

# 多媒体技术基础 音频实验

计 24 2012011335 柯均洁

## 一、 实验 1: A to X 声音转换

### ● 实验内容

1. 将某非著名相声演员的声音转换为 3 种不同的声音

guodegang.wav

2. 将著名评书演员单田芳的声音转换为 3 种不同的声音

shantianfang.wav

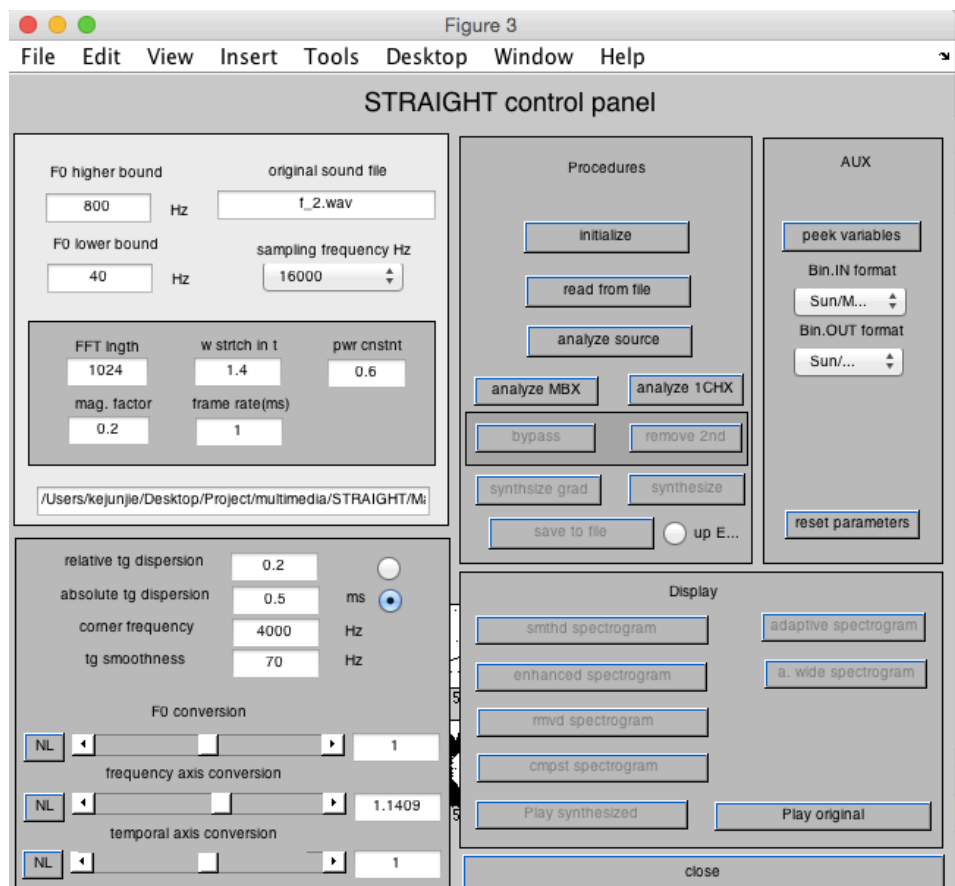
### ● 实验工具

Matlab\_Straight

### ● 实验过程

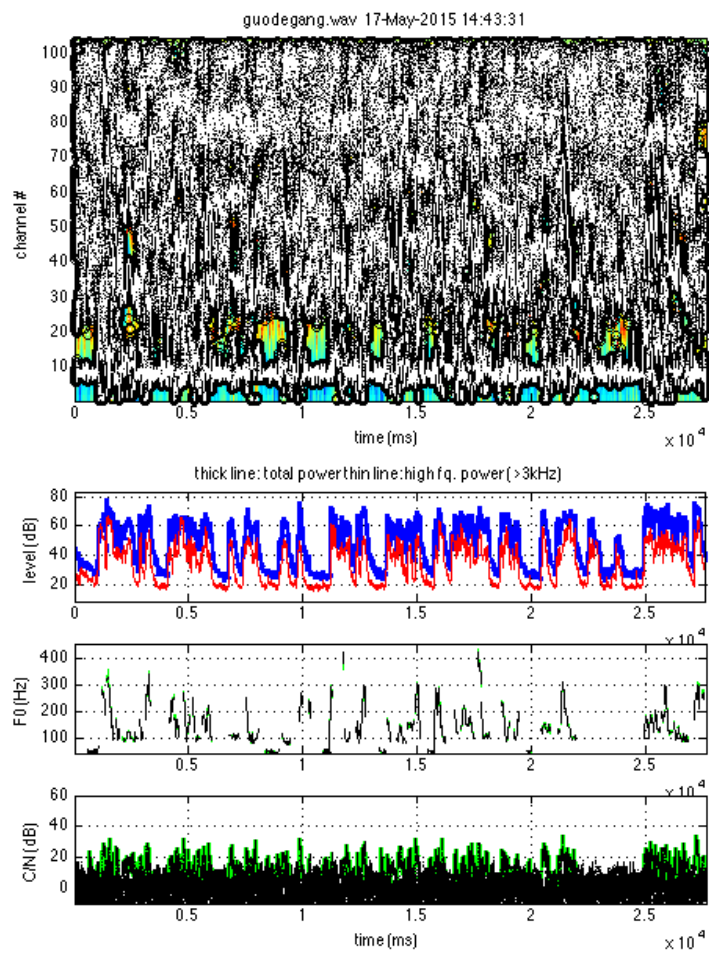
分析转换 guodegang.wav、shantianfang.wav

1. 打开 STRAIGHT，点击界面中间的 read from file

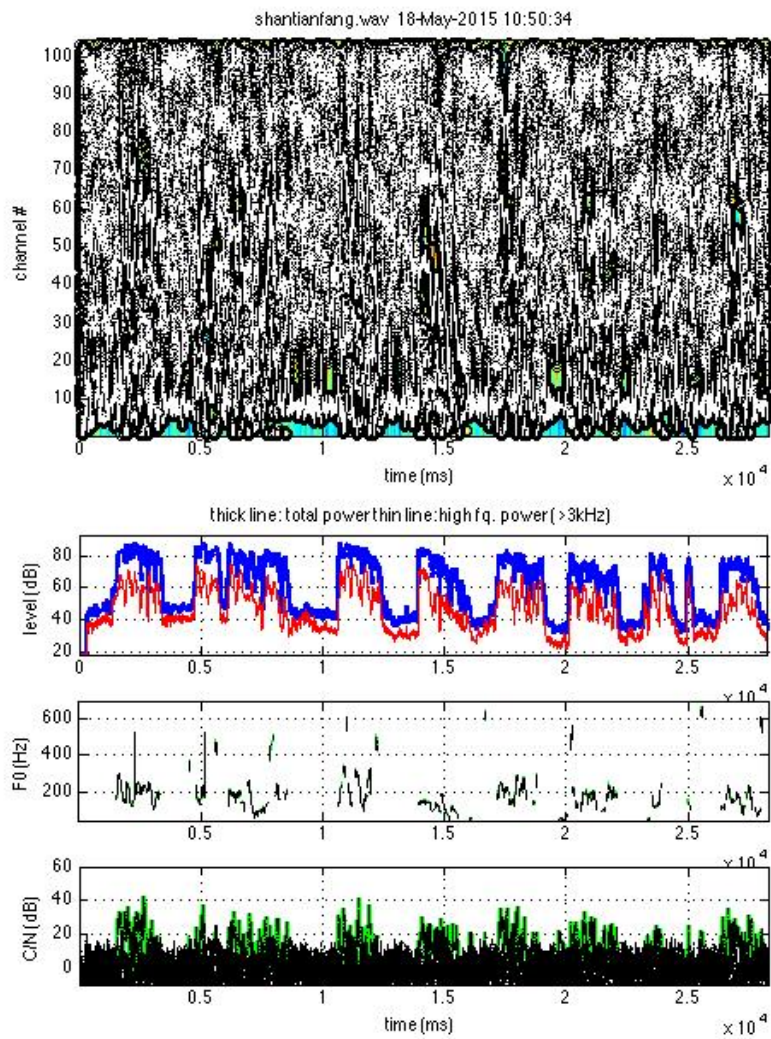


2. 点击 Analyze source，分别得到下图

➤ guodegang.wav

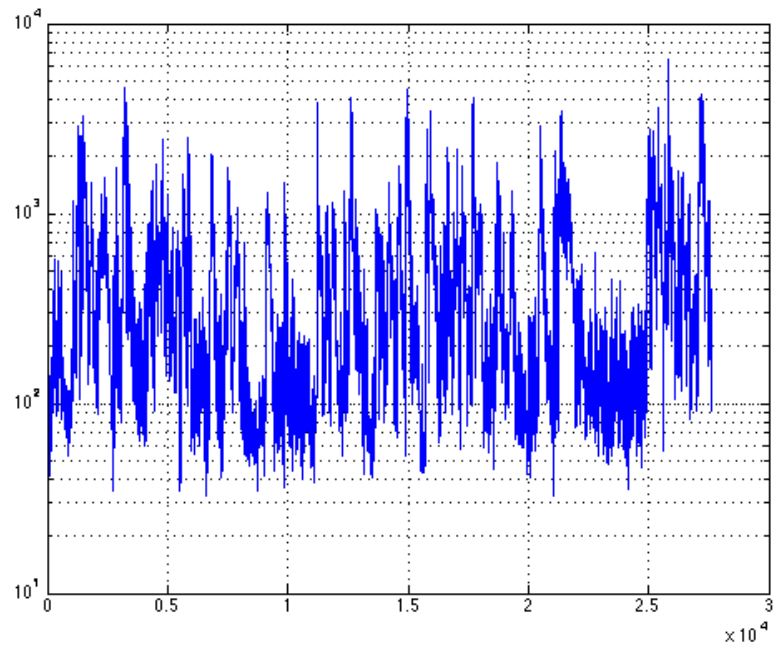


● shantianfang.wav

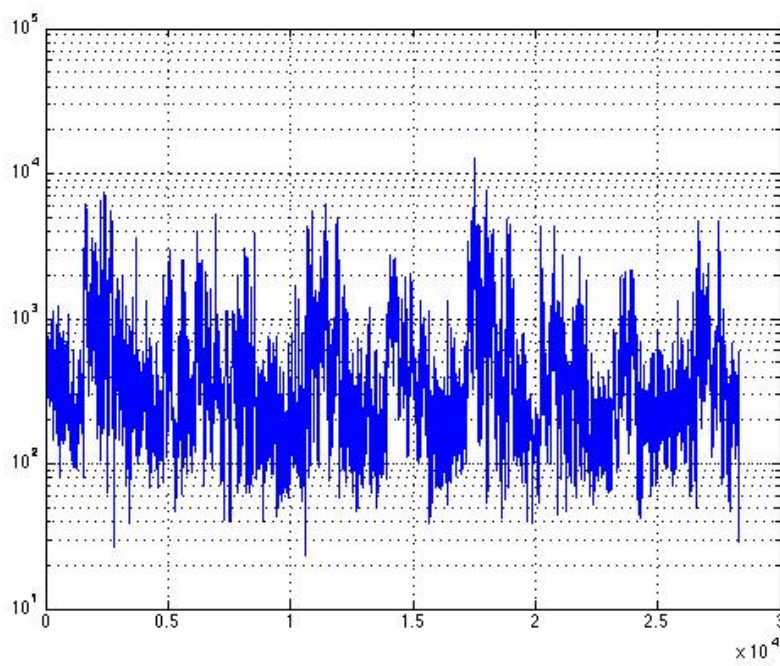


3. 点击 analyze MBX, 得到下图

➤ guodegang.wav

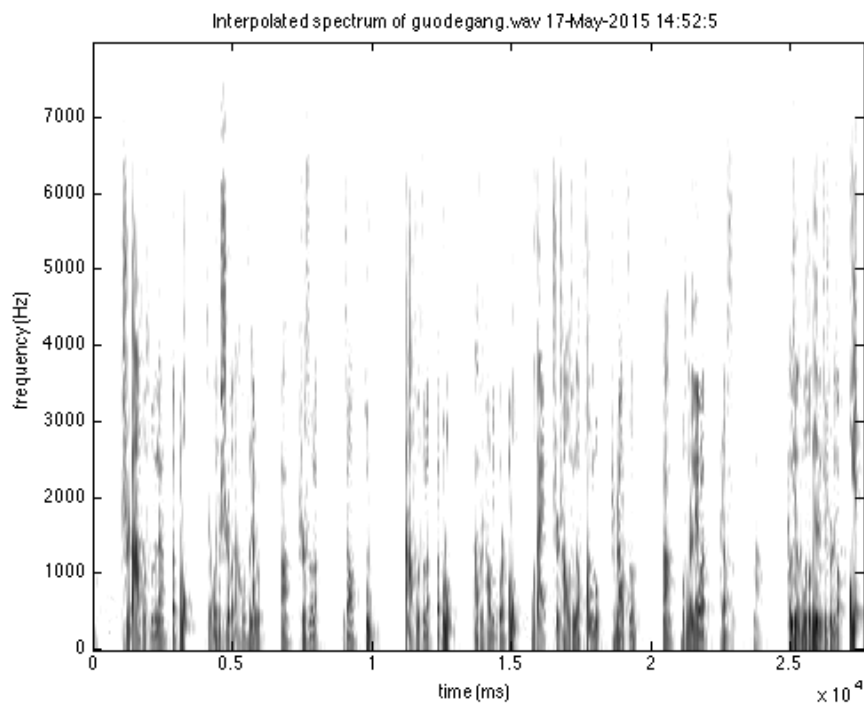


● shantianfang.wav

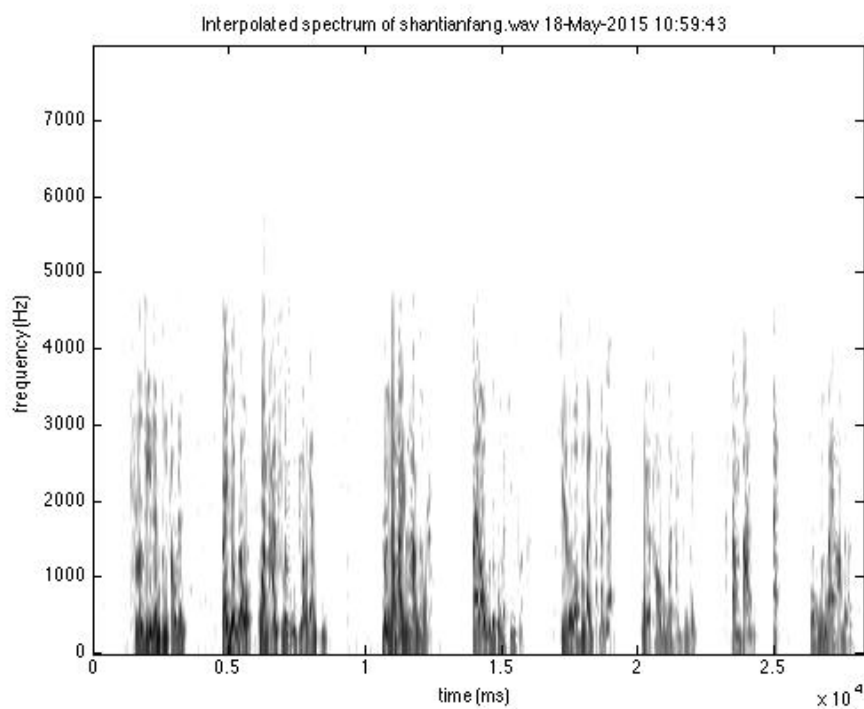


4. 点击 bypass, 接着点击 smthd spectrogram, 得到下图

➤ guodegang.wav



● shantianfang.wav



从上述图表可以看出，声音的基频分布在 100-300Hz 之间，频率大约在 50-2000 之间，声音的大小大约在 20-60db。

## 5. 声音转换

修改左下角的 F0、frequency、temporal，再点击 Synthesize

grad, 即可实现对声音的转换。分别尝试三种变换参数即可得到三种不同的转换结果

- 声音转换参数及结果

- i. 原始文件: guodegang.wav

- 1. 参数 1:

- $F0 = 2.2022$

- Frequency = 1.1996

- Temperal = 0.83176

- 变换后的声音文件为 guodegang\_1.wav

- 声音变得尖细了, 而且语速变快

- 2. 参数 2:

- $F0 = 0.6964$

- Frequency = 0.8897

- Temperal = 1.2549

- 变换后的声音文件为 guodegang\_2.wav

- 声音变得低哑了, 而且语速变慢

- 3. 参数 3:

- $F0 = 1.6927$

- Frequency = 0.96179

- Temperal = 1.2549

- 变换后的声音文件为 guodegang\_3.wav

- 声音变粗但是音调变高, 而且语速变快

- ii. 原始文件: shantianfang.wav

- 1. 参数 1:

- $F0 = 2.0217$

- Frequency = 1.2034

- Temperal = 0.84044

- 变换后的声音文件为 shantianfang\_1.wav

- 声音变得尖细了, 而且语速变快

- 2. 参数 2:

- $F0 = 0.7737$

- Frequency = 0.966

- Temperal = 1.2148

- 变换后的声音文件为 shantianfang\_2.wav

- 声音变得低哑了, 而且语速变慢

- 3. 参数 3:

- $F0 = 2.1309$

Frequency = 0.81028

Temperal = 0.84044

变换后的声音文件为 shantianfang\_3.wav

声音变粗但是音调变高，而且语速变快

## 二、 实验 2：A to B 声音转换

- 实验内容：

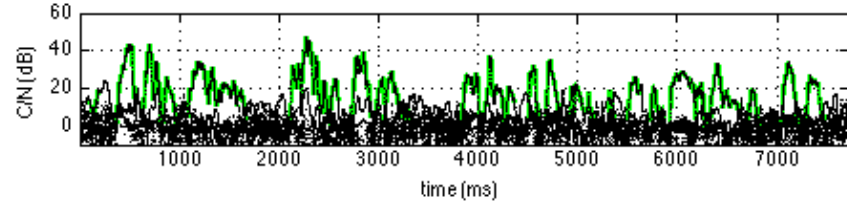
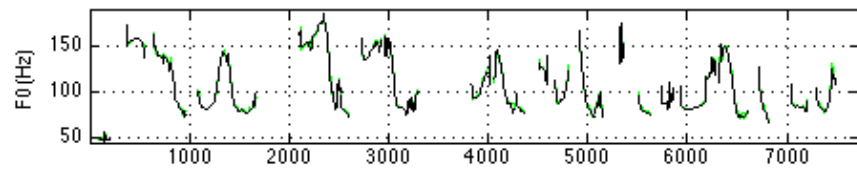
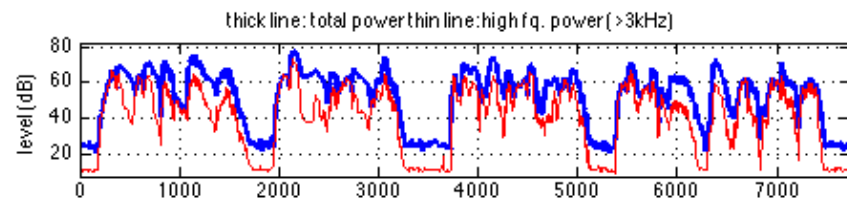
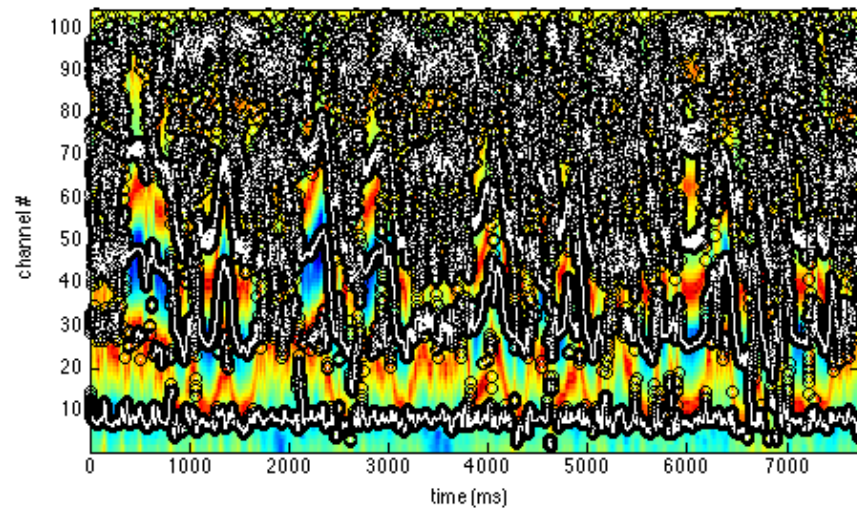
分析原始声音 A 和目标声音 B 的声学特征(F0/时长/频谱/共振峰等)差别，使用声音转换工具进行转换，修改 A 的声学特征，使其尽量接近于 B 的声音

- 实验过程：

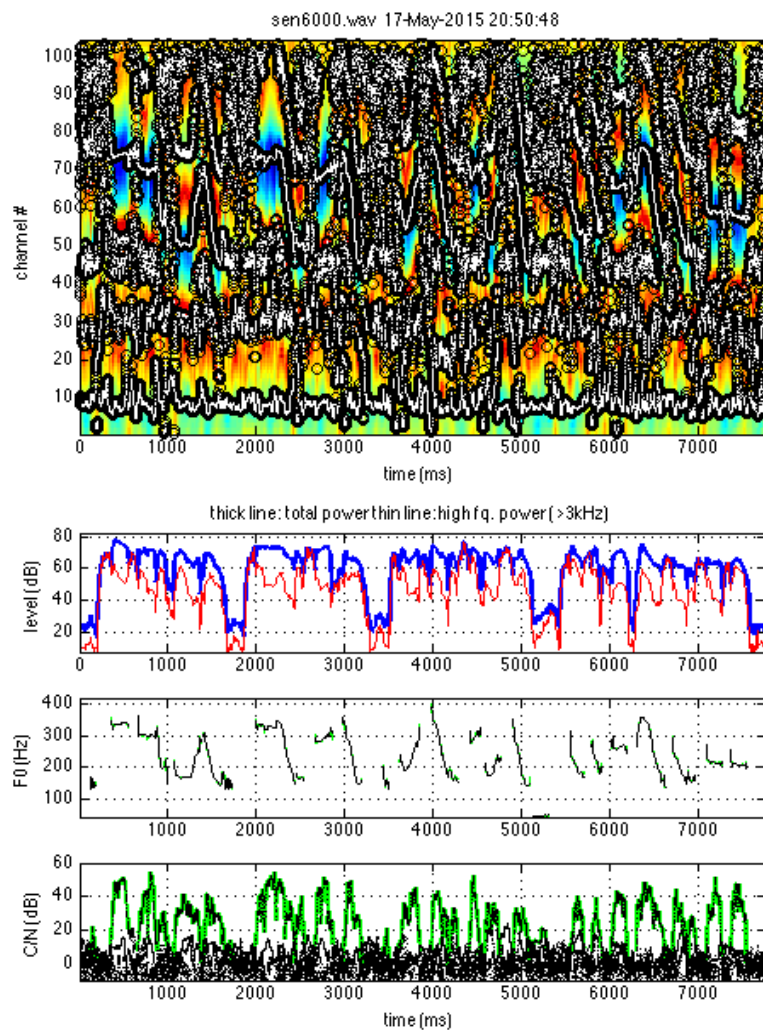
以 sen6000.wav 为例，介绍一下实验过程

1. 对原始声音 A 中的 sen6000.wav 及目标声音 B 中的 sen6000.wav 都进行 analyze source，分别得到下图：

sen6000.wav 17-May-2015 20:39:49

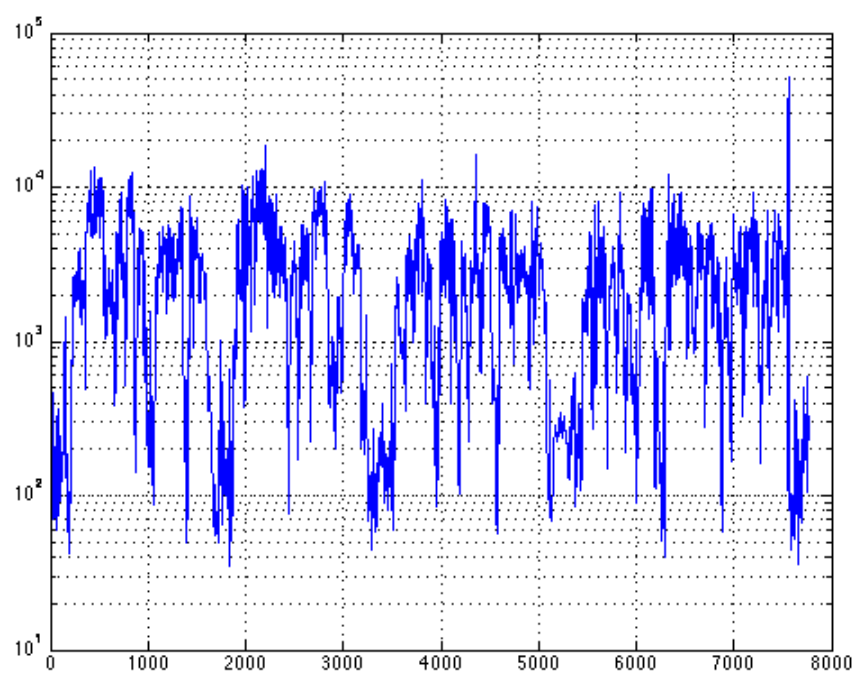
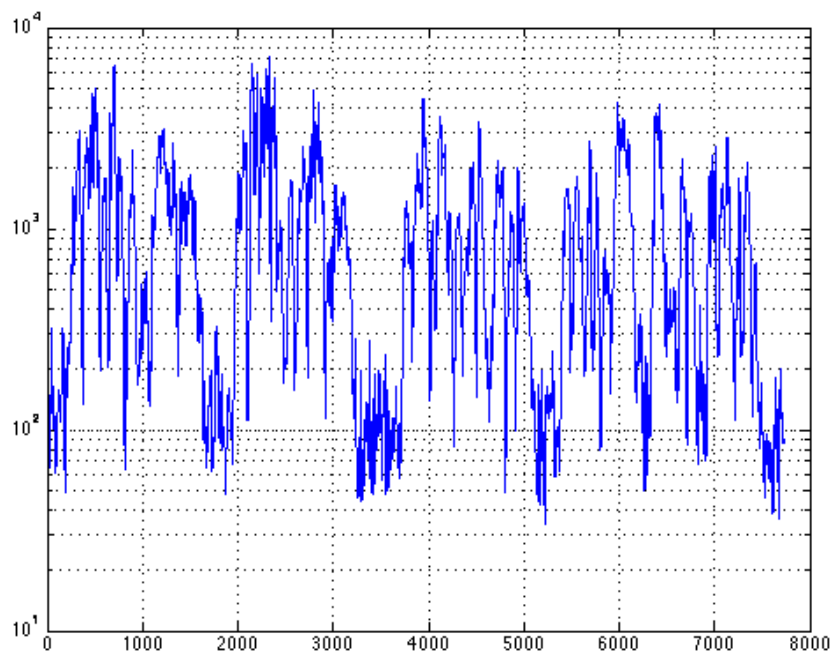




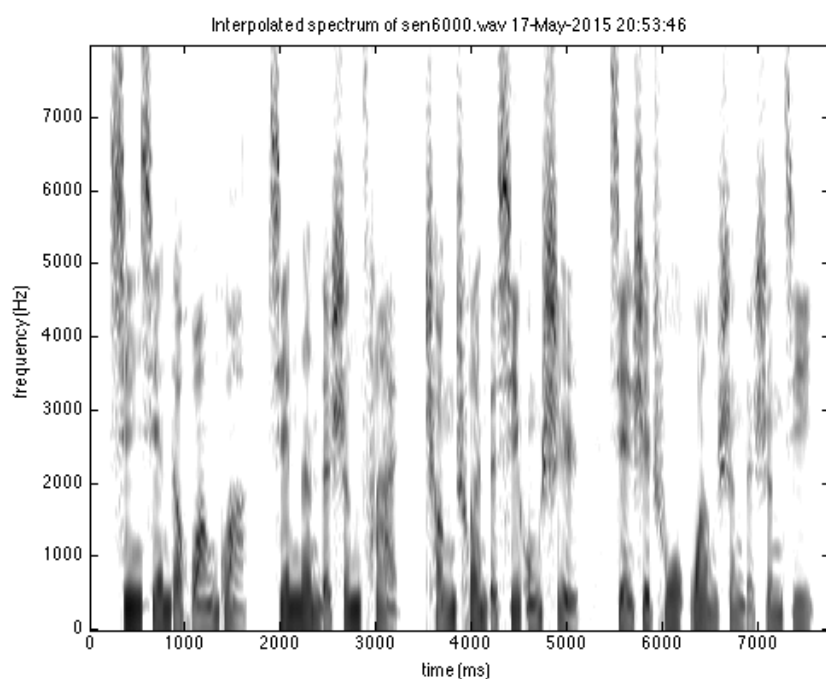
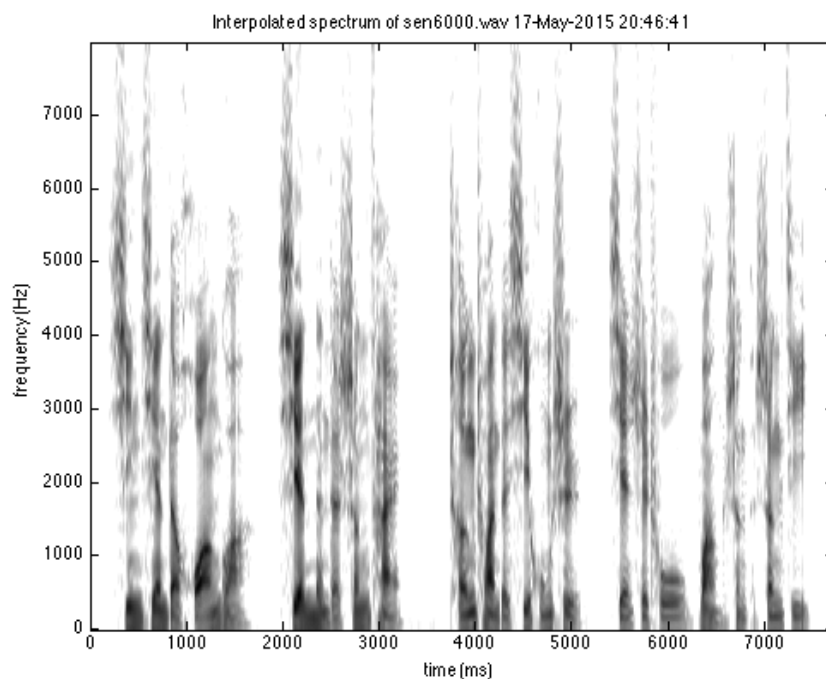


从上图可以看出，A、B 的信噪比、能量分布都比较相似，主要是在基频上的区别：A（男声）的基频在 100 到 150Hz 之间，B（女声）的基频在 200 到 300Hz 之间。B 的基频大约是 A 的两倍因此，可以试着将基频提高 2 倍左右

2. 对 A、B 都进行 analyze MBX，分别得到结果如下：



3. 对 A、B 都进行 smthd spectrogram，分别得到结果如下：



可以看出，B 的频率要比 A 的高，大约在 1.2 倍左右

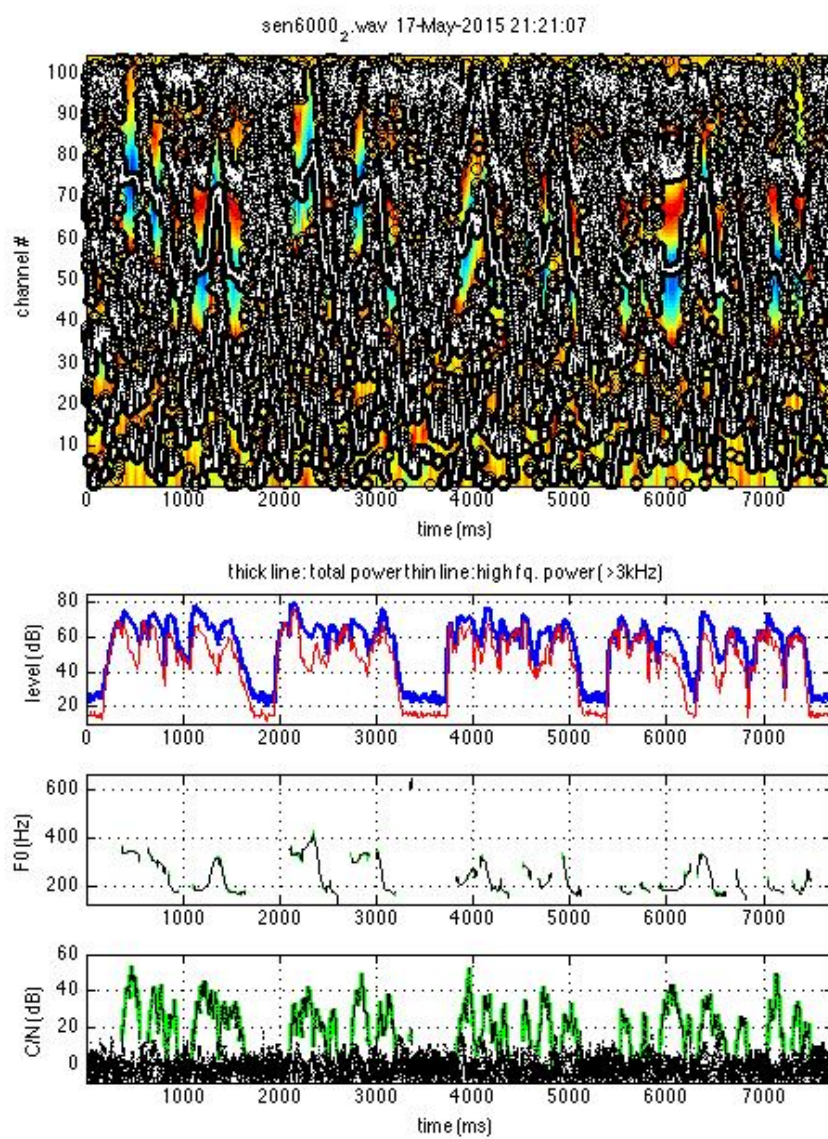
#### 4. 声音转换

通过上述分析并结合多次尝试，将基频提高为 2 倍，频率提高为 1.2 倍时可将 A 转换为非常接近 B 的声音

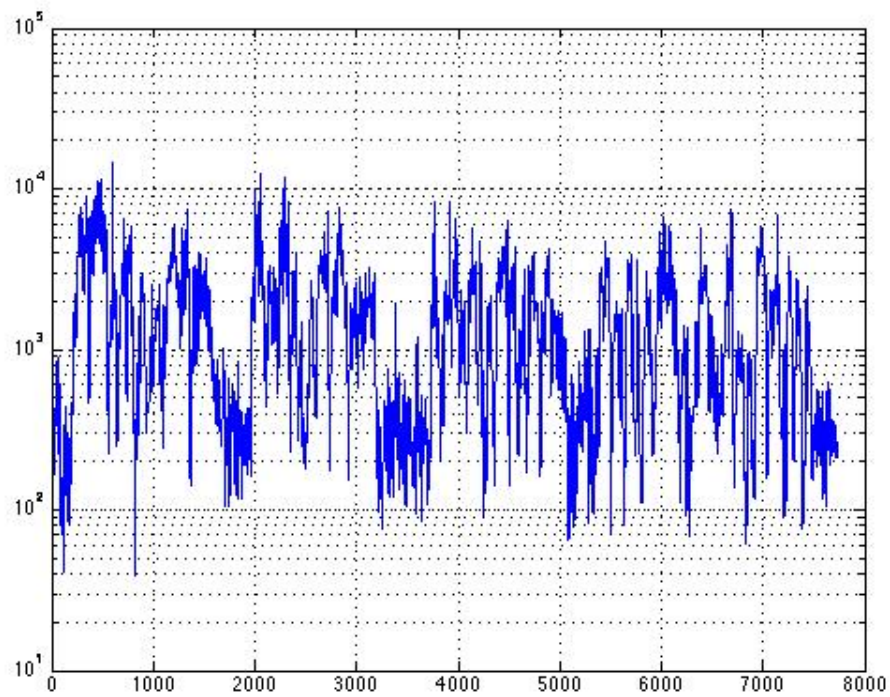
#### ● 实验结果

若从 A 转换得到的新的声音为 C，通过 STRAIGHT 对 C 进行分析：

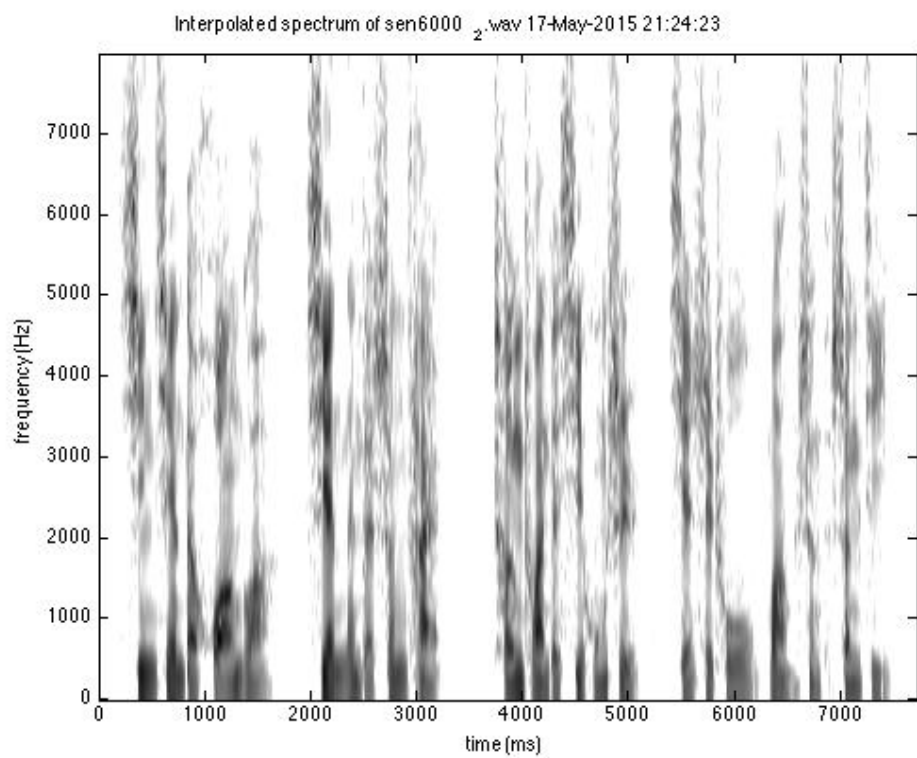
## 1. Analyze source



## 2. Analyze MBX



### 3. Smthd spectrogram



比较 C 与 B 的特征发现两者差别很小，参数基本相近

- 实验参数

1. sen6000.wav

- F0 = 2.2909
  - Frequency = 1.2734
  - Temperal = 1
- 2. sen6015.wav
  - F0 = 2.7542
  - Frequency = 1.3017
  - Temperal = 1
- 3. sen6028.wav
  - F0 = 2.5119
  - Frequency = 1.1922
  - Temperal = 1
- 4. sen6044.wav
  - F0 = 2.0893
  - Frequency = 1.1663
  - Temperal = 1
- 5. sen6147.wav
  - F0 = 2.2909
  - Frequency = 1.1922
  - Temperal = 1

### 三、 实验 3：声音转换评价

- 实验内容：

给出至少两种声音距离度量准则

对比实验 1 中的原始声音和转换后声音，利用你给出的声音距离度量准则，计算转换前后声音的距离

对实验 2，利用你给出的声音距离度量准则，给出原始声音 A 到目标声音 B 的距离以及 A 到转换后声音 A' 的距离，说明为什么 A' 比 A 听上去更接近 B 的声音

- 三种声音距离度量方法：

1. 基频均值（见 avgF0Dist.m）

使用 STRAIGHT 中的 exstraightsource 函数获得语音的基频向量，将基频向量的平均值只差作为一种距离度量。距离计算公式如下，其中

$F0(X)$  表示信号 X 的基频向量：

$$AvgF0(X) = mean(F0(X))$$

$$Dist(X,Y) = |AvgF0(X) - AvgF0(Y)|$$

➤ 对 guodegang.wav 得到如下所示结果

文件名	AvgF0	与原始声音距离
guodegang.wav	67.64	--
guodegang_1.wav	108.66	41.02
guodegang_2.wav	40.36	27.28
guodegang_3.wav	91.66	24.02

➤ 对 shantianfang.wav 得到如下所示结果

文件名	AvgF0	与原始声音距离
shantianfang.wav	72.21	--
shantianfang_1.wav	101.94	29.74
shantianfang_2.wav	48.69	23.52
shantianfang_3.wav	68.99	3.22

➤ 对实验 2，设原始声音 A，目标声音 B，转换后声音 C，得到如下所示结果

文件名	AvgF0(A)	AvgF0(B)	AvgF0(C)	Dist(A, B)	Dist(B, C)
sen6000.wav	62.06	143.01	126.33	80.94	16.68
sen6015.wav	60.66	144.52	143.74	83.85	0.78
sen6028.wav	59.05	145.00	127.75	85.95	17.26
sen6044.wav	65.57	138.76	120.50	73.19	18.26
sen6147.wav	66.20	147.68	133.46	81.48	14.22

2. 频谱的平均幅值（见 avgAmplitudeDist.m）

$$AvgAmplitude(X) = \frac{1}{n} \sum_{i=1}^n (abs(F[i]))$$

$$Dist(X, Y) = |AvgAmplitude(X) - AvgAmplitude(Y)|$$

➤ 对 guodegang.wav 得到如下所示结果

文件名	平均幅值	与原始声音距离
guodegang.wav	12.94	--
guodegang_1.wav	19.59	6.65
guodegang_2.wav	19.67	6.73
guodegang_3.wav	19.91	6.98

➤ 对 shantianfang.wav 得到如下所示结果

文件名	平均幅值	与原始声音距离
shantianfang.wav	57.71	--
shantianfang_1.wav	16.53	41.18
shantianfang_2.wav	17.14	40.57
shantianfang_3.wav	13.83	43.88

➤ 对实验 2，设原始声音 A，目标声音 B，转换后声音 C，得到如下所示结果

文件名	平均幅值 (A)	平均幅值 (B)	平均幅值 (C)	Dist(A, B)	Dist(B, C)
sen6000.wav	10.12	14.97	15.34	4.85	0.36
sen6015.wav	6.87	11.27	13.51	4.40	2.24
sen6028.wav	6.91	9.63	15.23	2.72	5.60
sen6044.wav	6.20	8.80	6.67	2.60	1.76
sen6147.wav	5.23	6.67	14.60	1.44	7.93

3. 频谱的平均频率（见 avgFrequencyDist.m）

$$AvgFrequency(X) = \frac{\sum_{i=1}^n (abs(F[i])) \times i}{\sum_{i=1}^n (abs(F[i]))} \times \frac{Fs}{n}$$

$$Dist(X, Y) = |AvgFrequency(X) - AvgFrequency(Y)|$$

➤ 对 guodegang.wav 得到如下所示结果

文件名	平均频率	与原始声音距离
guodegang.wav	1649.38	--
guodegang_1.wav	2025.00	375.62
guodegang_2.wav	1495.26	154.11
guodegang_3.wav	1670.13	20.76

➤ 对 shantianfang.wav 得到如下所示结果

文件名	平均频率	与原始声音距离
shantianfang.wav	1088.70	--



shantianfang_1.wav	1416.32	327.63
shantianfang_2.wav	1060.72	27.98
shantianfang_3.wav	1059.92	28.78

- 对实验 2，设原始声音 A，目标声音 B，转换后声音 C，得到如下所示结果

文件名	平均频率 (A)	平均频率 (B)	平均频率 (C)	Dist(A, B)	Dist(B, C)
sen6000.wav	2118.74	2854.14	2623.54	735.40	230.60
sen6015.wav	2047.15	2491.31	2560.67	444.16	293.46
sen6028.wav	2009.19	2064.00	2387.73	54.82	323.72
sen6044.wav	2170.38	2360.66	2517.70	190.28	157.04
sen6147.wav	2077.68	1919.32	2413.61	158.36	494.29

#### ● 实验结果

通过三种声音距离度量方法，分别考虑了声音信号的基频向量均值、频谱幅度均值、频谱频率均值，从基频、幅度、频率三个方面来考察原声音与转换声音的相似程度。可以看出，三种度量准则下实验 2 转换后的声音与目标声音的距离都要小于原声音与目标声音的距离

#### ● 实验总结

通过这次实验，我对音频处理的流程和方法有了更加深入的认识，将课堂上学习的知识与实践相结合，加深了对知识点的理解和掌握。感谢老师和助教的辛勤工作和热情指导！