian shape in the frequency domain. Thus, a narrower glottal pulse shape contains more high frequencies.

line 3) Each impulse in time becomes a Gaussian shape. In the frequency domain, we get an impulse train multiplied by a wide Gaussian function. If the glottal pulses in the time domain get narrower, the Gaussian envelope in the frequency domain gets wider and flatter. The implication is that narrower glottal pulses retain higher frequencies better than wider glottal pulses. The Rosetta Stone method yields this conclusion immediately, by inspection, using the rule that a narrower signal in the time domain results in higher frequency content in the frequency domain.

line 4) The output of a filter in the time domain is the convolution of the input signal with the impulse response of the filter. The vocal tract acts like a filter with the spectrum shown in the frequency domain [1]. The impulse response [2] is somewhat like a decaying sinusoid but with a few variations that give rise to peaks in the spectrum referred to as formants. The frequency ratios and relative heights of the formants play a primary role in encoding the type of vowel sound.

At this stage of the process, the spectrum of the impulse response is of more interest that the time-domain waveform itself, and a rough sketch of the impulse response would suffice.

line 5) The vowel sound for 'ah' in the time domain will be a repetitive waveform resulting from the convolution of the impulse response with the glottal pulse train. The convolution of the impulse response with the glottal pulse train will be a slightly smoothed version of repeated copies of the impulse response. This conclusion follows from the idea that convolution with an impulse makes a copy of a waveform, and convolution with a narrow pulse creates a slightly widened and smoothed copy of the impulse response.

Here, we are again more interested in the frequency domain spectrum, and the slight smoothing of the impulse response is omitted.

In the frequency domain, we see that the spectrum of the 'ah' sound is actually a set of spikes spaced by the voice pitch rate with an envelope corresponding to the filtering effect of the vocal tract.

From the graphical spectrum we may immediately draw several conclusions:

- The vocal tract filter is only "illuminated" at the harmonics of the voice pitch. The complete details of the vocal tract's frequency response are lost if the voice pitch remains exactly constant. Slight variation of the voice pitch can reveal more details of the vocal tract's frequency response.
- Taking this a step further, the peak of a formant frequency might lie between illuminated points of the vocal tract frequency response, which might make it difficult to understand which vowel sound was intended.
- A lower voice pitch gives more tightly spaced harmonics and illumination points for the vocal tract. Thus, a lower voice pitch may reveal more information about the vocal tract.
- As noted earlier, a narrower glottal pulse will increase the height of the peak at the higher-frequency formant, perhaps making it clearer relative to ambient noise. Conversely, raising the highest formant peak might alter the interpretation of what the vowel the sound is meant to be.
- Taken together, the preceding observations suggest that a person must learn how to control the interaction between their voice pitch,

glottal pulse shape, and vocal tract shape (that determines the filtering) in order to produce intelligible vowel sounds. The graphical Rosetta Stone method leads quickly to this conclusion.

REFS:

[1] http://hyperphysics.phy-astr.gsu.edu/hbase/Music/vowel.html accessed 01/14/2022.

[2] https://acousticscale.org/wiki/index.php/File:Vowel_Waveform_and_Spectrum.pdf.html accessed 01/14/2022.