Einfhrung in die Sprachverarbeitung

Timo Gerkmann

Winter Semester 2013-2014

1 Introduction

2 Speech Fundamental Frequency

As introduced in the previous sections, speech can be modelled as being produced by two types of excitiation. Unvoiced speech is rather noise- like, lacking a periodic structure. It is created from air flow being blown through the vocal tract by the lungs. The position of the vocal tract gives a spectral shape to this turbulent air flow. Voiced speech is generated by the glottus opening and closing, thus regulating this air flow in a periodic manner. The period of this opening and closing is referred to as the speech fundamental period, the inverse of which is the fundamental frequency.

Speech fundamental frequency is often synonymously used with the term pitch. It is important to note,however, that speech fundamental frequency is a quantitative value that is associated with the opening and closing of the glottus, whereas pitch is more qualitative, influenced by the loudness, length, as well as frequency of the speech.

Because fundamental frequency has this effect upon our perception of speech, it becomes an important parameter in speech signal processing and has important implications in speech coding, enhancement, modeling, and recognition. It is therefore necessary to develop tools in which this parameter can be estimated. For this, the advantages and disadvantages of several methods are explored.

2.1 Residual effect

It is first important to note that fundamental frequency can still be estimated from its harmonics, even when it is not, itself, present in the signal. This is exemplified by telephone speech which is generally band-pass filtered between 300Hz and 3400Hz in order to minimize bandwidth per user. This range preserves the formants necessary for speech comprehension, however does not include the fundamental frequency.

How the fundamental frequency is still preserved is obvious when we look at the superpostion of two harmonics of a fundamental frequency. If a signal has a fundamental frequency of 100Hz, there will be harmonics at 200Hz,300Hz, etc. If this signal is high-pass filtered at 150Hz, the perceived signal will be a superpostion of the harmonics above 200Hz. As can be seen in the accompanying figure, the sum of a 200Hz tone and 300Hz still displays a fundamental period of $\frac{1}{100Hz}$, however the 100Hz tone is still not present in the frequency spectrum.

2.2 Fundamental Freq Estimation by zero-crossing and peak measurement

Prone to errors and hard to automize in an algorithm.

2.3 Fundamental Freq Estimation by autocorrelation fucntion

.

We now define the autocorrelatiion function as:

$$\varphi_{xx}(\lambda) = E(x(n)x^*(n+\lambda) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv \tag{1}$$

A signal is white if succesive smaples of the signal are uncorrelated. This implies that it has a flat power spectral density and has only one peak at lag zero.

For speech, succesive samples are correlated therefore we will see peaks at lag zero and multiples of the fundamental period.

Window length must be longer than the fundamental period, but not so long that it cannot account for changes in the fundamnetal frequency.

3 Spectral Analysis of Speech Signals

3.1 Discrete Fourier Transform

WHy do we need it? We need to modify a signal. The properties we want to modify must be easily accesible. Example: fundamental freq is difficult to see in time domain, but easier to see in frequency domain.

How? We perform Fourier Analysis. We correlate our signal with sine and cosine functions to find its frequency content.

3.1.1 Fourier Series Decomposition

Only with Harmonics

3.1.2 Continuous Time Fourier Transform

Now with all values of omega

3.1.3 Discrete Time Fourier Transform

Discrete in Time Domain, but continuous in frequency domain Properties

3.2 Short Time Fourier Transform

4 Model of the Vocal Tract

4.1 Tube model of Vocal tract

SPeech sounds voiced and unvoiced

Needed Parametric Model for vocal tract filter function, and excitiation signal

Two state Model

We model the vocal tract as a tube(ignore the Velum) so nasals are not well modeled:

To blackboard:

Find mathematical model for vocal tract

Pressure waves in a tube

Tube Segments change shape

Solve the differential equations

- Approach for solution Combination of forward and backward travelling wave
- for pressure, forward and backward travelling waves add
- for velocity, forward and backward travelling waves subtract
-

$$ \tag{2} $$

4.2 Linear Prediction