## 基于源滤波模型短时傅立叶变换幅度谱的声音变换

**摘 要：**声音变换的一个重要评价标准是在改变声音的目标参数的时候如何保持其他参数的恒定。对声音节奏变换、声音音调变换和声音音色变换而言，目标参数分别是声音快慢速度、基音频率和声音共振峰的位置和带宽。本文提出了一种基于源滤波模型短时傅立叶变换幅度谱的声音变换方法。实验表明，利用本文提出的方法能有效地将基音参数和共振峰参数分开调整，取得了很好的声音变换效果。

**关键词**：声音变换，节奏变换，音调变换，音色变换，信号重构

**Abstract:** The key evaluation criterion in a specific voice modification is how to change target parameters while keeping other parameters constant. For time-scale modification (TSM), pitch modification (PM) and timbre modification (TM), the target parameters are respectively the tempo, pitch frequency and location and bandwidth of formants. This paper proposes a method for voice modification by short-time Fourier transform magnitude (STFTM) based on source filter model (SFM). Experiments using the proposed method show that pitch parameters and formants parameters are successfully separated. A pretty good performance is achieved.

**Keywords:** voice modification, time scale modification (TSM), pitch modification (PM), timbre modification (TM), signal estimation

1. **引言**

声音变换是一种用来改变声音特点的技术。这种技术广泛应用于娱乐产业以及用于增加声音合成数据库的多样性。例如，一个语言教学系统需要改变语音播讲的速度以使得发音更加清楚；一个语音合成系统（TTS）需要改变原始语音数据库中的声音以使得一个男生听起来像一个女声，从而在不增加语音数据库大小的情况下增加语音数据库的多样性。

声音变换一般包括四种类型：节奏变换，音调变换，音色变换，强度变换。声音节奏变换的难点在于如何只改变声音的播放速度而不改变音调和音色。声音音调变换的目标是压缩或扩展声音各次谐波间的空间距离而保持短时频谱包络以及声音节奏。声音音色变换的目标在于改变声音共振峰的位置和带宽的同时保持声音的节奏和基音频率。声音强度变换可以通过简单地把信号乘上一个强度因子来得到。研究人员提出了一系列声音变换的方法。如同步叠加法（SOLA）[1]，波形相似叠加法（WSOLA）[2]，声码器以及各种改进方法[3, 4]，峰值对齐叠加法（PAOLA）[5]。然而上述方法在改变声音基音频率的同时改变了声音共振峰的位置和带宽，反之亦然。结果使得你在改变音调的时候却改变了音色，从而导致一个男声听起来如同女声。声音的个性特点遭到了破坏。Portnoff提出了用短时傅立叶变换来进行节奏变换[6]。Griffin等人提出了利用修改短时傅立叶幅度谱来重构信号的方法处理节奏变换和音调变换[7]。在Griffin的基础上Xinglei等人又提出了实时语谱迭代转换法（RTISI）和超前实时语谱迭代转换法（RTISI-LA）[8]。这些算法在信号处理的实时性方面有了较大改善，但依然没有把声音的基音频率参数和共振峰参数区分开来。导致在进行音调变换的时候改变的原始声音载有的个性特点，在进行音色变换时声音的音调也发生了改变。

为了解决上述存在的问题，本文引入了源滤波模型，将声音信号的基音频率参数和共振峰参数区分开来。原始的声音信号被分解成声音激励信号和声道滤波器信号。然后再通过在频域修改声音激励信号的频谱，利用短时傅立叶幅度谱重构信号的方法来重构激励信号以及通过修改滤波器参数的方法来修改声音的基音频率参数和共振峰参数。最后再将修改后的声音激励信号和声道滤波器信号重新合成回新的声音信号。

1. **源滤波模型**

源滤波模型认为声音信号是由声带振动产生的激励信号经过声道滤波产生的。因此，声音信号可以被分解成激励信号和声道滤波器两部分。激励信号携带了声音的基音频率，其大小决定这音调的高低。滤波器幅度谱的峰值被称为共振峰，其位置和带宽影响着声音的音色。下图显示了源滤波模型的原理。



图1 源滤波模型

源滤波模型的思想可以通过倒谱分析或是线性预测分析来实现。本文采用线性预测分析来分解声音信号。假设代表一段离散声音序列，n=1.2.3..为序列号。可认为是由之前p个信号的加权值来预测。记的预测值为

 (1)

其中和均为离散实数序列，加权参数 (k=1, 2…p)可由莱文森-杜宾算法求解得到。p为线性预测分析的阶数。理论上，当p趋近正无穷时，预测声音信号无线接近原始声音信号。通常情况下，p取10到12。相应的，原始声音信号和其预测信号的误差定义为

 (2)

对（2）式两边取z变换得到



 (3)

于是在z域里，误差信号由原始声音信号和传递函数得到。这里的是一个全零点数字滤波器，代表了声道滤波器。通过调整零点，就可以调节声音共振峰的频率和带宽。误差信号则携带了声音的基音频率信息。

1. **基于短时傅立叶变换幅度谱的信号重构**

短时离散傅立叶变换能够将一个离散时间信号转化到频率域得到频谱信号。于是，我们可以在频率域对进行修改，然后再将修改后的频谱信号反变换回时域得到修改后的时间信号。然而频谱参数包括幅度谱和相位谱。而相位在实际操作中很不方便，所以很多时候我们需要直接利用幅度谱来重构新的信号。也就是说我们需要首先将时域信号转换到频域得到频域的幅度谱，然后根据需要对幅度谱进行修改得到，最后再利用修改后的幅度谱来重构信号。

Griffin等人提出了一种算法[7]从修改过的信号幅度谱来重构信号，使重构的信号的短时傅里叶变换幅度谱尽可能地接近原始信号的短时傅里叶变换幅度谱。定义原始信号和目标信号的相似性距离如下



 (4)

其中是原始信号的短时傅里叶变换幅度谱，为修改过的短时傅里叶变换幅度谱。

利用取代，Griffin等人给出了如下迭代方程

(5)

这里有

 (6)

其中为的短时傅里叶变换幅度谱。从数学上可以证明距离测度会随着迭代次数的增加而逐步减小[7]。一般情况下取迭代次数取4到5就能使得距离测度减小到可以接受的程度。

在本文的声音变换算法中，为了进一步减少计算负担，提高运算的实时性。我们采用了标准叠加公式（7）来取代Griffin的迭代公式(5)。

 (7)

1. **节奏变换，音调变换，音色变换**

这一节介绍如何将源滤波模型和信号的短时傅里叶变换幅度谱重构法结合起来进行各种声音变换。图2揭示了声音变换的原理，将声音节奏变换，声音音调变换，声音音色变换结合到一个统一的流程当中。

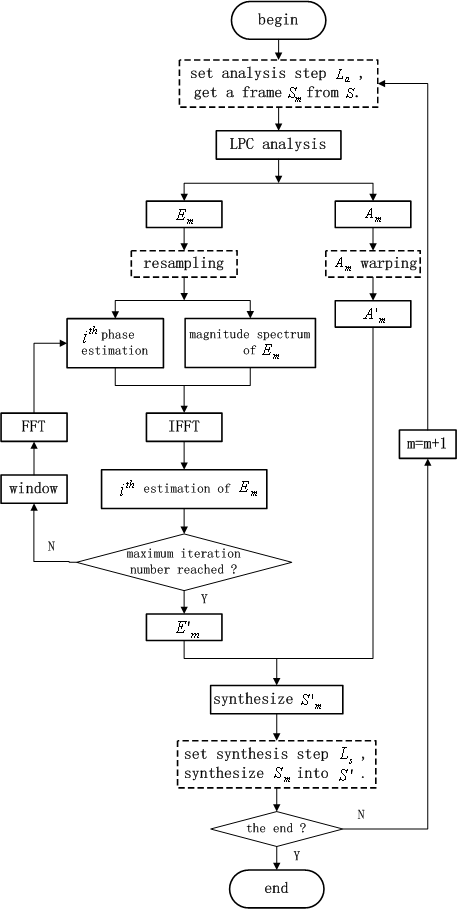


图2 声音变换流程图。实线框阐述了节奏变换，音调变换，音色变换的共同处理过程。虚线框表示这三种变换在这些处理过程当中有所不同。

实线框表示节奏变换，音调变换和音色变换都需要进行的处理步骤。虚线框表示三种变换在这些地方的处理有所不同。声音信号首先通过线性预测分析得到激励信号和声道滤波器参数。然后根据不同变换的需要利用短时傅里叶变换幅度法修改激励信号，并修改声道滤波器参数。最后将修改的激励信号和声道滤波器重新合成回时域信号就得到了变换后的声音信号。通过调整这一比率可实现声音节奏变换，通过调整重采样率，并使得可实现音调变换，通过调整的零点参数可实现音色变换。当然可以同时调整这些参数以实现特定的混合变换效果。

1. **声音节奏变换**

**参考文献**

[1] Wayman J L, Reinke R E, Wilson D L. High quality speech expansion, compression, and noise filtering using the sola method of time scale modification[C]. 1989.

[2] Verhelst W, Roelands M. An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech[C]. 1993.

[3] Laroche J, Dolson M. Improved phase vocoder time-scale modification of audio[J]. IEEE Transactions on Speech and Audio Processing. 1999, 7(3): 323-332.

[4] Dolson M. The Phase Vocoder: A Tutorial[J]. Computer Music Journal. 1986, 10(4): 14-27.

[5] Dorran D, Lawlor R, Coyle E. High quality time-scale modification of speech using a peak alignment overlap-add algorithm (PAOLA)[C]. 2003.

[6] Portnoff M. Time-scale modification of speech based on short-time Fourier analysis[J]. IEEE Transactions on Acoustics, Speech and Signal Processing. 1981, 29(3): 374-390.

[7] Griffin D, Jae L. Signal estimation from modified short-time Fourier transform[J]. IEEE Transactions on Acoustics, Speech and Signal Processing. 1984, 32(2): 236-243.

[8] Xinglei Z, Beauregard G, Wyse L. Real-Time Signal Estimation From Modified Short-Time Fourier Transform Magnitude Spectra[J]. IEEE Transactions on Audio, Speech, and Language Processing. 2007, 15(5): 1645-1653.