Journal of Image and Signal Processing 图像与信号处理, 2020, 9(4), 246-255
Published Online October 2020 in Hans. http://www.hanspub.org/journal/jisp
https://doi.org/10.12677/jisp.2020.94028

Hans汉斯

一种稳定、精准、实时的语音信号基频的检测与提取算法

章　森，曹瑞兴，邓海刚

天博电子信息科技有限公司，山东 青岛
Email: zhangsena@22tianbo.com

收稿日期：2020年9月25日；录用日期：2020年10月14日；发布日期：2020年10月21日

摘　要

针对语音基频检测与提取问题，融合了频域算法和时域算法的特点，提出了针对语音基频检测与提取的两步算法，首先基于频域算法的稳定性给出基频的一个粗估计，然后根据时域算法的精确性，再给出一个精确估计。该算法达到了稳定、精准、实时的目标。实验结果表明，该算法在汉语语音基频检测与提取方面的性能优于语音分析与处理专用软件Praat和Adobe Audition的相应功能。

关键词

语音信号处理，基频检测，特征提取，基音频率

Robust, Precise and Real-Time Algorithm for Speech Signal Pitch Detection and Extraction

Sen Zhang, Ruixing Cao, Haigang Deng

Tianbo Electronic Information Technology Co. Ltd., Qingdao Shandong
Email: zhangsena@22tianbo.com

Received: Sep. 25th, 2020; accepted: Oct. 14th, 2020; published: Oct. 21st, 2020

Abstract

A new two-step algorithm was proposed for speech pitch detection and fundamental frequency extraction. This algorithm first estimates a guess of the pitch based on the frequency analysis, and then calculates an accurate solution for the pitch based on time-domain analysis. This algorithm realized the expectation of robust, accurate and real-time. The experimental results show that the

文章引用: 章森, 曹瑞兴, 邓海刚. 一种稳定、精准、实时的语音信号基频的检测与提取算法[J]. 图像与信号处理, 2020, 9(4): 246-255. DOI: 10.12677/jisp.2020.94028

performance of this algorithm is better than that of Praat and Adobe Audition in Chinese speech pitch detection and fundamental frequency extraction.

Keywords

Speech Signal Processing, Pitch Detection, Feature Extraction, Fundamental Frequency

1. 前言

在语音信号分析与处理中，语音信号可视为一个动态非平稳随机过程，一般被分解为许多不同频率的正弦波的叠加，其中频率最低的正弦波即为基音频率(简称基频或基音，用 F0 或 f0 表示)，而其它频率较高的正弦波则为泛音。从言语的发音模型看[1]，基音周期是声带每开启和关闭一次的时长，基频则是基音周期的倒数。基音频率是语音信号的最重要的特征参数之一，在旋律辨识、声调辨识、语音合成和语音编码等研究领域起着非常关键的作用。然而，基频与语音信号本身一样，也是复杂多变的，不同的人发出的语音基频可能不同，同一个人说不同的字词时基频可能不同，同一个人在不同的时间说相同的字词时基频也可能不同。通常，基频与发音人声带的长短、薄厚、韧性、劲度和发音习惯等有关系，在很大程度上代表了个人的特征。此外，基频还随着人的性别、年龄不同而有所不同。一般来说，男性说话者的基音频率较低，而女性说话者和小孩的基音频率相对较高。可见，基频是随发音人、发音内容、发音时间等因素动态变化的。在汉语语音中，基频的变化主要表现在声调的变化，声调具有辅助辨义的作用。另外，汉语中存在着多音字现象，同一个字在不同的语境下具有不同的声调。因此准确可靠地进行基频检测对汉语语音信号的处理显得尤为重要。

基频检测与提取是语音信号处理领域的一个基础课题。由于语音波形和声带振动的频率变化范围大且十分复杂，目前还没有一种通用的方法能准确可靠地检测出任何人在任何环境下的基频。传统的基频检测和提取算法中，以时域的自相关函数(ACF)法和平均幅度差函数(AMDF)法最为经典[2] [3] [4]。ACF 方法是计算语音信号的自相关函数，通过 ACF 曲线在基音周期整数倍的位置存在较大峰值来估计基音。但是自相关函数法有时会因为噪声太多、复音、泛音等因素，导致判断错误[5]。AMDF 方法是计算语音信号的平均幅度差函数，其原理与 ACF 方法类似，该方法在语音信号的幅度或频率变化比较敏感时，基音检测精度明显下降。基于 ACF 方法改进的 YIN 算法，结合了概率模型或机器学习的方法，但在信噪比(SNR)太小时也经常出现基频检测错误。另外，常用的基频检测还有频域方法和混合法[6] [7]。频域方法的思想是首先经过快速傅里叶变换(FFT)将信号转为频谱，然后通过频谱分析得到基频。频域方法中比较典型的有泛音内积频谱法、倒频谱分析等。混合法结合了时域和频域信息，通过时域的自相关函数法和频域的频谱信息来检测音高，并从较为可能的音高中，利用动态规划找出最佳的音高作为基频，降低了单独使用时域或频域所造成的误差。另外，基于某些特殊变换(如小波变换、Hilbert-Huang 变换等)，提出了在特殊的变换域上根据对偶原理处理基频检测问题的多种方法[8]-[13]。

虽然对于基频检测与提取问题已经提出了许多解决方法或算法，但还没有一个能够同时达到稳定、精准、实时的要求。近年来，对该问题从不同的维度进行了深入的研究与探索，提出了许多新方法或思路。2015 年，Rupinder Kaur 等人[14]对该问题的最新研究进展进行了综述和分析，重点分析了 1999 年以

来提出的新方法。基于统计理论，Wohlmayr M 等人[15] [16]提出了基频检测的贝叶斯(Bayesian)概率框架，通过估计相邻帧间相关的谐波模型参数以及频谱变化实现基频检测。从模式识别的角度出发，Z. Jin 等人[17]提出了基于隐马尔科夫模型(HMM)的基频轨迹估计方法，该方法在稳定性方面有一点的优势。E Benetos 等人把基于 HMM 的基频检测方法应用于包含音乐(MIDI)成分的语音的多基频检测。Sen Zhang 等人[18]于 2010 年提出了基于时域峰值分析和波形相似度的基频检测方法，主要应用于嵌入式环境下的语音分析和处理。另外，机器学习和神经网络技术也被应用在基频检测中[16] [19]，其思想是首先使用卷积神经网络(CNN)来选取候选基音，然后根据语音信号的连续性，再用动态规划方法进行基音追踪，生成连续的基音轮廓。E.Vincent 等人[17] [20]提出了基于谐波分析检测和提取多基频的方法，主要用于从混合语音中分离出的人声和乐音。这样得到的基频结果比较稳定，但计算量较大。值得注意的是，国内外许多公司和科研机构(例如：华为、百度、中科院声学所和自动化所、清华大学、微软、谷歌等)在基频检测与提取领域申请了多项技术专利，多数优秀的专业语音处理软件中使用的基频检测与提取技术细节还没有公开。因此，关于该问题的综述分析可能存在某些遗漏。

根据发声时声带是否振动，语音信号一般分为浊音(Voiced)和清音(Unvoiced)。例如，英语中的元音，汉语中的韵母等一般都可以划分为浊音，而英语中的某些辅音和汉语中的某些声母可以划分为清音。基频的检测与提取一般针对浊音段，而清音段一般认为没有明显的基频特征。本文的工作主要针对汉语中的韵母语音段进行基频的检测与提取，内容包括四部分：第一部分前言，主要介绍基频检测与提取的典型算法及最新进展。第二部分算法设计与实现，介绍一种稳定的、精准的、实时的基频检测与提取算法，讨论了这个算法的设计思想与实现问题。第三部分实验结果，针对不同信噪比的语音数据，对比分析了基于本文算法与基于语音分析工具 Praat (荷兰阿姆斯特丹大学开发的一款多功能语音学专业软件，主要用于对数字语音信号进行分析、标注、处理及合成等实验)得出的基频结果，验证了本文算法在稳定性、精准性和实时性方面的特点。最后一部分结论，总结了本文提出的基频检测与提取算法的优缺点以及进一步改进的方法。

2. 算法设计与实现

本文为了叙述方便，首先说明几个概念和操作。

语音帧 F：一段数字化语音数据的样本点，一帧语音的时长一般是 0.2~0.35 秒，这样一帧语音大约包含 4~5 个基音周期。本文采用的语音帧长为 0.32 秒。如果采样频率是 16 kHz，对应的一帧的语音样本点数为 512 个。

语音序列 S：语音信号可以看作一维的流数据，数字化的语音信号可以表示为一个数字序列 S，即 $S[1],S[2],\cdots,S[n]$。

频谱序列 P：从语音序列 S 中选取一帧语音数据 F，对 F 进行快速傅里叶变换，即得到频谱序列 P。

峰值点：对序列 X 及其下标 j，如果满足：

$$X[j-1]<X[j] \text{ 且 } X[j+1]<X[j]$$

则 X[j]称为一个峰值点，j 称为这个峰值点的下标，X[j]称为峰值。

值得注意的是，一帧内的峰值点可能不是最大值点，最大值点也可能不是峰值点。

据不完全统计，目前为止语音的基频检测与提取算法至少存在几十种。但是如何评估比较这些算法还没有一个客观公正的标准。根据大量的应用实例的需求，对一个基频检测与提取算法的评价主要包括如下几个方面：

稳定性：无论在高信噪比还是低信噪比的情况下，算法都能检测和提取到语音信号的基频。

精准性：检测和提取到的语音信号的基频比较精准。

实时性：时延小，算法能够几乎实时进行处理。

复杂性：算法容易在不同的计算平台上实现，特别是在计算能力较弱的移动平台和嵌入式系统上。另外，算法中应该尽量少的包含各种变换和逆变换。

实际上，基频检测与提取算法既可以在频域上实现，也可以在时域上实现。频域上的算法一般稳定性较好，但精准性较差(一般误差在 10%左右)；时域算法与频域算法正好相反。融合频域算法和时域算法的特点，本文提出了针对语音基频检测与提取的两步算法(two-step fundamental pitch, TFP)，其主要思想是基于频域算法的稳定性，首先给出基频的一个粗估计，然后根据时域算法的精确性，再给出一个精确估计。实验结果表明，该算法在汉语语音基频检测与提取方面的性能达到了稳定、精准、实时的目标。

经过大量语音学的研究和实验，人类基频的范围约在 70~350 Hz，男声的基音频率一般在 100~200 Hz，女声则在 200~350 Hz。男声的平均基频为 160 Hz，误差范围(−24, 24)，女声的平均基频为 297 Hz，误差范围为(−36, 36)。另外，年龄也会影响基频范围。老年人的基频较低，儿童的基频较高。本文的算法对基频的检测范围为 93~375 Hz。

如果在语音序列 S 中选取了语音帧 F，那么本文提出的语音基频检测与提取算法如下，而且算法的每一步都是易于实现的：

a) 对语音帧 F 做快速傅里叶变换，即得到频谱序列 P = FFT(F)。如果采样频率是 16 kHz，FFT 的长度是 512，那么 FFT 的频率分辨率是 31.25 Hz。

b) 在频谱序列 P 上提取两个峰值点 PV1，PV2，一个最大值点 PK，其中 PV1 是 93~218 Hz 之间的峰值点，PV2 是 218~375 Hz 之间的峰值点，PK 是 93~375 Hz 之间的最大值点。注意，PV1，PV2 可能不存在，但 PK 一定存在。

c) 基频 F0 粗估计。如果 PV1 存在，则用 PV1 估计 F0；否则，如果 PV2 存在，则用 PV2 估计 F0；如果 PV1，PV2 都不存在，则用 PK 估计 F0，得到基频 F0 的粗估计 CF0。

d) 粗估计 CF0 的调整。针对粗估计 CF0 是基于峰值点 PV1 给出的，分为两种情况进行处理：第一种是 PV2 与 PK 重合，且 PV1 不是 PV2 的半频，则用 PV2 估计 F0。第二种是如果 PV2 存在，且 PV2 的峰值比 PV1 的峰值大，则用 PV2 估计 F0。

e) 对于基频的粗估计 CF0，在时域进一步精准估计。首先建立频率 Hz 与时域的语音序列 S 的下标之间的对应关系，这种关系是非线性的，为易于实现，用分段线性函数进行了简化，共分为三段，即 100~200 Hz，200~300 Hz，300~400 Hz，经过计算，可以得到如下分段函数关系式：

$$Y = -0.80X + 240 \tag{2.1}$$

$$Y = -0.27X + 134 \tag{2.2}$$

$$Y = -0.13X + 92 \tag{2.3}$$

其中，输入 X 为基频的粗估计 CF0，输出 Y 为 CF0 在语音序列 S 中对应的下标。如果 CF0 位于区间 100~200 Hz，则用函数式(2.1)；如果 CF0 位于区间 200~300 Hz，则用函数式(2.2)；如果 CF0 位于区间 300~400 Hz，则用函数式(2.3)。

f) 根据基频的粗估计 CF0 以及上述分段函数关系式，得到 CF0 在语音序列 S 中对应的下标 SI.在语音序列 S 中的下标 SI 附近搜索峰值点 SV1，且在下标 2 * SI 附近搜索峰值点 SV2。如果 SV1 和 SV2 都存在，则根据 SV1 和 SV2 的下标计算出 CF0 的精确估计 F0。

g) 如果粗估计 CF0 与精确估计 F0 相差 15%以上，直接以粗估计 CF0 代替精确估计 F0。

h) 返回精确估计 F0 作为本帧语音的基频。

对上述算法的分析可以看出，算法的时延是一帧语音，即 0.32 秒，能够满足实时性的需求。算法中避免了各种门限阈值的设定，泛化能力强，稳定性好。通过频域的粗估计和时域的精确估计，得到的基频估值比较精确。另外，算法用到的计算主要包括 FFT 和各种峰值点的搜索，容易在不同的计算平台上实现。事实上，我们是在语音分析与处理工具 OpenVoice 中基于 Java 语言实现的该算法。需要说明的是，本算法直接对原始语音进行处理，未做高频滤波。这是考虑到滤波操作使得原始语音波形的峰值点可能发生变化，在时域中提取基频时可能产生偏差。

3. 实验结果

为了评估本文提出的基频检测与提取算法的性能，我们在大量的不同语音数据上进行了实验，这些实验语音数据包括各种信噪比的男声、女声，原始语音是 16 kHz 采样、16 bit 量化，基频检测范围为 93~375 Hz。不同的信噪比语音是通过将原始语音归一化后加入不同噪声而产生的。

下面的实例是一段大约 2 秒的汉语语音(女声)，其中包含 5 个汉字。对这段语音 S 通过添加噪声生成三个不同信噪比的语音：S_0 dB，S_10 dB，S_20 dB。对这些实例分别用 Praat，Adobe Audition 及本文的 TFP 算法进行基频检测。从下图 1~12 可以比较直观的看出三种算法在不同信噪比下检测与提取基频的性能。图 1~4 是用 Praat 对语音段 S，S_0 dB，S_10 dB，S_20 dB 进行基频检测与提取的结果，其中语谱图区域中的蓝色曲线(多段)表示 Praat 计算出的基频曲线。

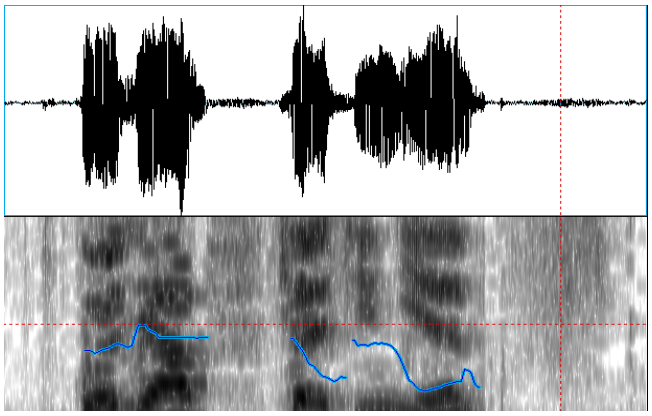

Figure 1. The Waveform and the pitch (blue curve, Praat)
图 1. 原始语音波形(上)及基频曲线(下，蓝色)

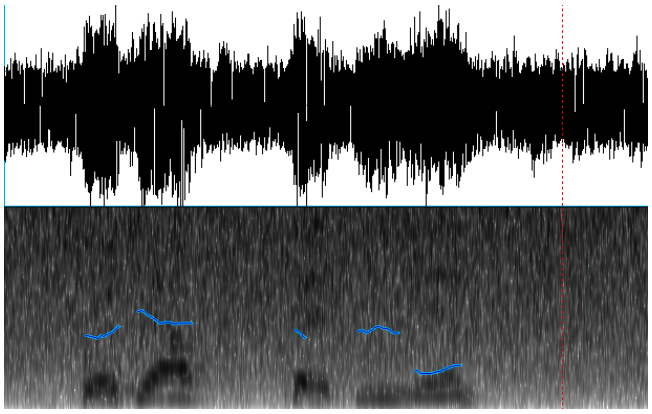

Figure 2. The Waveform of SNR = 0 db and the pitch (blue curve)
图 2. 信噪比 0 dB 语音波形(上)及基频曲线(下，蓝色)

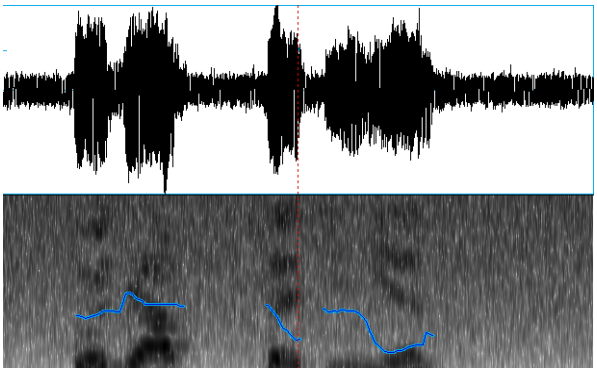

Figure 3. The Waveform of SNR = 10 db and the pitch (blue curve)
图 3. 信噪比 10 dB 语音波形(上)及基频曲线(下，蓝色)

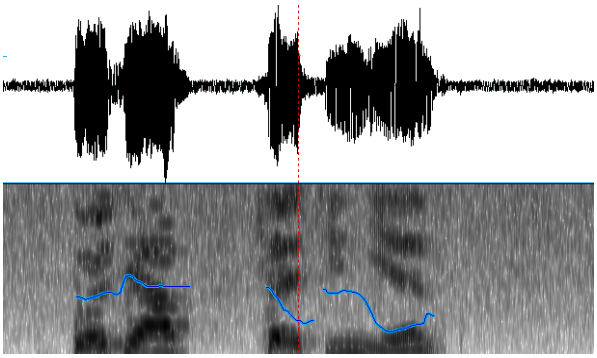

Figure 4. The Waveform of SNR = 20 db and the pitch (blue curve)
图 4. 信噪比 20 dB 语音波形(上)及基频曲线(下，蓝色)

从图 1 可以大致看出，Praat 对原始语音 S 做的基频分析基本正确，仅在基频曲线的端点附近误差较大。当信噪比较大时(图 3，图 4)，Praat 基频分析的性能下降不多，在信噪比等于 10 dB 时(图 3)端点的基频数据存在部分丢失问题。但在信噪比较小时(图 2 SNR = 0 dB)时，基频数据丢失严重。

下图 5~8 是用语音分析与处理专用软件 Adobe Audition 对语音段 S，S_0 dB，S_10 dB，S_20 dB 进行基频检测与提取的结果，其中语谱图显示区域中的蓝色曲线表示 Adobe Audition 计算出的基频曲线。从图 5~8 可以看出，总体上基频数据丢失问题严重(丢失 50%以上)，误差很大。因此，在基频的检测与提取性能方面，Adobe Audition 比 Praat 差。

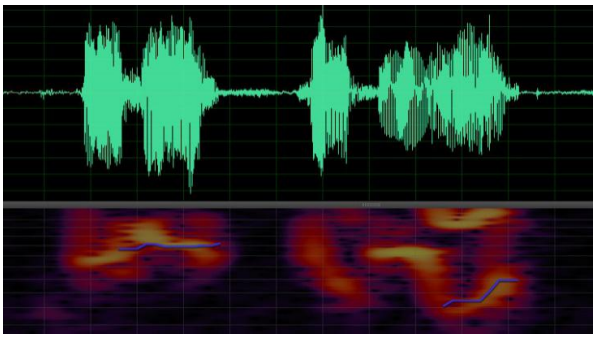

Figure 5. The Waveform and the pitch (blue curve, Audition)
图 5. 原始语音波形(上)及基频曲线(下，蓝色)

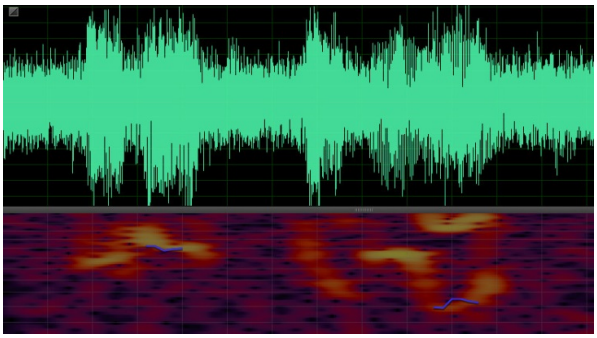

Figure 6. The Waveform of SNR = 0 db and the pitch (blue curve)
图 6. 信噪比 0 dB 语音波形(上)及基频曲线(下，蓝色)

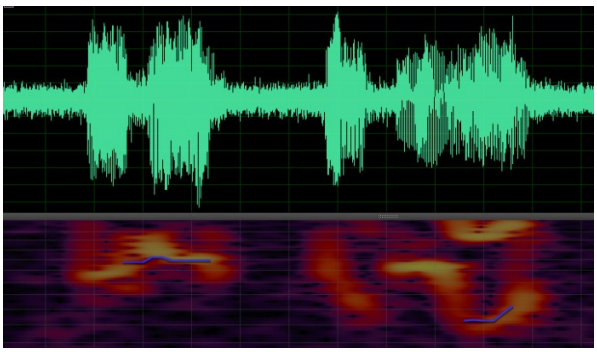

Figure 7. The Waveform of SNR = 10 db and the pitch (blue curve)
图 7. 信噪比 10 dB 语音波形(上)及基频曲线(下，蓝色)

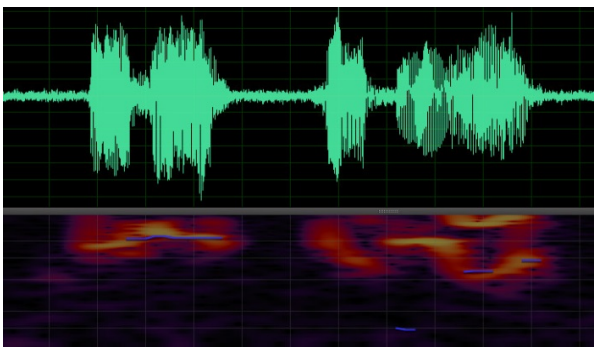

Figure 8. The Waveform of SNR = 20 db and the pitch (blue curve)
图 8. 信噪比 20 dB 语音波形(上)及基频曲线(下，蓝色)

下图 9~12 是用本文的 TFP 算法对语音段 S，S_0 dB，S_10 dB，S_20 dB 进行基频检测与提取的结果，其中黑色曲线中加灰色框的部分表示计算出的基频曲线。从图中可以看出，总体上基频曲线与 Praat 得到的几乎一致。在信噪比较小时(图 10 SNR = 0 dB)时，TFP 算法得到的基频数据除个别点外，仍然非常接近真实值。

为了更准确的比较三种算法在基频检测与提取方面的性能，我们对上述语音段 S 进行了人工分析与基频标注。实际上，这段语音有 5 个汉字，但最后两个汉字连读，所以有 4 段基频曲线，按照从左到右的顺序可以分别标记为 S1，S2，S3，S4。这四段基频曲线的数据通过四种方法(人工分析，Praat，Adobe

Audition，TFP)得到的结果如下：

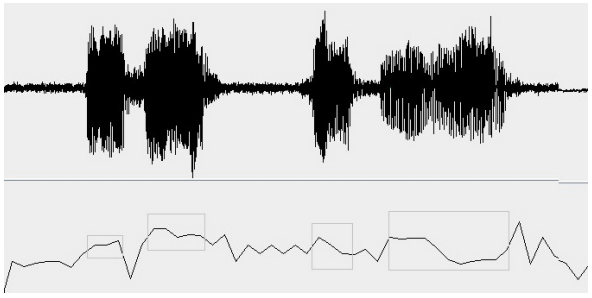

Figure 9. The Waveform and the pitch (black curve, TFP)
图 9. 原始语音波形(上)及基频曲线(下)

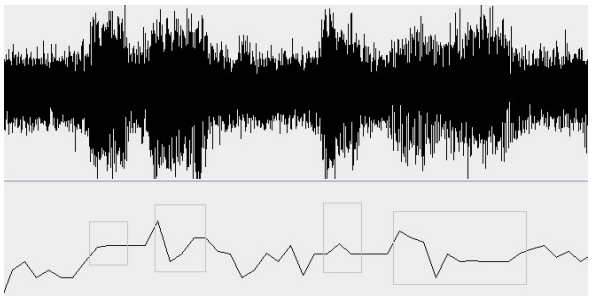

Figure 10. The Waveform of SNR = 0 db and the pitch (black curve)
图 10. 信噪比 0 dB 语音波形(上)及基频曲线(下)

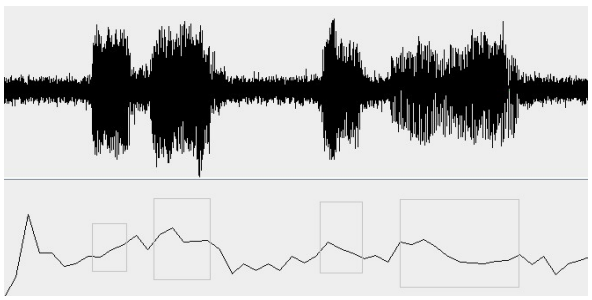

Figure 11. The Waveform of SNR = 10 db and the pitch (black curve)
图 11. 信噪比 10 dB 语音波形(上)及基频曲线(下)

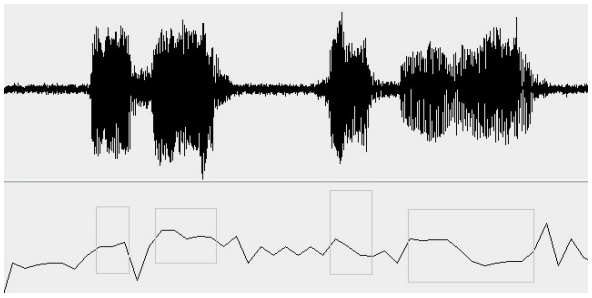

Figure 12. The Waveform of SNR = 20 db and the pitch (black curve)
图 12. 信噪比 20 dB 语音波形(上)及基频曲线(下)

人工分析与基频标注结果(Hz)：

S1：242，224，228，232，242

S2：280，280，262，266，258，250

S3：250，212，188，182

S4：250，242，232，226，180，166，150，156，160，164

Praat 基频标注的结果(Hz)：

S1：231，226，230，235，243

S2：278，282，270，262，256，256

S3：255，201，184，178

S4：252，241，239，222，175，160，151，150，155，165

Adobe Audition 基频标注的结果(Hz)：

S1：x，x，x，x，x (x 表示数据丢失，不可用)

S2：264，266，258，256，260，264

S3：x，x，x，x

S4：x，x，x，x，162，168，168，168，182，198

TFP 基频标注的结果(Hz)：

S1：242，219，229，230，242

S2：281，281，256，258，262，250

S3：250，219，188，186

S4：250，242，246，239，188，165，148，156，163，163

基于上述数据，可以计算出 Praat 和 TFP 在四个基频段 S1，S2，S3，S4 上相对于人工标注的平均误差(Adobe Audition 基频标注的结果较差，忽略之)。见下表 1。表 1 中的数据表示两个算法 Praat 和 TFP 在每个基频段上计算基频数据的平均误差，单位是 Hz。可以看出，TFP 的结果优于 Praat，特别是在强噪声的情况下，TFP 更加稳定。我们所做的大量实验也表明，TFP 是一个稳定、精准、易于实现的语音基频检测与提取算法。

Table 1. The pitch errors (Hz) by Praat and TFP
表 1. Praat 及 TFP 基频标注的误差表(Hz)

	S1	S2	S3	S4
Praat	3.8	4.0	6.0	5.1
TFP	1.6	3.3	2.8	4.2

4. 结论

本文提出的语音基频检测与提取算法 TFP 可用于各类语音信号处理场景，其性能优于 Praat。该算法还存在一些需要继续研究和改进的地方，例如该算法的基频检测范围为 90~375 Hz，对儿童、老人和经过声乐训练的人的语音基频的检测可能存在较大偏差。另外，我们的算法验证主要基于大量的汉语语音数据，对其他语言的验证不足。本文提出的汉语语音基频检测与提取算法 TFP 已经与 2020 年 5 月 12 日获得了国家发明专利，专利号：ZL201910670840.1。

参考文献

[1] 张金光. 语言发音模型研究综述[J]. 计算机工程与应用, 2018, 54(12): 27-34.

[2] Zhao, H. and Gan, W.J. (2013) A New Pitch Estimation Method Based on AMDF. *Journal of Multimedia*, **8**, 618-625. https://doi.org/10.4304/jmm.8.5.618-625

[3] Lin, Q.G. and Shao, Y.W. (2018) A Novel Normalization Method for Autocorrelation Function for Pitch Detection and for Speech Activity Detection. *Interspeech*, Hyderabad, 2-6 September 2018, 2097-2101. https://doi.org/10.21437/Interspeech.2018-45

[4] Kim, J.W., Salamon, J., Li, P., *et al*. (2018) CREPE: A Convolutional Representation for Pitch Estimation. 2018 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*), Calgary, 15-20 April 2018, 161-165. https://doi.org/10.1109/ICASSP.2018.8461329

[5] 陈萧, 徐波. 改进的用于口语处理的基频提取算法[J]. 清华大学学报(自然科学版), 2017, 57(1): 95-99.

[6] Kasi, K. and Zahorian, S.A. (2002) Yet Another Algorithm for Pitch Tracking. *IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing* (*ICASSP*), Orlando, Vol. 1, 361-364. https://doi.org/10.1109/ICASSP.2002.5743729

[7] Gonzalez, S. and Brookes, M. (2014) PEFAC-A Pitch Estimation Algorithm Robust to High Levels of Noise. *IEEE*/*ACM Transactions on Audio*, *Speech*, *and Language Processing*, **22**, 518-530. https://doi.org/10.1109/TASLP.2013.2295918

[8] Hajimolahoseini, H., Amirfattahi, R., Gazor, S., *et al*. (2016) Robust Estimation and Tracking of Pitch Period Using an Efficient Bayesian Filter. *IEEE*/*ACM Transactions on Audio*, *Speech*, *and Language Processing*, **24**, 1219-1229. https://doi.org/10.1109/TASLP.2016.2551041

[9] Huang, F. and Lee, T. (2013) Pitch Estimation in Noisy Speech Using Accumulated Peak Spectrum and Sparse Estimation Technique. *IEEE Transactions on Audio*, *Speech*, *and Language Processing*, **21**, 99-109. https://doi.org/10.1109/TASL.2012.2215589

[10] Zhang, X.L., Zhang, H., Nie, S., *et al*. (2016) A Pairwise Algorithm Using Deep Stacking Network for Speech Separation and Pitch Estimation. *IEEE*/*ACM Transactions on Audio*, *Speech*, *and Language Processing*, **24**, 1066-1078. https://doi.org/10.1109/TASLP.2016.2540805

[11] Miller, N. (2003) Pitch Detection by Data Reduction. *IEEE Transactions on Acoustics Speech & Signal Processing*, **23**, 72-79. https://doi.org/10.1109/TASSP.1975.1162642

[12] Du, S.C., Sugiura, Y. and Shimamura, T. (2018) Combining Zero Replacement Speech Enhancement with Lag Window Method for Pitch Detection. 2018 *IEEE* 3*rd International Conference on Communication and Information Systems* (*ICCIS*), Singapore, 28-30 December 2018, 53-57. https://doi.org/10.1109/ICOMIS.2018.8645016

[13] Bahja, F., Martino, J.D., Elhaj, E.I., *et al*. (2016) A Corroborative Study on Improving Pitch Determination by Time-Frequency Cepstrum Decomposition Using Wavelets. *Springerplus*, **5**, 564. https://doi.org/10.1186/s40064-016-2162-0

[14] Kaur, R. and Kumar, N. (2015) Review on Multi Pitch Detection in Speech. *International Journal of Scientific Research in Computer Science and Engineering*, **3**, 6-10.

[15] Chu, W. and Alwan, A. (2012) SAFE: A Statistical Approach to F0 Estimation under Clean and Noisy Conditions. *IEEE Transactions on Audio*, *Speech*, *and Language Processing*, **20**, 933-944. https://doi.org/10.1109/TASL.2011.2168518

[16] Nielsen, J.K., Christensen, M.G., Jensen, S.H., *et al*. (2013) Default Bayesian Estimation of the Fundamental Frequency. *IEEE Transactions on Audio*, *Speech*, *and Language Processing*, **21**, 598-610. https://doi.org/10.1109/TASL.2012.2229979

[17] Jin, Z.Z. and Wang, D.L. (2011) HMM-Based Multipitch Tracking for Noisy and Reverberant Speech. *IEEE Transactions on Audio Speech and Language Processing*, **19**, 1091-1102. https://doi.org/10.1109/TASL.2010.2077280

[18] Zhang, S. and Shirai, K. (2000) Visual Approach for Automatic Pitch Period Estimation. *IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing*, Istanbul, Vol. 3, 1339-1342.

[19] Niu, Y.D., Chen, F. and Chen, J. (2019) The Effect of F0 Contour on the Intelligibility of Mandarin Chinese for Hearing-Impaired Listeners. *The Journal of the Acoustical Society of America*, **146**, 85-91. https://doi.org/10.1121/1.5119264

[20] 宋黎明, 李明, 颜永红. 谐波显著度的基频提取方法[J]. 声学学报, 2015, 40(2): 294-299.