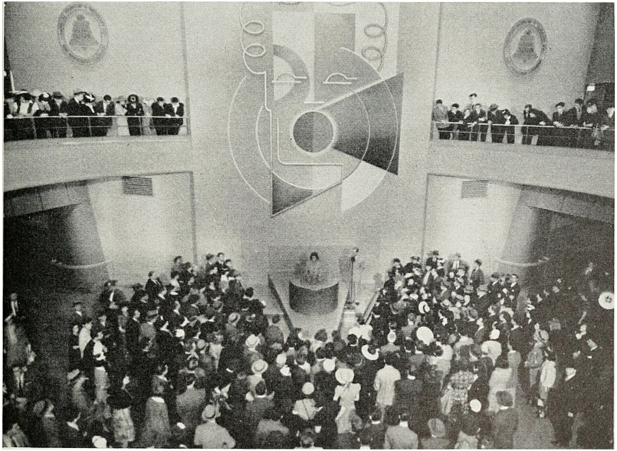
Lecture 4: The Vocoder

Mark Hasegawa-Johnson

ECE 537: Speech Processing Fundamentals

Outline

First Speech Synthesis: The Voder, 1939



Flickr Commons, https://commons.wikimedia.org/wiki/File:VODER_demonstrated_on_1939_New_York_World_Fair_-_The_VODER_fascinates_the_crowds_-_Bell_Telephone_Quarterly_(January_1940).jpg

→Watch Video←

First Speech Coder: The Vocoder, 1940

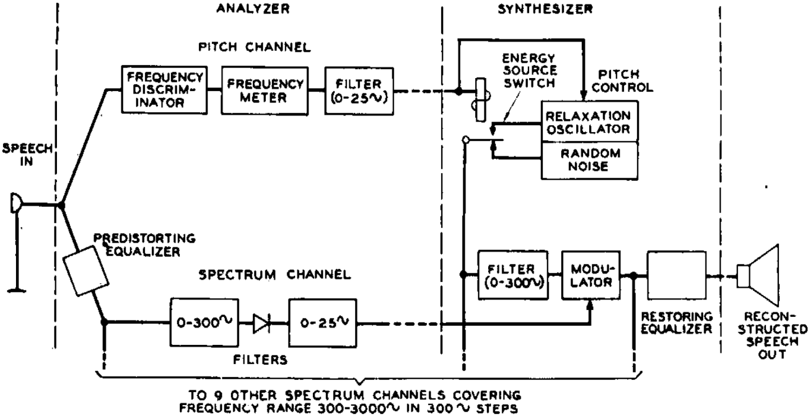


FIG. 2. Schematic arrangement of the Vocoder.

What background do you need in order to understand the vocoder?

- What is the spectrum of a "relaxation oscillator"?
  - To answer this question, we need to learn about the Discrete-Time Fourier Series (DTFS).
  - What happens when you bandpass filter it?
  - What happens when you adjust its level?
- What is the spectrum of a "random noise"?
  - To answer this question, we need to learn about autocorrelation and the power spectrum.
  - What happens when you bandpass filter it?
  - What happens when you adjust its level?

Recommended Textbooks

For this article, Licklider's article, and Atal's article, you'll find any of these four textbooks extremely useful. I learned from the 1978 text; these days I mostly refer to the 2010 text; the 2002 text has a little additional material that is missing in the others, but nothing critical for this course:

- Thomas F. Quatieri, *Discrete-Time Speech Signal Processing*, 2002
- Lawrence R. Rabiner & Ronald W. Schafer,
  - *Digital Processing of Speech Signals*, 1978
  - *An Introduction to Digital Speech Processing*, 2007
  - *Theory and Applications of Digital Speech Processing*, 2010

Outline

Voiced Speech, Step 1: Relaxation Oscillator

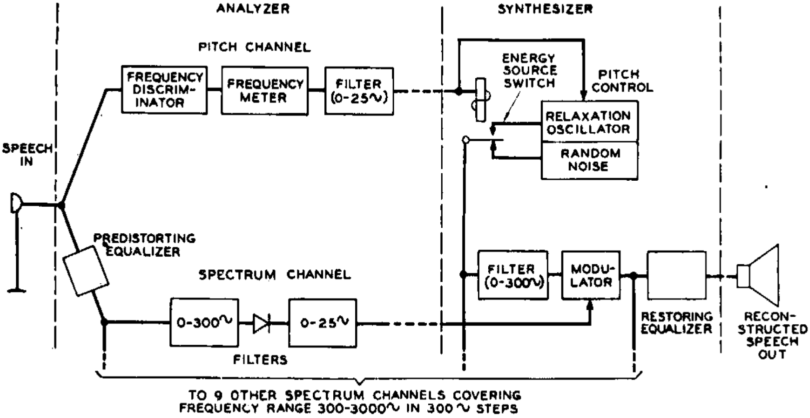
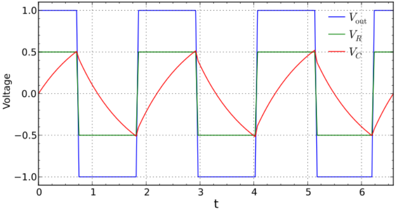


Fig. 2. Schematic arrangement of the Vocoder.

What is a “Relaxation Oscillator”?

A relaxation oscillator generates a square wave by switching the output voltage every time the internal capacitor voltage exceeds half the output (e.g.: for automobile turn signals).



Spectral Analysis: Transforms You Know

	Continuous Time, Infinite Frequency	Discrete Time, Periodic Frequency
Periodic Time, Discrete Frequency	Continuous Time Fourier Series (CTFS)	Discrete Fourier Transform (DFT), a.k.a. Discrete-Time Fourier Series (DTFS)
Infinite Time, Continuous Frequency	Continuous-Time Fourier Transform (CTFT)	Discrete-Time Fourier Transform (DTFT)

Discrete-Time Fourier Series

Suppose a signal, $x[n]$, is periodic, with period N, then it can be written:

$$x[n] = \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi kn}{N}},$$

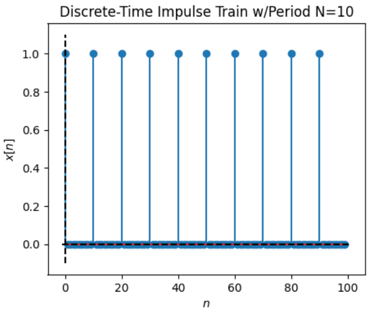
where the coefficients are

$$X_k = \frac{1}{N}\sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi kn}{N}}$$

Example: Impulse Train

For example, suppose $x[n]$ is the discrete-time impulse train

$$x[n] = \sum_{m=-\infty}^{\infty} \delta[n - mN]$$



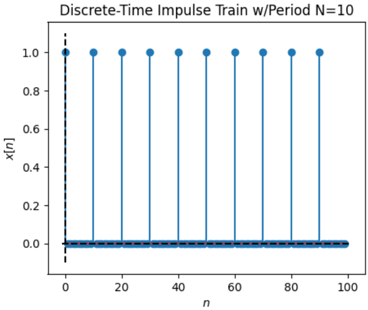
Example: Impulse Train

Fourier Series formulas say we can write $x[n]$ as

$$x[n] = \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi kn}{N}},$$

where the coefficients are

$$X_k = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi kn}{N}}$$



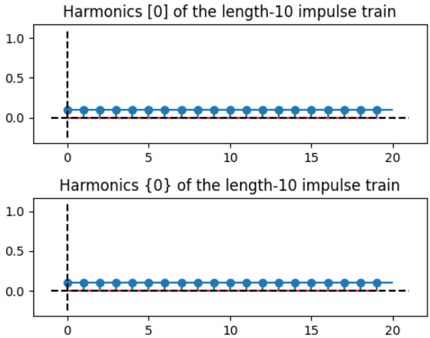
Example: Impulse Train

$$X_k = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi kn}{N}}$$

$$= \frac{1}{N}\left(x[0] + \sum_{n=1}^{N-1} x[n]e^{-j\frac{2\pi kn}{N}}\right)$$

$$= \frac{1}{N}(1+0)$$

$$= \frac{1}{N}$$



Outline

Properties of the Discrete-Time Fourier Series

- Linearity
- DTFT
- Convolution
- Parseval's Theorem

Linearity

If $x[n]$ and $y[n]$ have the same period,

$$x[n] = \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi kn}{N}}$$

$$y[n] = \sum_{k=0}^{N-1} Y_k e^{j\frac{2\pi kn}{N}},$$

then for any real or complex constants a and b,

$$z[n] = ax[n] + by[n] \quad \leftrightarrow \quad Z_k = aX_k + bY_k$$

DTFT

If a signal has a Fourier series,

$$x[n] = \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi kn}{N}},$$

then its discrete-time Fourier transform (DTFT) is given by

$$X(\omega) = 2\pi \sum_{k=0}^{N-1} X_k \delta\left(\omega - \frac{2\pi k}{N}\right)$$

where $\delta(\omega)$ is the Dirac delta.

Convolution

Suppose $x[n]$ is periodic with N,

$$x[n] = \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi kn}{N}},$$

and

$$y[n] = h[n] * x[n].$$

Then $yn]$ is also periodic with N, and

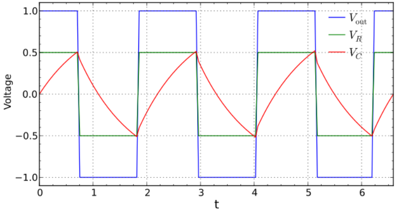
$$Y_k = X_k H\left(\frac{2\pi k}{N}\right),$$

where $H(\omega)$ is the DTFT of $h[n]$.

Convolution Example: Square Wave

For example, suppose we want the Fourier series of an even-symmetric square wave with period N:

$$y[n] = \begin{cases} 1 & |n - mN| \le \frac{L-1}{2} \\ & \forall \text{ integer } m \\ -1 & \text{otherwise} \end{cases}$$



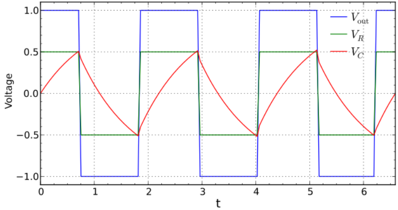
Convolution Example: Square Wave

We can find its Fourier Series by realizing that

$$y[n] = -1 + 2x[n] * h[n],$$

where $x[n]$ is an impulse train, and $h[n]$ is a rectangle:

$$h[n] = \begin{cases} 1 & -\frac{L-1}{2} \le n \le \frac{L-1}{2} \\ 0 & \text{otherwise} \end{cases}$$



CC-SA 3.0, `https://commons.wikimedia.org/wiki/File:Opamprelaxationoscillator.svg`

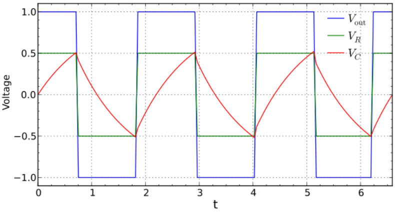
Convolution Example: Square Wave

We can find its Fourier Series by realizing that (except for the constant offset term, Y_0),

$$Y_k = 2H\left(\frac{2\pi k}{N}\right) X_k$$

where $X_k = \frac{1}{N}$ (because it's an impulse train!), and

$$H(\omega) = \text{DTFT}\left\{\text{rectangle}\right\}$$
$$= \frac{\sin(\omega L/2)}{\sin(\omega/2)}$$



CC-SA 3.0, `https://commons.wikimedia.org/wiki/File:Opamprelaxationoscillator.svg`

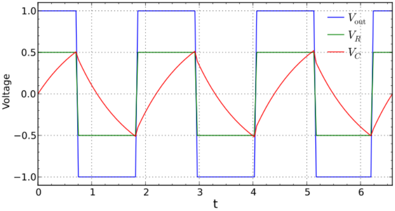
Convolution Example: Square Wave

Plugging in $\omega = \frac{2\pi k}{N}$, we get that, for $k \neq 0$,

$$Y_k = 2\frac{\sin(\pi k L/N)}{N\sin(\pi k/N)}$$

This is called a $1/f^2$ spectrum, a.k.a. https://en.wikipedia.org/wiki/Brownian_noise, because

$$|Y_k|^2 = \begin{cases} 0 & kL = \text{integer multiple of } N \\ \frac{1}{\mathcal{O}\{k^2\}} & \text{otherwise} \end{cases}$$



https://commons.wikimedia.org/wiki/File:Opamprelaxationoscillator.svg

Parseval's Theorem

…and speaking of $|Y_k|^2$…

Parseval's theorem says that power in the time domain equals power in the frequency domain:

$$\frac{1}{N}\sum_{n=0}^{N-1}|y[n]|^2 = \sum_{k=0}^{N-1}|Y_k|^2$$

This is especially useful after filtering. If $y[n] = h[n] * x[n]$, it might be difficult to calculate power in the time domain, but in the frequency domain, you just use $Y_k = H\left(\frac{2\pi k}{N}\right) X_k$, square, and add.

Outline

Voiced Speech, Step 2: Bandpass Filter the Buzz

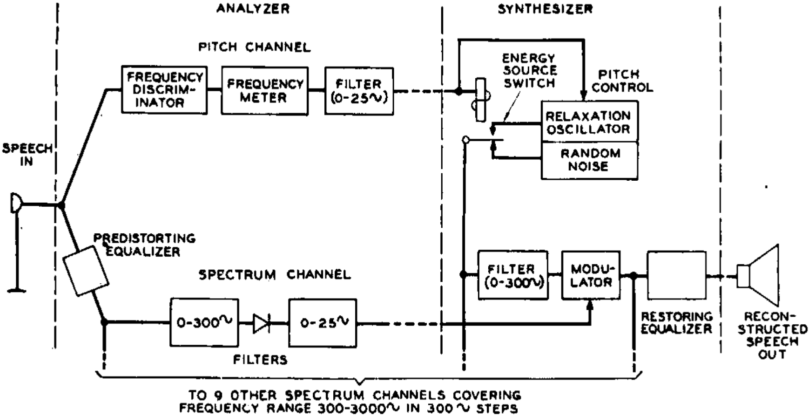


FIG. 2. Schematic arrangement of the Vocoder.

Filtering a Periodic Signal

Suppose $x[n]$ is periodic:

$$x[n] = \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi kn}{N}},$$

and we bandpass filter it with a filter $h[n]$:

$$y[n] = h[n] * x[n],$$

then $y[n]$ is periodic with Fourier series coefficients given by:

$$Y_k = H\left(\frac{2\pi k}{N}\right) X_k$$

Bandpass Filtering a Periodic Signal

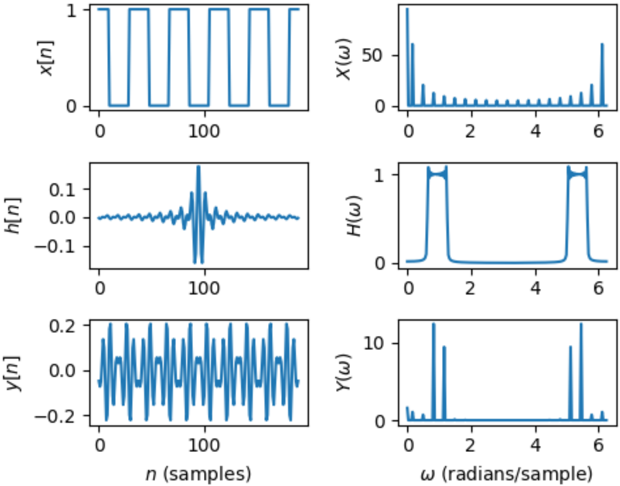
In particular, suppose that $H(\omega)$ is an ideal bandpass filter with center frequency α and bandwidth β:

$$H(\omega) = \begin{cases} 1 & \alpha - \frac{\beta}{2} \le |\omega - 2\pi m| < \alpha + \frac{\beta}{2}, \quad \forall \text{integer } m \\ 0 & \text{otherwise} \end{cases}$$

Then Y_k is just a selection of a few of the harmonics of X_k:

$$Y_k = \begin{cases} X_k & \alpha - \frac{\beta}{2} \le \left|\frac{2\pi k}{N} - 2\pi m\right| < \alpha + \frac{\beta}{2}, \quad \forall \text{integer } m \\ 0 & \text{otherwise} \end{cases}$$

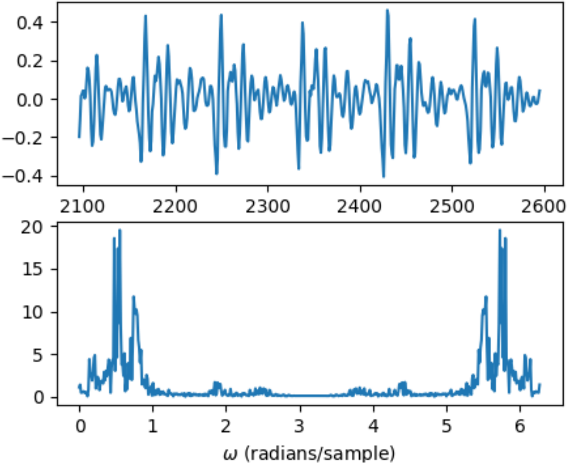
Voiced Speech, Step 2: Bandpass Filter the Buzz



Outline

Real Speech

Waveform of an open back vowel, and its magnitude FFT



→Listen←

How to Synthesize a Vowel

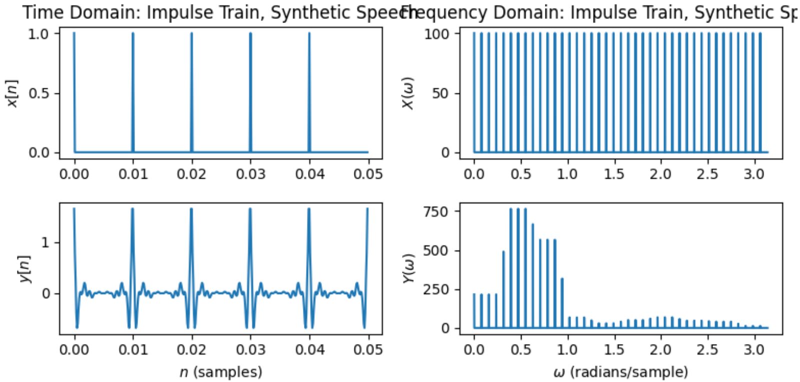
Start with an impulse train (Dudley used a relaxation oscillator = square wave, but an impulse train is better because it has a flat spectrum):

$$X_k = 1$$

Filter by a set of 10 bandpass filters $H_l(\omega)$, each about 300Hz wide, then adjust the amplitude of each one (A_l) to match the amplitude of the speech signal in the same band:

$$y[n] = \sum_{l=1}^{10} A_l \sum_{k=0}^{N-1} H_l\left(\frac{2\pi k}{N}\right) X_k$$

Synthetic Speech



→Listen← Spectrum is about right. Pitch is too monotone. Time-domain $h[n]$ should be causal to better represent speech.

Outline

Conclusions: Attributes of DTFS that made Vocoder Possible

1. The Fourier Series of an impulse train is $X_k = 1$
2. Convolution:

$$y[n] = h[n] * x[n] \quad \leftrightarrow \quad Y_k = H\left(\frac{2\pi k}{N}\right) X_k$$

3. Linearity:

$$Z_k = aX_k + bY_k \quad \leftrightarrow \quad z[n] = ax[n] + by[n]$$

4. . . . and one more that will help on the homework: Parseval's theorem.

$$\frac{1}{N}\sum_{n=0}^{N-1} |x[n]|^2 = \sum_{k=0}^{N-1} |X_k|^2$$