

ICT 응용기술 Project 결과 발표

김성환
신현준

Project 진행 방향

1. **WORLD - a high-quality speech analysis, manipulation and synthesis system**를 이용한 목소리 변조 시스템

1. **Matlab**를 이용한 음성 변조 방법

WORLD 를 이용한 목소리 변조 시스템

Vocoder?

통신을 위한 음성 압축 기술,

휴대폰 등 다양한 기기에 사용되고 있다.

음성 파형을 직접 보내는 것이 아니라 매개 변수를 보내고,
수신 측에서는 그 매개 변수에서 원본 음성을 합성한다.

음악용 보코더는 이 기술을 응용한 것



WORLD 를 이용한 목소리 변조 시스템

WORLD?

A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications

Masanori MORISE, Fumiya YOKOMORI, Kenji OZAWA

265회 인용, github 463 Stars

WORLD?

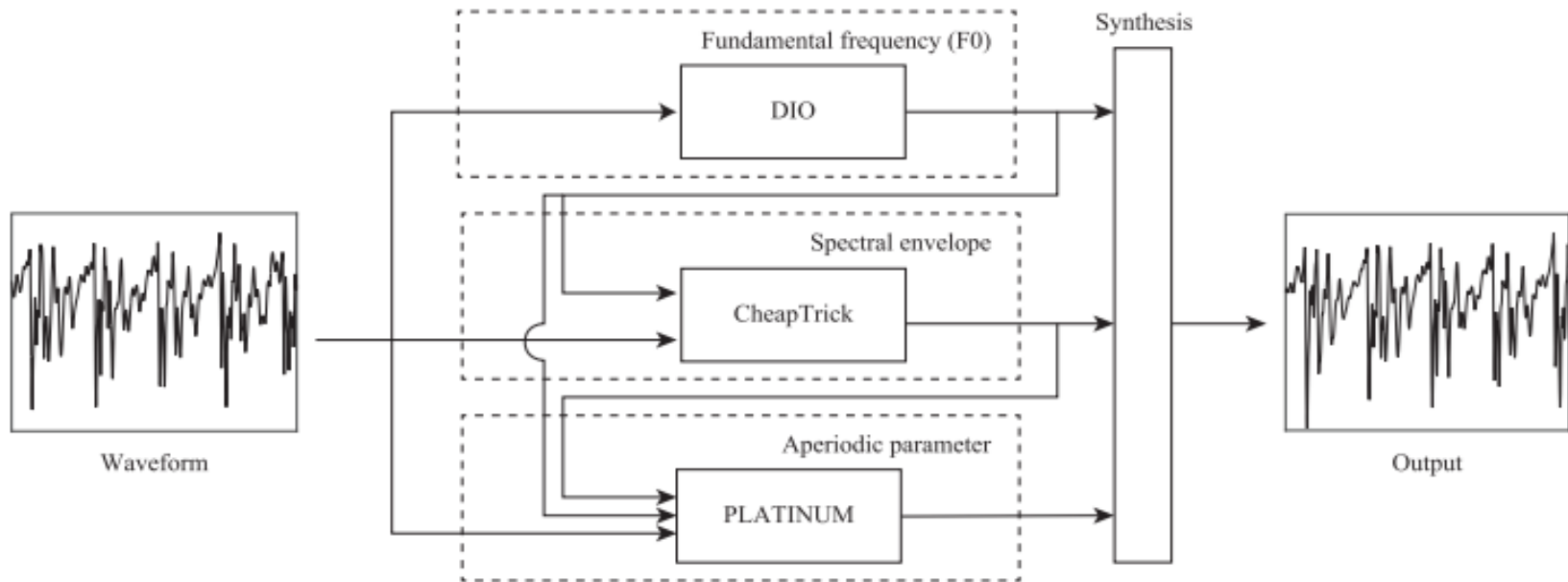


Fig.1 Overview of the developed system. WORLD consists of three analysis algorithms for determining the F0, spectral envelope, and aperiodic parameters and a synthesis algorithm incorporating these parameters.

Fundamental frequency(F0) estimation Alogorithm : DIO

F0 is defined as the inverse of the smallest period of a periodic signal

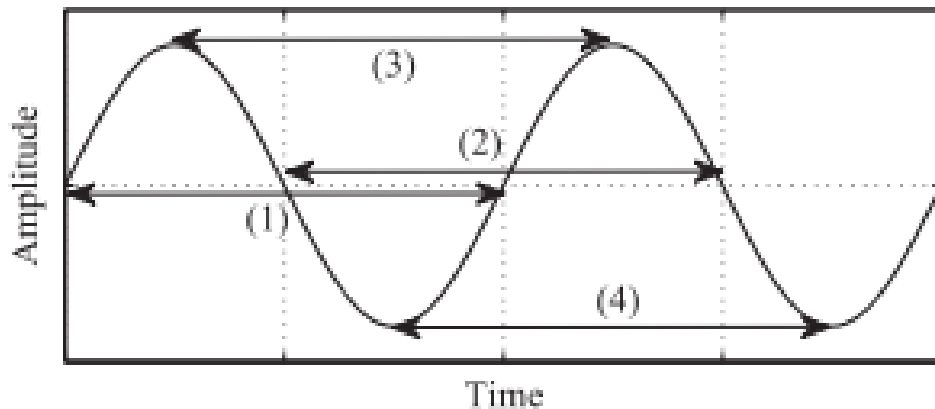
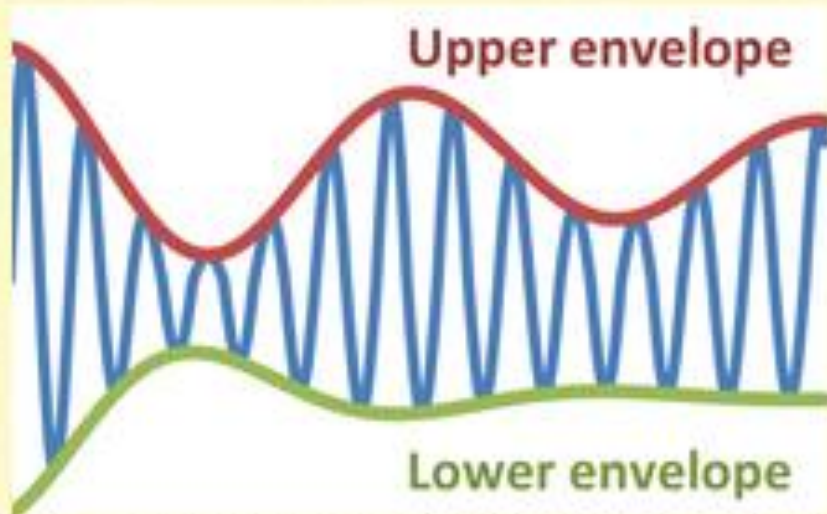


Fig. 2 Four intervals used for calculating an F0 candidate and its reliability. If the filtered signal only consists of the fundamental component, the four intervals indicate the same value.

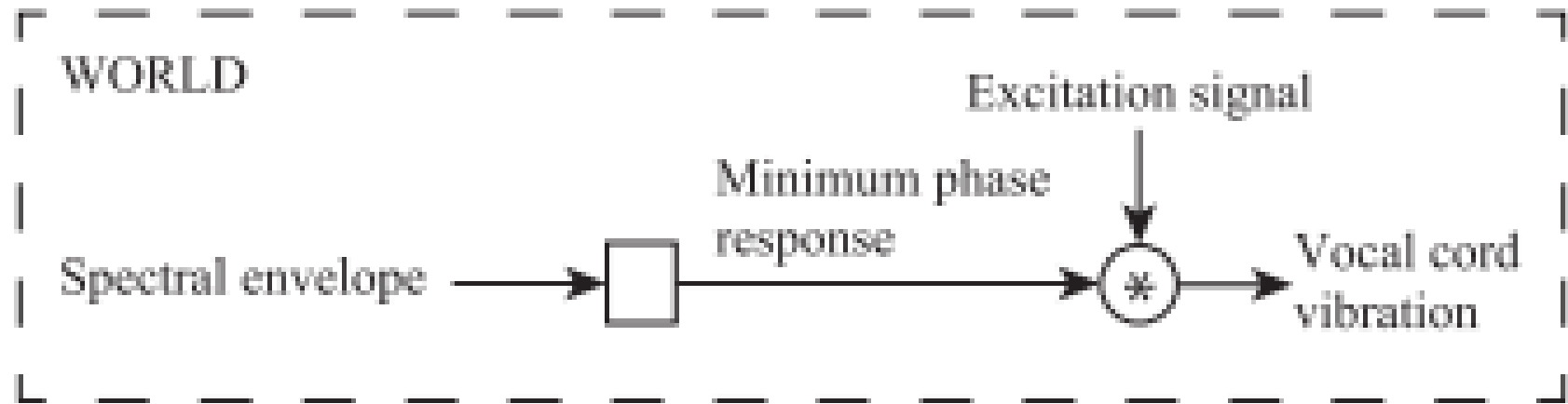
Spectral Envelope Estimation Algorithm : CheapTrick

Envelope : 파형의 최상점 또는 최하점들을 부드러운 곡선으로 연결한 선



[http://en.wikipedia.org/wiki/Envelope_\(waves\)](http://en.wikipedia.org/wiki/Envelope_(waves))

Aperiodic Parameter Extrction : PLATINUM



Evaluation

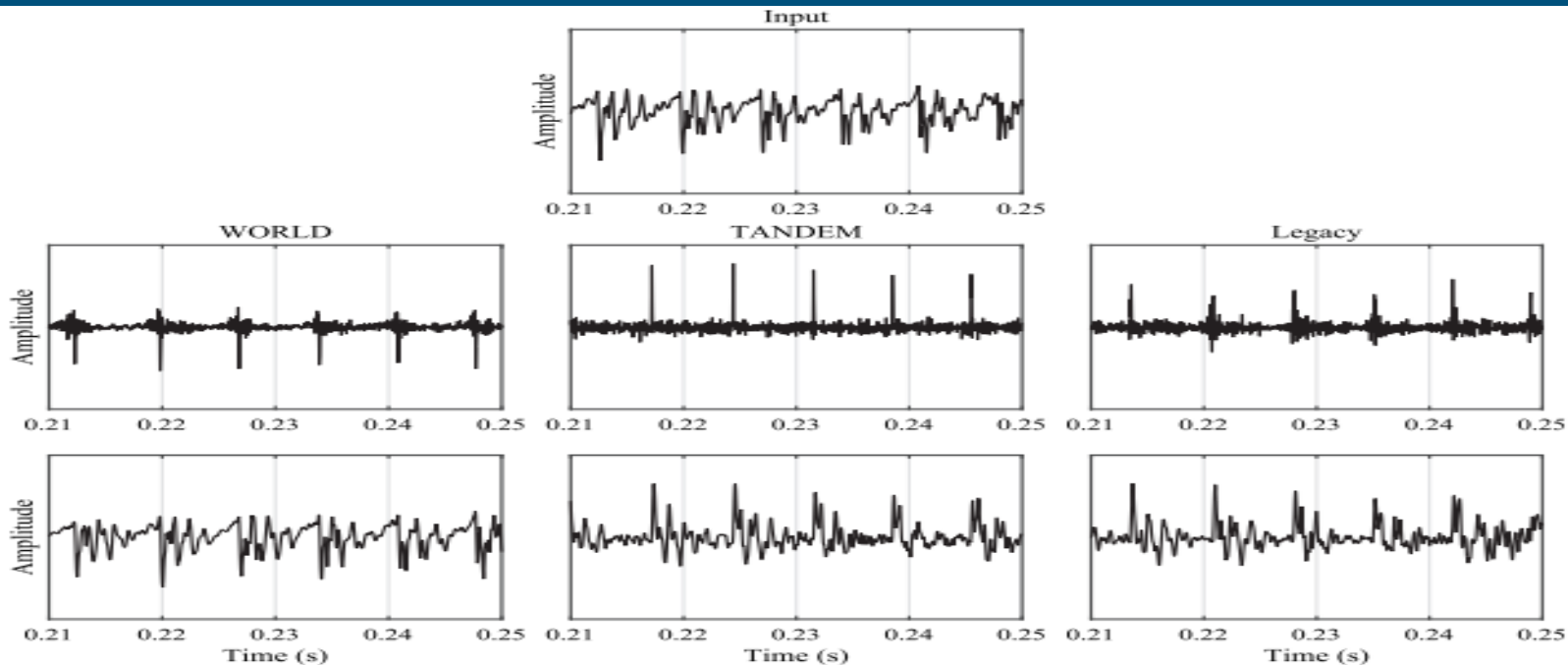
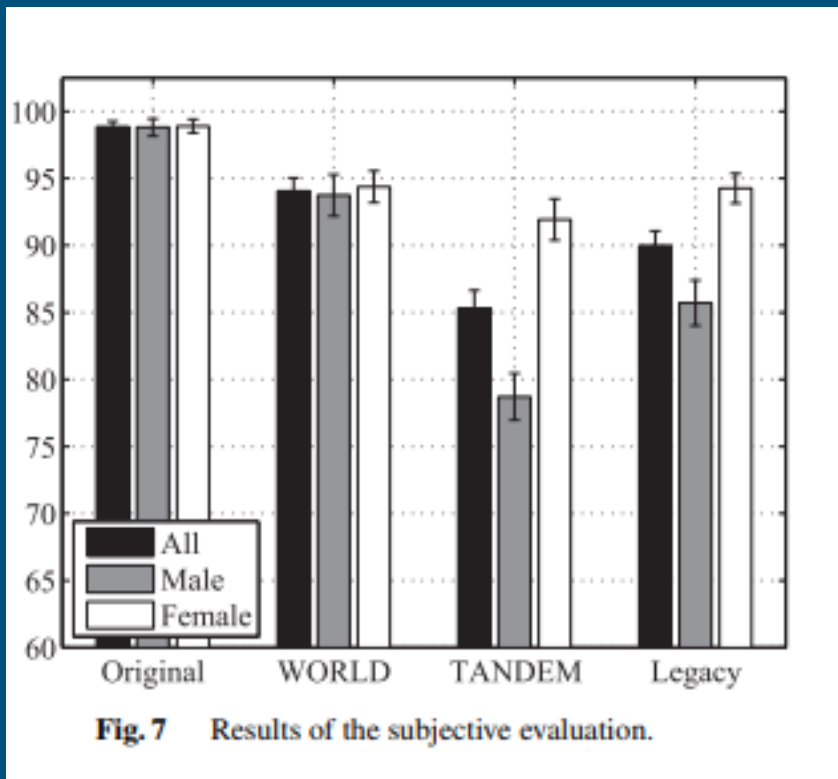


Fig. 4 Examples of input waveform, excitation signals, and synthesized waveforms. Top: Input waveform. Middle: Excitation signals before convolution. Bottom: Synthesized waveforms.

Evaluation



Implementation

```
void ParameterModification(int argc, char *argv[], int fs, int f0_length,  
    int fft_size, double *f0, double **spectrogram)  
{  
    double shift = atof(argv[3]);  
  
    for (int i = 0; i < f0_length; ++i) f0[i] *= shift;
```

주파수를 배로 늘리는 인자 주파수가 높아지면 남성 목소리가 여성화될 것으로 판단

WORLD 를 이용한 목소리 변조 시스템

Implementation

Micro

인자값

```
complete.
hj@DESKTOP-KJIBV0U: /mnt/c/GitHub/ICT_Project/World/build$ ./test valueo2d.wav output.wav 1
File information
Sampling : 22050 Hz 16 Bit
Length 17500 [sample]
Length 0.793651 [sec]

Analysis
Harvest: 298 [msec]
CheapTrick: 14 [msec]
D4C: 60 [msec]

Synthesis 1 (conventional algorithm)
WORLD: 14 [msec]

Synthesis 2 (All frames are added at the same time)
WORLD: 15 [msec]

Synthesis 3 (Ring buffer is efficiently used.)
WORLD: 14 [msec]
complete.
hj@DESKTOP-KJIBV0U: /mnt/c/GitHub/ICT_Project/World/build$ ./test valueo2d.wav output.wav 2
File information
```

Matlab을 이용한 음성 변조

목소리는 남자의 경우 120Hz, 여자의 경우 220Hz의 주파수를 가지고 있는 점을 이용, 주파수를 변조하여 다른 성별의 목소리를 생성해낸다.

1. 샘플을 STFT(Short Time Fourier Transform)을 이용해서 주파수 도메인으로 변환

- STFT : 일정 크기의 window를 끌고 다니면서 전체 데이터가 아닌 그 window 안에 들어온 작은 부분만 Fourier Transform

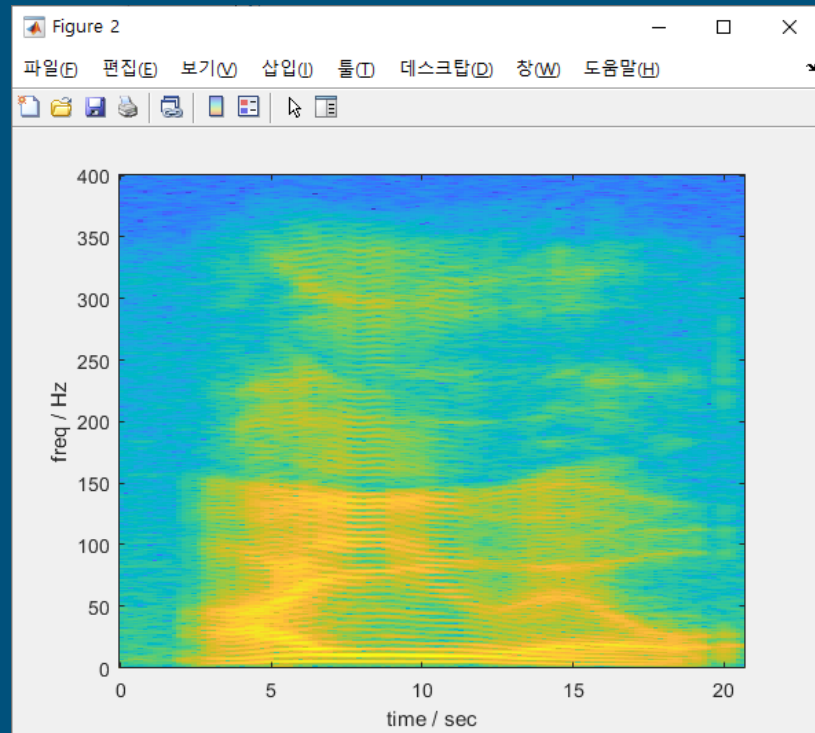
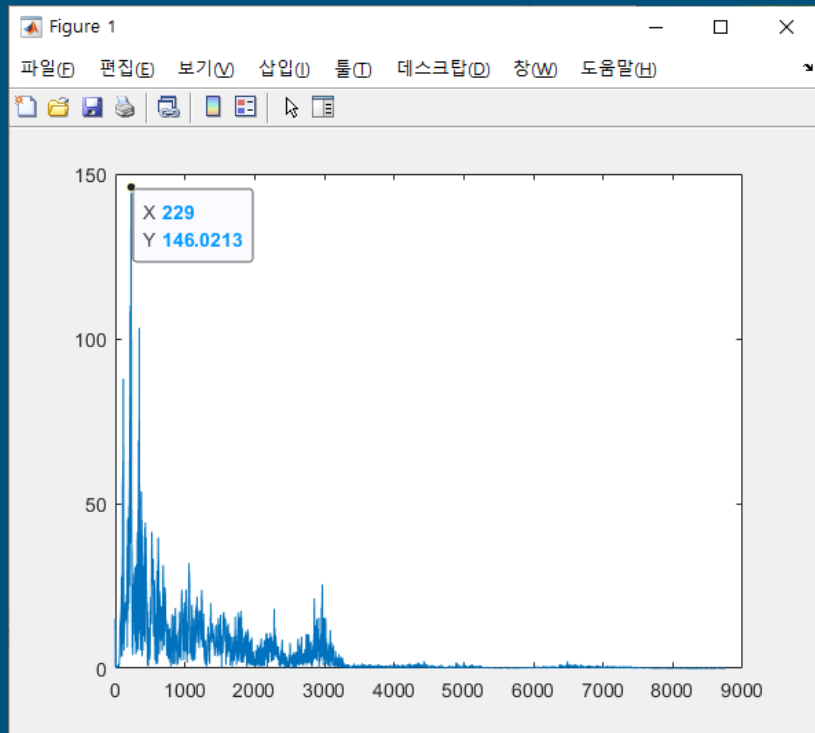
2. Phase vocoder 사용

- 분석 - 재합성 의 원리
- 음원을 레고 블럭처럼 다루는 것으로 음을 여러 개의 bin으로 나눈 후 다시 합성

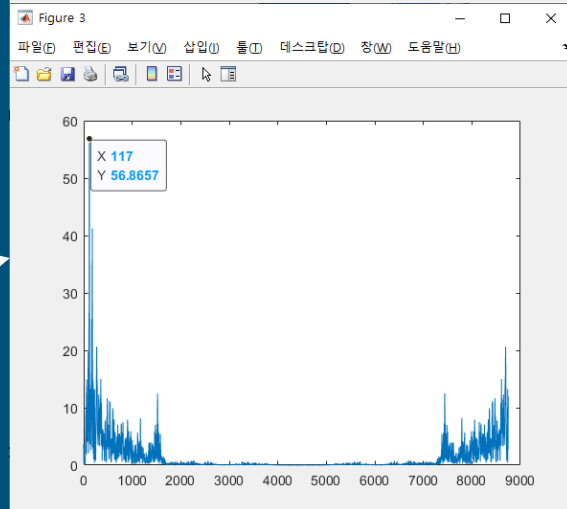
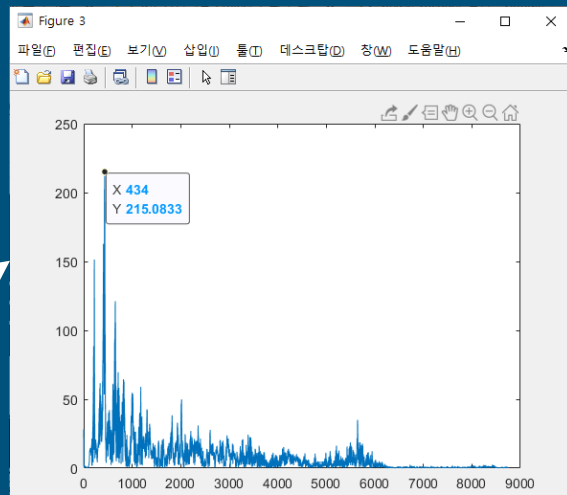
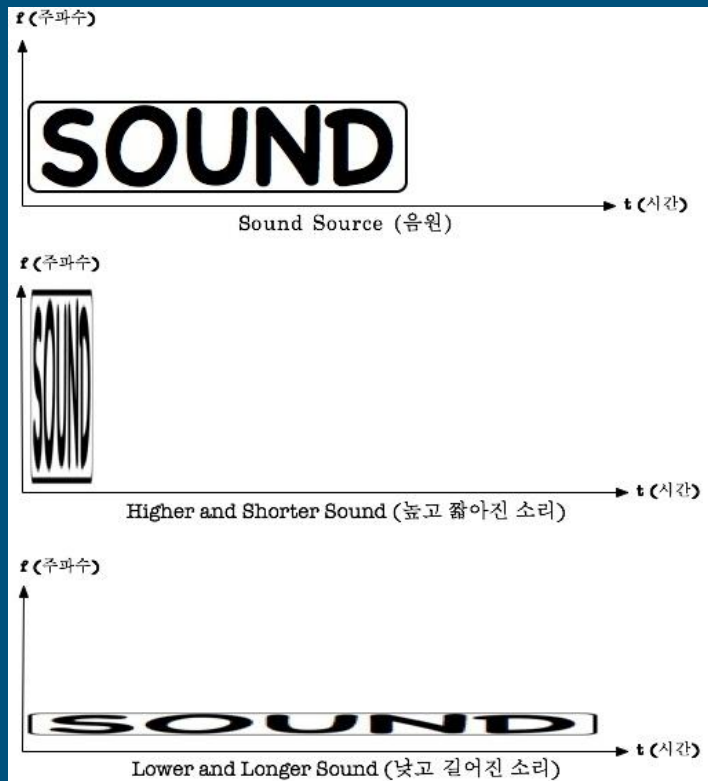
3. ISTFT(Inverse Short Time Fourier Transform)을 이용해서 원래 소리 형태로 변환

FFT결과

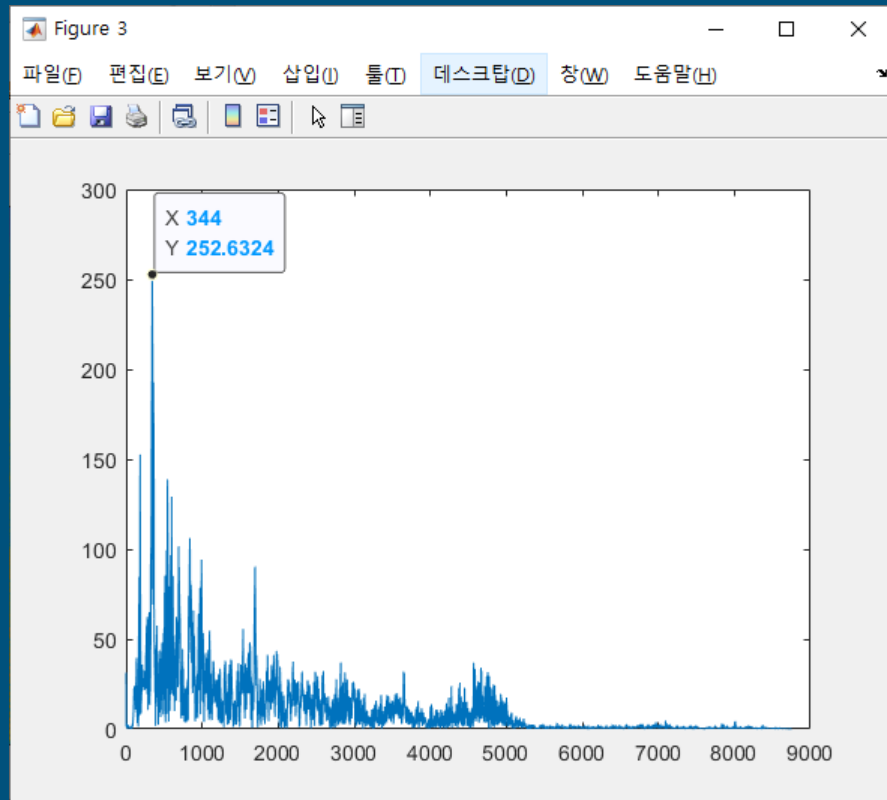
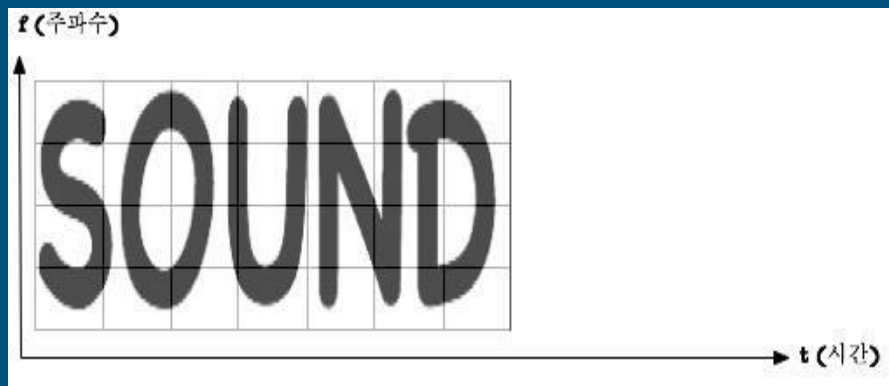
STFT 결과



주파수만 바꿨을 경우



변환된 목소리의 FFT 결과



평가

- 단순히 주파수만 바꾸는 경우 시간 도메인도 영향을 받아 음성 변조가 안 되는 것을 확인
- STFT 를 사용하여 시간 도메인도 함께 분석하여 주파수를 변조해야 하는 것을 확인
- 하지만 단순히 주파수만을 바꿨다고 여자처럼 말하는 것이 아니고 높은 목소리로 말하는 것이므로 다른 사람처럼 말하는 것은 또 다른 방법 필요