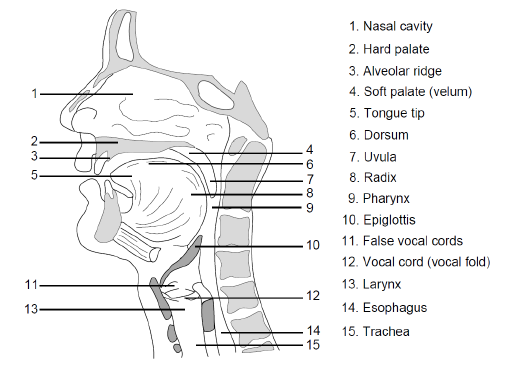
Physiology of Voice Production

# Components of the Speech Production System

* Lungs, or sub-glottal respiratory system. Acts as the source of energy which supplies airflow to the larynx through the trachea (windpipe).
* Larynx (voice box). Houses the vocal folds which open/close quasi-periodically. Opening between the vocal folds is the ‘glottis’.
* Vocal tract or supra-glottal area. Consists of the nose, mouth and pharynx.

Source filter model – modulated airflow from the glottis is the input to the filter, known as the glottal flow or glottal volume velocity waveform. Vocal tract is a variable acoustic filter which varies in shape when person speaks, giving the sound timbre, effectively shaping the spectral characteristics of the source.

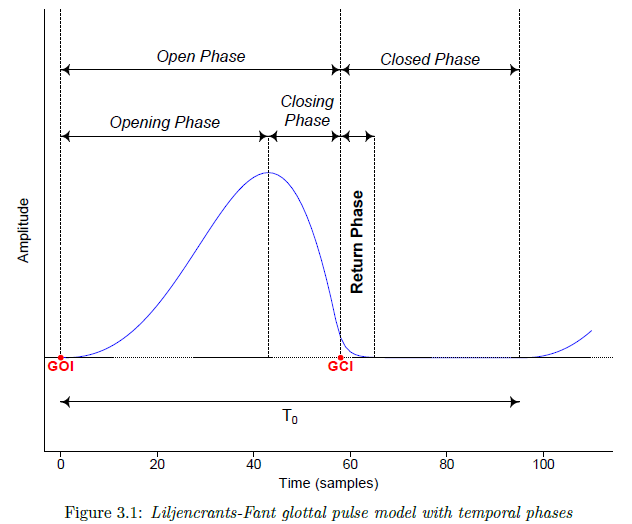


# Lungs and Respiratory System

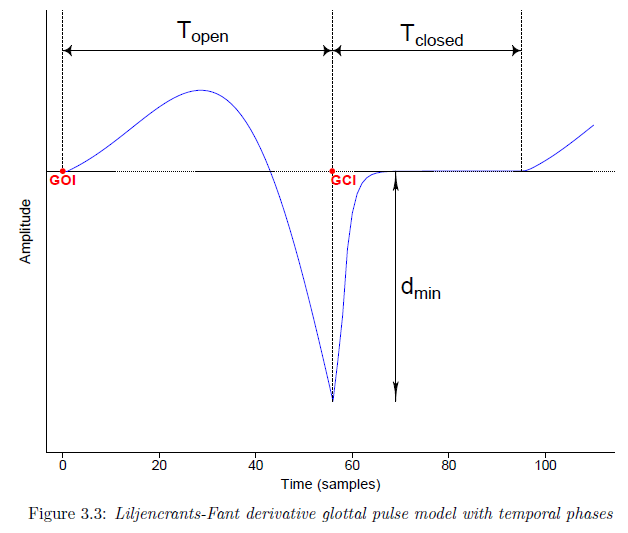
* Respiratory system = lungs, airways and respiratory muscles.
* Main purpose of lungs is the exchange of gases (oxygen/CO2). Oxygen is absorbed into the bloodstream during inhalation and CO2 is expelled during exhalation.
* Airways form path from lungs to body’s exterior, which includes the trachea, bronchi, and the glottal and supra-glottal organs.
* Collectively referred to as the sub-glottal respiratory system.

Glottal Source Signal

Glottal source signal/glottal flow is the input to the source-filter model. Can be obtained from a speech signal via source-filter deconvolution. Glottal waveform is continuous and differentiable, except for at instances of glottal closure (GCI). Below shows the Liljencrants-Fant parametric model, which can be fit to the glottal flow.



is the fundamental period of glottal cycle, is fundamental freq. of vocal fold vibration. The open-phase is the duration in which the vocal folds are open, closed phase is when the vocal folds are closed. Temporally divided by instance of maximum glottal flow. The derivative of the glottal flow is shown below.



The speed quotient (SQ) is a measurement of glottal pulse skewness, which quantifies the asymmetry of the glottal pulse,

Due to the mechanics of the vocal folds, abduction is longer than adduction, and so . Helpful for detection of voice pathologies (e.g. vocal fry) and provides a measure for glottal efficiency.

The open quotient (OQ) quantifies the width of the glottal pulse in relation to the duration of the glottal cycle,

The OQ reflects the behaviour of the vocal folds. Closely related to the close quotient (CQ) and the contact quotient (Qx). .

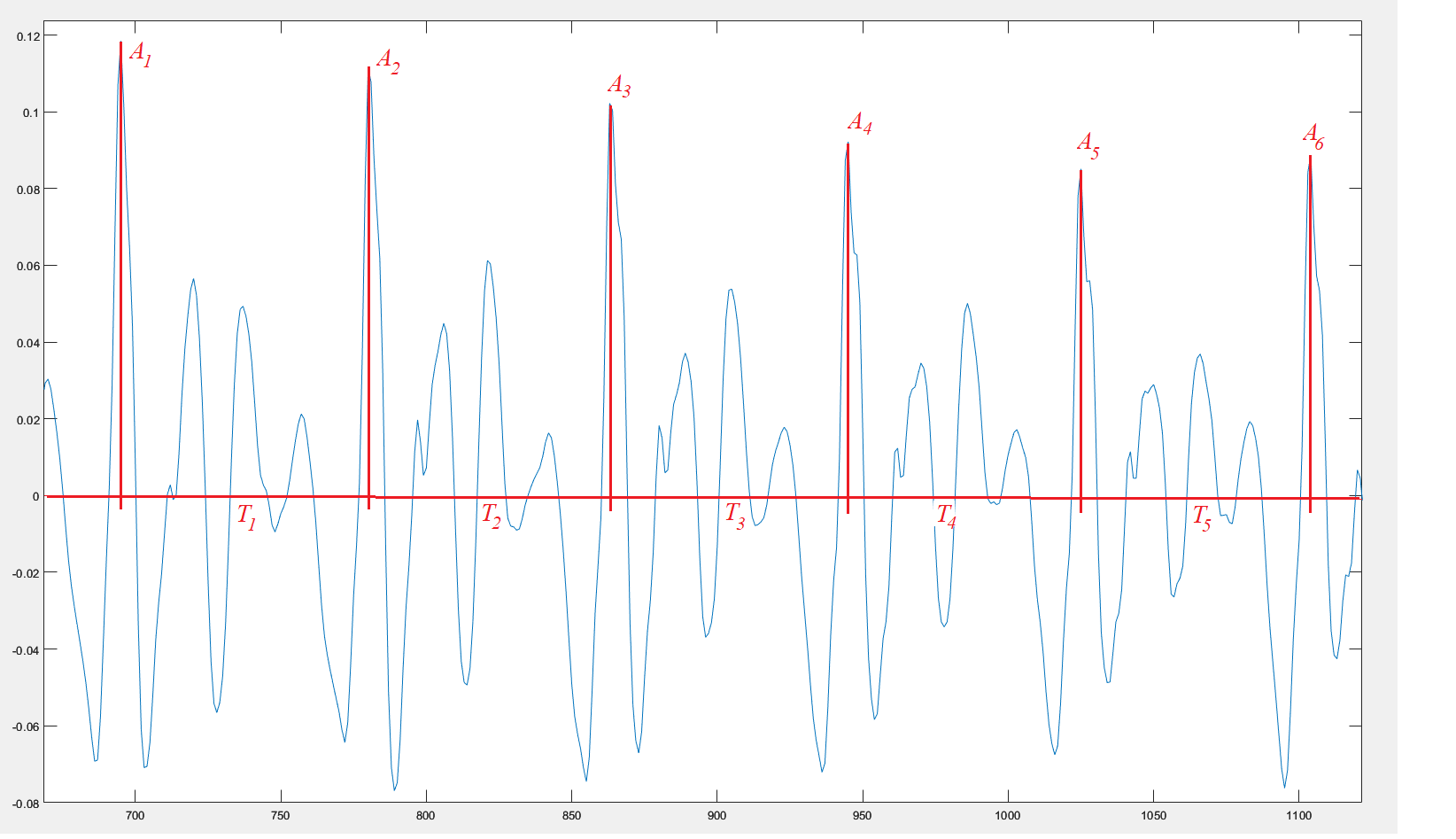
The closing quotient (and opening quotient) are similarly defined as

An alternative to the ClQ is the normalised amplitude quotient (NAQ),

The NAQ is more reliable and robust as it does not require estimating glottal opening and closing instances. is the peak-to-peak amplitude of the glottal flow, is the minimum negative amplitude of the glottal flow derivative.

Shimmer and Jitter

Consider a sample of voiced speech (vowel) shown below.



The shimmer quantifies cycle-to-cycle perturbations in the peak amplitudes. The local/relative shimmer quantifies the average absolute difference in amplitude between consecutive cycles as a percentage,

The shimmer can also be quantified using high order measures, which are called amplitude perturbation quotients (APQs). The APQ3, for example, quantifies the average absolute difference in amplitude between one cycle and the average of its neighbouring cycles on either side,

This can be generalised to the -th order APQ where is an odd positive integer,

where . The APQ quantifies the average absolute difference in amplitude between one cycle and the average of its neighbouring cycles on either side.

# Jitter

The jitter quantifies cycle-to-cycle perturbations in the fundamental frequency. The local/relative jitter quantifies the average absolute difference in frequency between consecutive cycles as a percentage,

The higher order measures of jitter are called period perturbation quotients (PPQs), except for the third order measure, which is called the relative average perturbation (RAP). The RAP is given by

and the -th order PPQ, for some positive off integer , is

where .

# MATLAB Implementation

The Troparion toolbox (<https://github.com/Mak-Sim/Troparion>) can be used for the calculation of both relative/local shimmer and jitter, as well as higher order APQs and PPQs.

clear;

addpath('..\Speech samples');

addpath('Troparion-master\IRAPT\IRAPT\_web');

filename = 'vowel.wav';

[s,Fs]=audioread(filename);

[Fo, ~, time\_marks] = irapt(s, Fs, 'irapt1','sustain phonation');

% Segmentation of signal onto fundamental periods

[periods] = WM\_phase\_const(s,Fo,time\_marks,Fs);

[amplitudes]= amp\_extract(periods,s);

% Perturbation parameters calculation

J\_loc = shim\_local(periods);

J\_rap = shim\_apq3(periods);

J\_ppq5 = shim\_apq5(periods);

J\_apq11 = shim\_apq11(periods);

J\_apq17 = shimmer\_apq(periods,17);

S\_loc = shim\_local(amplitudes);

S\_apq3 = shim\_apq3(amplitudes);

S\_apq5 = shim\_apq5(amplitudes);

S\_apq11 = shim\_apq11(amplitudes);

S\_apq17 = shimmer\_apq(amplitudes,17);

% Plot the speech signal

n = 0:size(s,1)-1;

plot(n,s)

For the vowel sound shown above, the calculated perturbation parameters are:

J\_loc =3.4907 J\_rap =0.8259 J\_ppq5 =0.7433 J\_apq11 =2.4222

J\_apq17 =5.9698 S\_loc =6.0298 S\_apq3 =2.6811 S\_apq5 =4.5740

S\_apq11 =8.0217 S\_apq17 =5.7511