# Investigation of Speaking and Singing Timbre Difference


## By: Haiqin Yin and Shuizening Li

University of Rochester


## Introduction

This project explores the sonic difference between spoken and singing voice in various aspects such as power, harmonic content, periodicity, pitch contour and consonant length, etc. by using functions in Matlab to visualize the sonic difference.

## Properties

### Consonant length

Consonant is "*a speech sound produced by occluding with or without releasing (p, b; t, d; k, g), diverting (m, n, ng), or obstructing (f, v; s, z, etc.) the flow of air from the lungs (opposed to vowel ).*" Because consonants often occur at the beginning and the end of a syllable and contributes to the power of the sound in forms of plosives, we can analyze the difference in length along with other attributes of consonants in speech voice and singing voice by visualizing the waveforms of the voices.

### Harmonic content/formant

One of the most important attributes of one's voice that contributes to the timbre of one's voice. Age, gender and body condition all have an impact on the harmonic content in one's voice. Comparing harmonic content in speech voice and singing voice is effective because the fundamental frequency won't be the factor since the fundamental frequency in speech voice doesn't match up with the singing voice most of the time.

### Pitch contour

Singing and speaking voices both have a underlying pitch contour, which outline the intonation and the melody of the voices. The melody of singing is divided into discrete pitches defined by different temperament. In contrast, the pitch contour of a natural spoken utterance does not follow any temperament and glides rather continuously.

### Rhythm/Periodicity

The main difference in rhythm between speaking and singing voice is periodicity. The rhythm of a singing voice is more strict than speaking voices as it sets boundaries to where the words need to be spoken and it could also break the pronunciation of a word into syllables. Speaking voices, on the other hand, only have restrictions on the word and phrase.

### Power

Power indicates the amplitude and energy of the speaking and singing voices can be analyzed with the power spectral density (PSD), which illustrates the strength of energy as a function of frequency



## Experiment

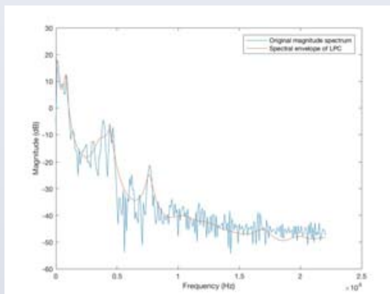
The subject, male, 22 years-old, read and sang the first line of the following 6 songs: *Don't Take Me Alive, Hey Jude, Layla, Long Distance Runaround, Patience, Pennyroyal Tea.*

## Result & Discussion

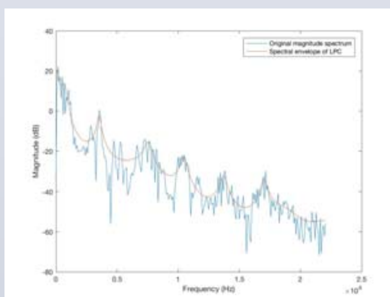
**Harmonic Content /Formant**

The LPC graph shows that in speaking voices, the power of harmonics decreases in a linear manner as the order of the harmonics goes up. On the other hand, in singing voices, the power of harmonics decreases faster in lower orders (i.e. 2nd and 3rd harmonic) and decreases slower in higher orders so there is more higher order harmonics present in singing voices than in speaking voices. In practice, these differences make the speaking voice sound fuller, darker and more masculine and singing voice sound brighter and less masculine.

Speaking:




Singing:



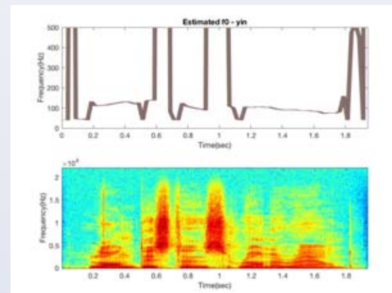

**Consonant Length**:

We used YIN Algorithm together with amplitude spectrum analysis to measure the length of consonants in speaking and singing voices. In YIN algorithm, because consonant has lower correlation with pitch, we identify the parts in the audio that don't have stable pitches as consonants and the parts with stable pitches and harmonics as vowels. On average, the consonants at the beginning of each syllable in singing voices are 50%-60% longer than the consonants in speaking voices.

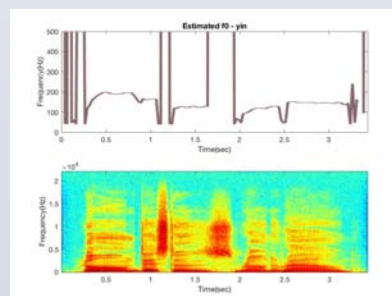
**Pitch Contour & Periodicity**

We used YIN Algorithm, a method of autocorrelation with modification, to find the pitch contour and periodicity of the voices. The diagram shows that the singing voices changes more successively, while speaking voices have some noticeable pauses in between pitches. The pitch of speaking voices does not changes a lot over time. It stabilized at a certain range of frequency of 100-200Hz. As for periodicity, the gap in between pitches shows that singing voices are more periodic than speaking voices.

Speaking:



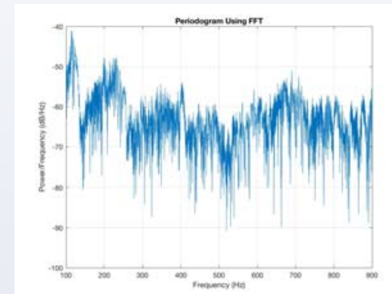

Singing:



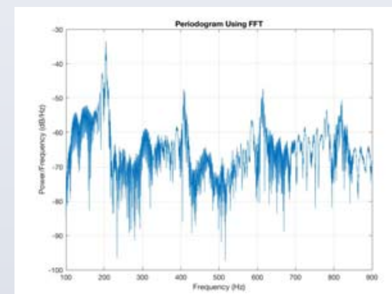

**Power (Power Spectral Density)**

Since the fundamental frequency range of a male voice is from 100-900Hz, we only obtained the power spectral density in this range. In general, the power spectral density of speaking voices is lower than singing voices, but they both follow a similar pattern. For singing voice, it is easier to determine the peaks at around 200Hz, 400Hz, 600Hz, which indicate the fundamental frequency and harmonics. Overall, speaking voices shows a higher power level for frequencies under 150Hz. The power spectral density graph also shows that the power of singing voices shifts more frequently, while the speaking voices changes more randomly than singing voices. This shows that the voices when people sing have more energy and are tend to be more centralized and stabilized than voices when people speak.

Speaking:




Singing:




## Conclusion & Application

The LPC shows that speaking voice is more prominent in second and third harmonics while singing voice excites more higher harmonics. The PSD shows that speaking voice is more constant in power while singing voice is more dynamic. YIN Algorithm shows that the pitches shift more smoothly in singing voices and speaking voices have more sudden changes and pauses. YIN Algorithm and the Amplitude Spectrum Analysis show that the consonants are usually longer in singing voices than in speaking voices.

This project finds several qualitative and quantitative difference between speaking and singing voices. These different features can be used in voice command detection, singing detection in music applications and singing synthesis, etc.


## Acknowledgement

Thanks to people who assisted on the research on this project, especially Dr. Zhiyao Duan, who introduced YIN Algorithm to us