

基于大规模语音样本声调分析软件的设计与实现

丁玮明

云南师范大学汉藏语研究院, 云南 昆明

收稿日期: 2022年5月17日; 录用日期: 2022年6月21日; 发布日期: 2022年6月30日

摘要

基于大规模语音样本的声调分析软件在基础语音分析功能的同时, 创造性地设计并实现了批量处理通过大容量词表调查获取的大量单音节词语语音样本数据的程序, 分析流程包括: 音频载入、提取基频、确定声调段、筛选F0数据、数据预处理、数据归一化、调值估计、数据聚类, 整个过程全自动运行。该方法的应用价值在于能够与广泛使用的小样本分析方法形成参照、有效地挖掘出声调系统细微的变异与演变信息、为音系整理工作提供些许便利。

关键词

大样本, 语音分析, 声调分析

Design and Implementation of Tone Analysis Software Based on Large-Scale Speech Samples

Weiming Ding

Institute of Sino-Tibetan Studies, Yunnan Normal University, Kunming Yunnan

Received: May 17th, 2022; accepted: Jun. 21st, 2022; published: Jun. 30th, 2022

Abstract

Besides the basic speech analysis function, the tone analysis software based on large-scale speech samples creatively designs and implements the program for batch processing of a large number of voice sample data of monosyllabic words obtained through a large volume of word list. The analysis process includes: audio loading, F0 extraction, tone segment determination, F0 data screening, data preprocessing, data normalization, tone value estimation, data clustering. The whole process is fully

automatic operation. The application value of this method lies in that it can form a reference with the widely used small sample analysis method, effectively dig out the subtle variation and evolution information of the tone system, and provide some convenience for phonological system arrangement.

Keywords

Large Sample, Speech Analysis, Tone Analysis

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

对于有声调的语言或方言而言,其声调系统包括调值、调型和调类三个方面。调值是用数字表示的声调音值,一般用五度法表示;调型是声调的音高模式类型,即高低、升降的变化模式;调类指声调的分类,是根据调值归纳出来的类别。

借助语音实验手段的声调分析一般是根据预设的调类选取多组声调例词,同组声调例词尽可能遵循音位对立原则,如依据普通话的四声选取巴(bā)、拔(bá)、把(bǎ)、爸(bà)作为一组例词。然后寻找发音人朗读这些例词,并使用录音工具记录读音。再将录制的音频文件导入语音分析软件,确定声调负载段并从中提取基频(F0)。通常在所有语音样本中提取相同数量的基频值,将这些基频数据进行归一化处理,分别计算各个调类中所有语音样本相同点位归一化值的均值,将这些均值连成线即得到各个调类的音高曲线。最后根据音高曲线及其曲线上点的数值确定各个调值和调型。

以上分析方法是声调系统实验分析的主流方法,不同的研究者的具体操作可能有所不同,但分析流程基本一致。这种基于声调例词的分析方法我们称之为小样本分析,样本经过研究者的精挑细选,使用小容量的词表获取语音数据。小样本分析方法的优点在于能得出规整的声调格局图,赋予每个调类确定的调型和调值,对于语音系统的描写十分有利。其弊端在于规整的声调格局将复杂的发音活动简单化,难免会掩盖其中细微的声学信息。

为弥补小样本分析的不足,设计了基于大规模样本的声调分析程序,即从大量单音词的音频文件中提取的基频数据,基于这些数据,用自建的数据处理模型对每一个单字音进行调值估计,并对估计值相同的单字音进行自动归类,并进行频数统计。

2. 基于大样本调值估计程序的总体思路和应用价值

2.1. 总体思路

与小样本分析相比,大样本分析的对象不再是精心挑选的声调例词,而是用大容量词表(2000词)调查得到的全部单音节词的读音,对每一个单字音的五度调值进行估计。

估计程序的总体设计思路是:从调查所得的大量单字音的.wav音频文件中以固定的时长间隔提取基频,并筛选出属于声调负载段的数据。将得到的原始基频数据进行缺失填充、异常检测以及归一化等预处理之后,根据归一化值和曲线特征为每个语音样本推断调型、估计调值,将调值的估计值相同的单字音归并到一起,统计所有估计值包含的单字音数量(频数)。程序能够将每个单字音的数据集以Excel或CSV格式保存到本地磁盘,分门别类存放在相应的调值估计值命名的文件夹中。同时保存下来的还有各调值估计值的频数统计。

为了方便五度值的计算，原始基频的归一化方法选用的是 T 值法，该方法由南开大学的石锋教授提出，计算公式如下：

T 值法的公式是[1]：

$$T = \frac{\lg x - \lg b}{\lg a - \lg b} \times 5 \quad (\text{公式 1.1})$$

其中 x 是当前基频值， a 是所有基频中的最大值，即调域上限， b 是所有基频中的最小值，即调域下限。T 值公式将所有对数化的基频数据经转换后映射到[0, 5]的区间，将发音人的调域范围五等分，T 值处于[0, 1]为 1 度，[1, 2]为 2 度，[2, 3]为 3 度，[3, 4]为 4 度，[4, 5]为 5 度。当 T 值位于区间的临界位置时，需要根据实际情况做出判断。

不同发音人的调域上下限是不同的，T 值法不具备消除人际差异的能力，因此调值估计程序仅适用于来自单一发音人的大样本数据。如有多个发音人，需要多次运行。

2.2. 应用价值

2.2.1. 为基于小样本单字调分析结论提供参照

大样本分析方法的设计目标并非取代被广泛使用的小样本分析方法，而是与之形成互补。首先二者的应用价值有所区别，小样本方法用于系统性描写，大样本方法用于复杂声学信息的挖掘。其次，如果调值估计程序的基频归一化和五度调值转换的方法与小样本分析的方法保持一致，二者能够形成相互参照，既做一致性印证，也能做差异性比较。

2.2.2. 为挖掘声调的音位变体提供帮助

小样本分析基于少量的、精心挑选过的样本数据，并不利于发掘声调的音位变体。估计程序对大样本做出的调值估计和分类汇总，有助于弥补小样本分析的这一缺陷。一个调类在实际发音活动中可能会产生多种音高模式，基于大样本的分析有助于发现这些模式，并从中发掘出声调的音位变体和演变趋势方面的信息。

2.2.3. 为音系整理工作提供些许便利

音系的整理一般基于 2000 词、3000 词、5000 词等大容量词表的调查，这类词表往往依据义类排序，对于声调来说其实是乱序的。对于缺乏研究基础的语言，从中整理出声调系统的是比较费时费力的。由于调值估计程序包含了聚类的工作，即将估计值相同的单字音聚合在一起。相比于按照词表的顺序听音记音，按照聚类结果的顺序听记再分类整理或许能达到事半功倍的效果。

3. 基于大样本调值估计程序的设计

图 1 显示的是调值估计程序的分析流程：

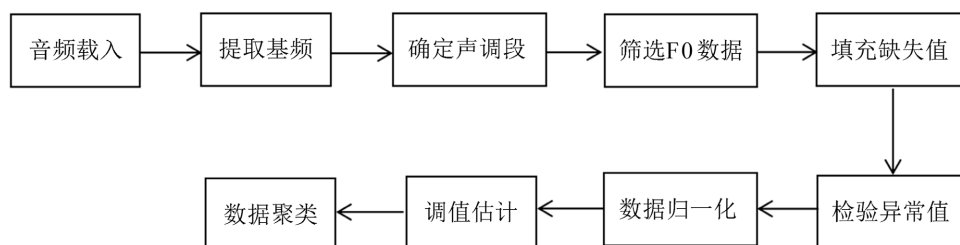


Figure 1. Flow chart of tone value estimation program

图 1. 调值估计程序流程图

3.1. 音频载入与基频提取

估计程序的第一步是载入存储单字音的.wav 音频文件并进行采样，一个音频文件仅存储一个单字的录音。采样率保留音频文件原本的采样率，既不降采样，也不过采样，得到采样后的音序列数据，并将其引进基频提取方法得到基频。

基频提取采用的 pYIN 算法，该算法于 Matthias Mauch 和 Simon Dixon 于 2014 年发布[2]，p 代表概率(probability)，YIN 指的是 De Cheveigné、Alain 和 Hideki Kawahara 于 2002 提出的 YIN 算法[3]。pYIN 是引入概率方法对 YIN 算法的改进，而 YIN 算法又是从基频提取常用的自相关函数法(Auto-correlation Function, ACF)发展而来。

YIN 算法大致的思路是：1) 针对 ACF 对振幅敏感导致出现频率或过高估计的弊端，创造性的采用差分函数(Difference Function, DF)替代自相关函数。2) 这种变换将 ACF 的乘积式变成了 DF 的减法式，寻找局部峰值的目标变换成寻求局部谷值，可能导致算法零时延处的谷值而忽略周期性波谷，对此改进 DF 提出累积均值归一化差分函数(Cumulative Mean Normalization Difference Function, CMNDF)，使得第一个真正的谷值出现前函数值始终保持较大的数值。3) DF 虽然能改善 ACF 取到比周期性峰值更高的值，但其自身不能完全杜绝取到比周期性谷值更低的值的错误，设置一个绝对阈值避免过低估计。4) 采用抛物线插值法解决信号不满足周期不是采样周期的倍数导致估计值仅有采样周期的一半的错误。5) 在周期值附近选择最优局部估计。YIN 取自汉语“阴阳”的“阴”，核心思想是用 DF 替代 ACF，这二者本身又是相互关联、一体两面的，该算法的设计者认为这类似于东方阴阳哲学。

pYIN 算法又称为概率 YIN 算法，对 YIN 算法的改进是放弃了 YIN 算法的绝对阈值，而改用服从贝塔分布的概率阈值，从而避免阈值本身小于周期性谷值而造成的错误。另外，pYIN 还加入了一个新步骤——基于隐马尔科夫模型(Hidden Markov Model, HMM)的音高追踪，定义了清音和浊音两种状态观测概率和转移概率，起到了浊音检测的功能。

基频提取不像基于小样本的单字调分析那样在声调负载段中平均取固定数量的点，而是以每间隔 11.6 ms 选取一个点。程序同时也实现了在声调段内平均提取固定数量基频点的方法，当样本量很大时程序的运行效率明显低于按固定的时间间隔提取。

3.2. 声调负载段的确定与基频数据的筛选

在声调分析过程中，不会选取算法提取到的全部基频值，而只是取得声调负载段的基频。在人工确定声调负载段的过程中，除了会观察语音可视化分析软件展示的音高曲线之外，还会观察语谱图。声调负载段在元音段，同时也是语谱图颜色较深、能量充足的音段。

让程序像人根据语谱图准确区分元音和辅音，去掉弯头降尾，截取出声调的稳定段是不容易实现的，因此只能退而求其次，通过能量来控制声调段的起终点。因此，在提取基频数据的同时，还需获取相应时间点上的能量数据。

时间点上的能量数据是“频率 - 能量”的二维数据，每隔一个频率间隔获取一个能量值，这个间隔就是带宽。时间点的能量我们选取带宽为 23.2 ms 的窄带语图上各频率段的能量最大值，之所以选择最大值，是因为最大值的数值最为显著。不同的元音能量的频率分布不一样，如果取时间点上的能量均值，兼顾到所有的元音，频率的上限最高要设定到 10,000 Hz，除了能量集中区之外，其他频率大量的能量值数值很小，经过平均之后能量充足的音段和能量衰减音段的数值差异性变小，不利于设定确定声调负载段的能量阈值。

我们使用的开发工具声音能量的 dB 值的取值范围是[-80, 0]，根据反复实验观察，认为-20 dB 到 0 dB 是能量充足段，因此将声调的结尾固定在最后一个不小于-20 dB 能量处。之所以设定的是最后一个而不

是第一个，是考虑到凹型曲折调的能量变化特点，凹调在谷值附近的能量会很小，甚至小于-20 dB，如果设为第一个很可能取到下降段的值就结束提取，造成错误。

由于 pYIN 算法带有浊音检测机制，当声母是清音时，算法得出的声调段的开头基本是可靠的。然而，当声母是浊音，尤其是鼻音或浊擦音时，浊辅音部分也会被提出基频，而这些都是声调分析中所不需要的，因此对于确定声调段的开头设置了-15 dB 能量阈值，如果算法得到的第一个基频点对应的能量大于-15 dB，使用算法得到的基频；否则就找到第一个大于-15 dB 的时间点上的基频作为声调段的开头。

为了尽可能得到声调的稳定段，对于通过能量阈值得到的声调段还会取前后 20% 的部分进行二次检验。前 20% 的部分重复确定开头的过程，后 20% 重复确定结尾的过程。

为了一定程度地消除弯头降尾的影响，会选择舍弃根据能量阈值确定的负载段首尾各 10% 的基频点。下图展示了一个声调段截取的示例，音频是云南省红河州金平县瑶族优勉话中的单音词“雪[bwen²⁴]”。

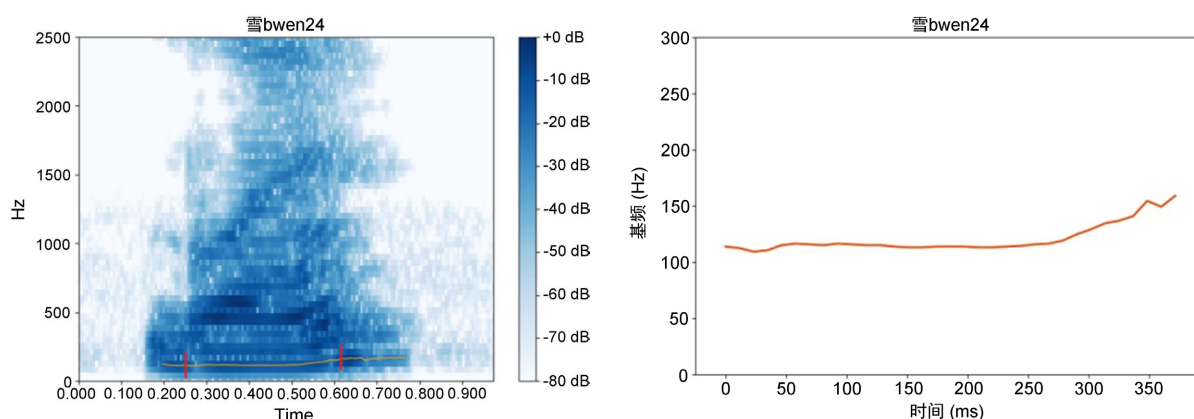


Figure 2. Example diagram of automatic capture of tone load segment

图 2. 自动截取声调负载段示例图

图 2 左侧图像的蓝色部分为窄带语图，橙色部分为基频折线，两条红色竖线的范围是截取的部分。可以看到原负载段中浊音声母也被计算了基频(全浊声母有时有鼻冠音色彩)，大部分被自动排除，负载段右端长长的一段降尾也被排除。右图是截取出来的基频折线，可见这种固定阈值的方法虽然不能完全得到完美的声调负载段，但是把大部分一般认为不参与声调分析的部分去除，基本能保证调值估计的可靠性。

3.3. 缺失数据填充与异常数据检测

3.3.1. 缺失值填充与填充值异常

算法提取基频时有时会出现缺失部分，其原因是实际发声的部分被算法判定为不发声，继而不显示音频。此时我们选择 pYIN 算法自带的缺失填充手段，它会根据音高趋势的猜测缺失部分可能的基频值，达到填充缺失值的目的。

但在含有下降趋势的声调中，算法很可能会将基频探测下限作为填充值，这里我们设定的基频探测范围是 75~500 Hz，这个值是不能接受的。因此，出现这类填充值，会作为异常数据处理，出现异常数据的单字音将不纳入后续的数据归一化与调值估计的分析过程。

3.3.2. 缺失值填充与填充异常

拟合吻合度异常指的是提取并筛选的基频值连成的基频折线与二次多项式拟合曲线的吻合度不足，这是为了排除出现基频过高估计或过低估计的问题，这种情况会造成基频曲线出现“大起大落”的现象。

经过模拟数据的多次实验，对基频数据做二次多项式拟合，然后将拟合数据和原始数据进行卡方检验，用卡方检验的 P 值来定义拟合曲线和原始基频曲线的吻合度。经过大量测试，如果声调曲线是正常的，得到的 P 值大部分会是 1.0 或近似于 1.0，一般会大于 0.95；如果出现“大起大落”的情况，P 值则极难达到 0.95。以下分别展示吻合度异常和吻合度正常的示例图 3。

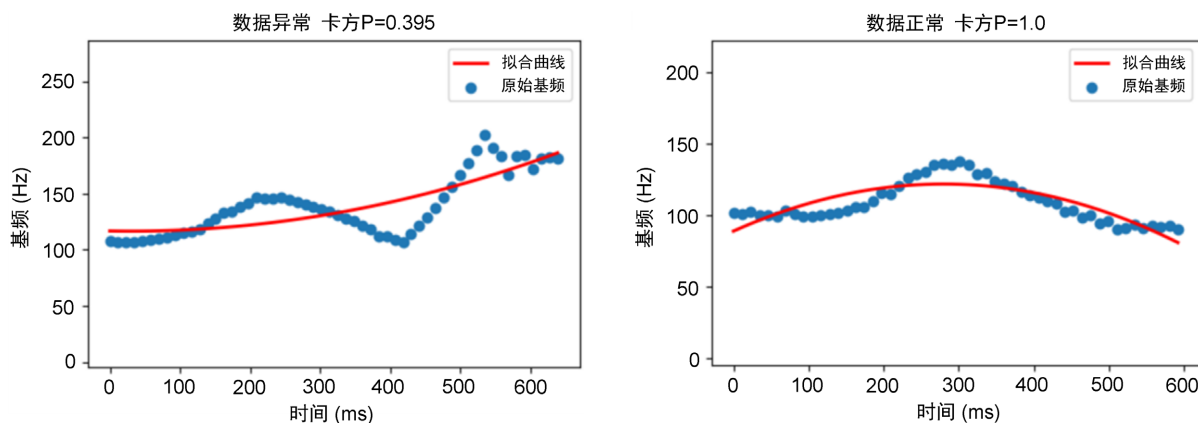


Figure 3. Example diagram of fitting degree

图 3. 拟合吻合度示例图

图 3 反映的是测试音频得出的数据，左侧显示的是末尾出现倍频错误，声调出现大幅度起落的示例，卡方检验报告的 P 值为 0.395，远小于 0.95；右侧显示的是一个正常的的数据，报告的 P 值为 0.999...，极其近似于 1。

衡量拟合优度的通常方法是考察 R 方值，但是实验证明 R 方值不如卡方检验的 P 值有效，尤其是对于平调，很容易出现吻合度极高但 R 方很低的情况，因为平调的形状近似于直线，抛物线模型对平调而言是过拟合的。

卡方检验通常被认为适用于定类数据相关分析的非参数假设检验方法，这里使用该检验方法，是把拟合数据和原始基频数据两组离散数据看作以时间尺度上的定类变量。卡方检验的原假设是一组数据之间具有相关性，作为两组数据的吻合度检验，则不以拒绝原假设的 0.05 为参考指标，而反过来以接受零假设的概率值 0.95 作为参考标准。

声调曲线总体上符合二次多项式的形态，李倩等(2020)将一种被称为“增长曲线分析”新的方法引入声调分析中时，采用的也是二次多项式的拟合模型[4]。后续的自动判断调型、调值的步骤也会借助二次多项式拟合。

3.3.3. 缺失值填充与填充异常

定义的第 2 类异常值是离群值，即数据中的极端值，这里选取四分位法进行离群值检测。具体流程根据计算获取的所有单字音基频值的下四分位数(Q1)和上四分位数(Q3)，并计算二者的差值得到四分位距(IQR)，将 $[Q3+3IQR, Q1-3IQR]$ 的区间定义可接受的数据范围，超出这一范围的值定义为离群数据。

缺失填充和异常检测数据预处理的全部过程，在基于大样本的分析过程中，三类异常数据只要出现一类，该样本就会被剔除。

3.4. 数据归一化与调值估计

数据归一化方法选用 T 值法，T 值法将原始数据经转换之后映射到 $[0, 5]$ 的区间上，方便五度值转换，然后基于五度数据估计各个单字音的调值。T 值转换公式以及五度值转换方法 2.1 中已有介绍，此处不再

赘述。

估计调值首先要判断调型，判断调型的方法是：首先对 T 值进行二次多项式拟合，得出形如 $y = ax^2 + bx + c$ 的函数，然后根据拟合曲线的特征得出调型的判断，在依据调型寻找声调目标，最终得出调值。

3.4.1. 平调的调型与调值判断

平调需同时满足两个条件中的其中一个，1) 所有 T 值的五度值相同，2) T 值的极差小于 0.5。设置第 2 个条件针对的是处于调域分界线附近的平调。

调值判断的方法是，将 T 值转换为 5 度值后，如果所有五度值在同一五度空间调值是十分确定的。出现两个及以上五度值，T 值数据极有可能是围绕在调域分界线附近分布的，此时将其视为带有临界值的数据，在调值估计值后面标注“(临)”。

3.4.2. 曲折调的调型与调值判断

不满足上述平调条件的单字音需要进一步判断其调型，根据抛物线对称轴的位置判断曲折调和非曲折调。曲折调抛物线的对称轴应该会在曲线中部，对于这个中部我们定义为曲线的 30%~70%处。即对称轴在声调段拟合曲线中间 30%~70%处，认为是曲折调；如果对称轴在 T 值拟合曲线的上述范围之外，认为是升调或降调。

曲折调的凸凹很容易区分，只需考察二次项系数 a 的正负，a 大于 0 说明先降后升，是一个凹型调；a 小于 0 说明先升后降，说明是凸型调。调值判断寻求其声调目标，首先找到极值的五度值，如果是凹调寻找以极值点为中心两侧的最大值，如果是凸调寻找以极值点为中心两侧曲线的最小值。

3.4.3. 升、降调的调型与调值判断

既不满足平调曲线特征，也不满足曲折调曲线特征的就是升调或降调。一般情况下，只需比较两端的 T 值，左侧大于右侧是降调，右侧大于左侧是升调。然而，为了稳妥起见，防止微升或微降的调型未被完全排除的弯头或升尾出现相反的结论，如一个上升幅度很小的升调弯头的 T 值甚至大于结尾降尾 T 值的情况，仍然用拟合曲线的特征判断调型，主要观察的是对称轴，因为对称轴所在的值为拟合曲线的极值。具体判断程序是：如果对称轴在中间段(这里用曲线 30%~70%处定义)左侧，且对称轴处的 T 值小于末端 T 值，说明是升调，反之是降调；如果对称轴在中间段右侧，且对称轴处的 T 值大于首端 T 值，说明是升调，反之是降调。最后一步是寻求 T 值的最大值和最小值的五度值，从而得出调值。

3.4.4. 临界值的检查与标注

当 T 值在临近调域分界线时，对调值的判断可能出现两可的情况，因此对 T 值进行临界值检查。对于临界值的定义，我们参考刘俐李在《江淮方言声调实验研究和折度分析》一书中所提出的用 ± 0.1 的“界域”来定义边界的观点[5]，检查作为用于确定调值的声调目标的 T 值是否在“界域”之内。

对于平调来说，所有的 T 值都是决定调值的声调目标，对此我们规定只要存在处于界域内的数据，就会被视为临界调值，因此只需要判断 T 值的极大值和极小值是否在界域内即可。升调和降调定位调值的声调目标是极大值和极小值，曲折调的声调目标是极值和以极值点为中心两侧曲线的局部极值。临界值检查就是判断这些 T 值点是否在界域内。

如果出现声调目标的 T 值在界域内的情况，我们选择直接标记出来，在调值估计值后标记“(临)”，与未被标记的估计值作为两类分别存储单字音数据集。

3.5. 数据聚类与频数统计

程序将各单音词样本的数据集根据调值估计结果自动聚类，将所有数据导出到本地。被标记的调值

作为单独的一类，如记为“41”和记为“41(右)”的单字音数据分别存储在两个文件夹中，在频数统计时也作为两类分别统计。

图 4 展示的是转储本地 Excel 的云南金平瑶族优勉语单音词“霜[sɔŋ³³]”，图中显示表头字段和前 5 条记录。

	A	B	C	D	E	F	G	H	I	J
1		时间(ms)	基频(Hz)	曲线拟合值	最大能量(dB)	是否为插值	原时间点(s)	是否离群	T值	T拟合值
2	0	0	129.8304842	131.4616627	-3.686840057	0	0.325079365	0	2.590361446	2.656042891
3	1	11.60998	130.5825845	131.208935	-3.486141205	0	0.336689342	0	2.620481928	2.645813772
4	2	23.21995	130.5825845	130.9697484	-3.085037231	0	0.34829932	0	2.620481928	2.636133773
5	3	34.82993	130.5825845	130.7441031	-3.111053467	0	0.359909297	0	2.620481928	2.627002896
6	4	46.43991	131.3390417	130.531999	-2.76566124	0	0.371519274	0	2.65060241	2.618421139

Figure 4. Example diagram of excel table for normal data

图 4. 正常数据 Excel 表示例图

字段中的“时间(ms)”和“原时间点(s)”分别指当前被截取的声调负载段的时间点和在原始音频文件中的时间点，前者的单位是毫秒，后者的单位是秒。“最大能量(dB)”指的是 23.2 ms 带宽下的该时间点上各频率段的最大能量，用于确定声调负载段，筛选基频数据。“曲线拟合值”和“T 值拟合值”分别是对原始基频数据和 T 值数据进行二次多项式拟合后该点函数值，第一次拟合用于异常检测，第二次拟合用于调型判断。“是否为插值”和“是否离群”分别标记当前时间点的基频数据是否为缺失填充值和是否为离群值，用 0 和 1 表示，0 表示“否”，1 表示“是”。

频数是指该调值估计值下包含的单字音数量，同时也会计算频率，即在全部数据量中的占比情况。频数统计数据也会以指定的数据表格形式存储指定的本地目录。该数据表有三个字段，分别是“调值估计”、“频数”、“频率”。

如果有未通过异常检测的数据，也会生成一个数据表，其他与正常数据的表同名的字段含义相同。“是否插值”字段 0 表示“否”，1 表示“是”，-1 表示“插值异常”；“是否离群”字段 0 表示“否”，1 表示“是”。这些数据表单独保存在一个目录中。图 5 展示的数据表显示表头字段和最后 5 条记录。

	A	B	C	D	E	F	G	H
1		时间(ms)	基频(Hz)	曲线拟合值	最大能量(dB)	是否为插值	原时间点(s)	是否离群
13	11	127.7098	103.0465	98.2654288	-10.65491295	0	0.325079365	0
14	12	139.3197	93.94983	93.34158109	-13.24061775	0	0.336689342	0
15	13	150.9297	86.15238	87.93160316	-17.20194054	1	0.34829932	0
16	14	162.5397	79.00208	82.03549503	-11.60949802	1	0.359909297	0
17	15	174.1497	75	75.6532567	-18.06777954	-1	0.371519274	0

Figure 5. Example diagram of sheet for abnormal data

图 5. 异常数据表示例图

程序会自动在本地保存各估计调值的频数统计表，并为每个估计调值单独创建目录，将属于该估计调值的样本数据存入其中。图 6 展示程序自动处理云南金平瑶族优勉语发音人 970 条单音词的语音样本得出的估计调值频数统计，为节省空间表格经过转置，仅展示前 10 条数据。

	A	B	C	D	E	F	G	H	I	J	K
1	调值估计	33	32	33(临)	42	324	343	324(临)	44	32(临)	232
2	频数	134	78	61	55	52	40	28	21	21	21

Figure 6. Example diagram of transposed frequency table

图 6. 频数统计表(转置)示例图

程序也会报告大样本计算得出的一些基础数据，包括处理的样本数、F0 数据量、F0 的极值(调域上下限)以及均值和标准差，并转储本地 TXT 文件，如图 7 所示。

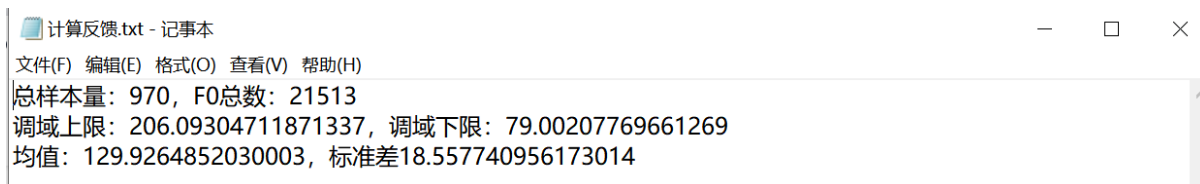


Figure 7. Example picture of basic data report

图 7. 基础数据报告示例图

4. 声调分析软件的实现

4.1. 主界面与基础语音分析功能的实现

分析软件的实现使用 Python 语言，该语言拥有丰富而强大的数据科学工具包，且能够快速便捷地开发图形化界面。McFee 等开发者创造的开源音频分析工具包 Librosa [6]提供了各类声音信号分析的功能模块供其他开发者调用。声调分析软件的主界面如图 8 所示。

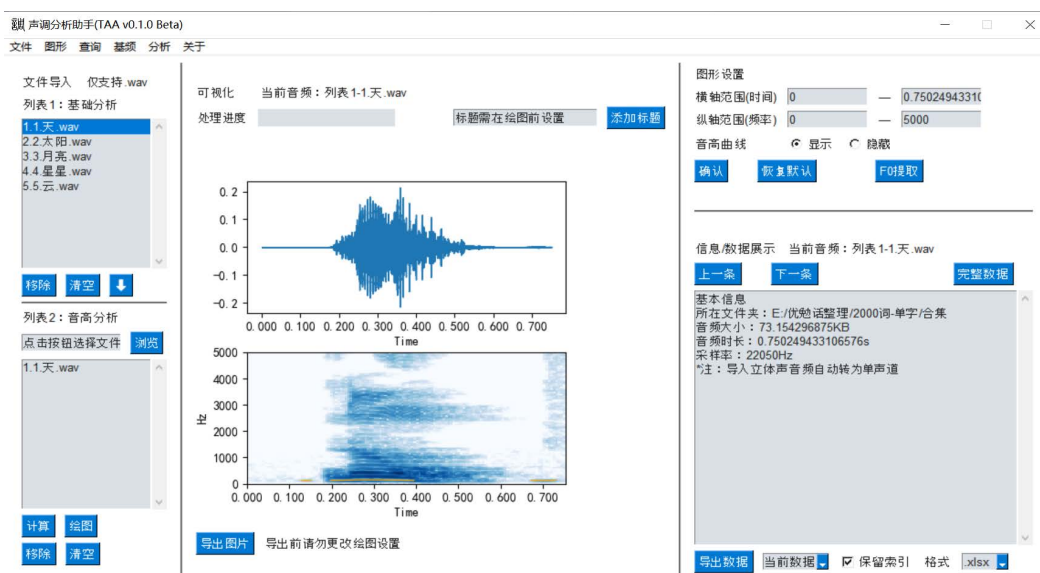


Figure 8. Example picture of basic data report

图 8. 基础数据报告示例图

界面上的菜单栏、按钮、文本框、下拉框等元素的功能含义已经有较为直观呈现，不再一一介绍。左侧的两个文件列表，列表 1 用于单个语音样本的分析，列表 2 主要支持多个语音样本的分析，上述调值估计程序即适用于列表 2 中的语音样本。

中间可视化模块展示的是云南金平瑶族优勉语单音词“天[lun⁴²]”的波形图和窄带语图，语图上有音高曲线。顶部菜单栏中的“图形”可以更改图形展示模式，除了以上展示模式，还可选择只展示波形图、频谱图或音高曲线一种图形，得到更大、更清晰的图片，频谱图支持修改带宽，图像支持保存到本地。

右侧的图形设置模块中的“F0 提取”按钮用于在任意时间范围内提取基频数据，支持两种提取模式，一是设置固定的时间间隔，如每隔 10 ms 提取一个基频；二是设置固定的基频点数量，如等距离提取 10

个基频点。该设置位于顶部菜单栏“基频”中，数据支持保存到本地。图 9 展示在“天[lun⁴²]”音频中的 0.25~0.38 秒之间提取 10 个基频点的运行效果。

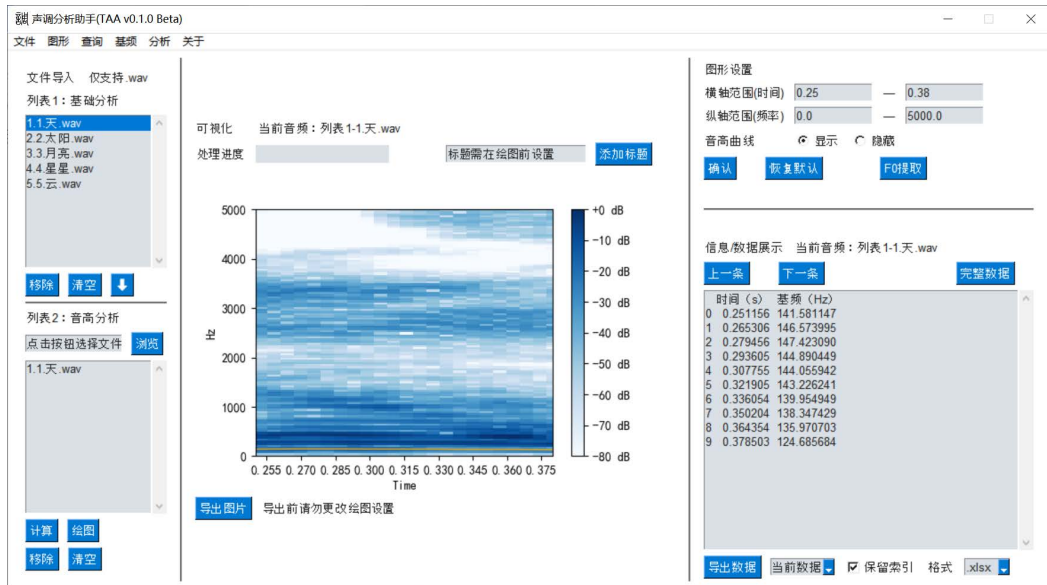


Figure 9. Example picture of F0 extraction function
图 9. F0 提取功能示例图

右下方的文本框显示了提取出来的基频数据。这一功能有效支持了广泛使用的基于声调例词表的小样本分析方法。

导入列表 2 中的语音文件仅显示声调负载段的音高曲线，负载段的确定遵循 3.2 中介绍的方法，支持上述两种基频提取模式。在列表 2 中导入多个音频，点击“绘制”即可将它们的音高曲线绘制在一张图上，数据支持批量导出，如图 10 所示。

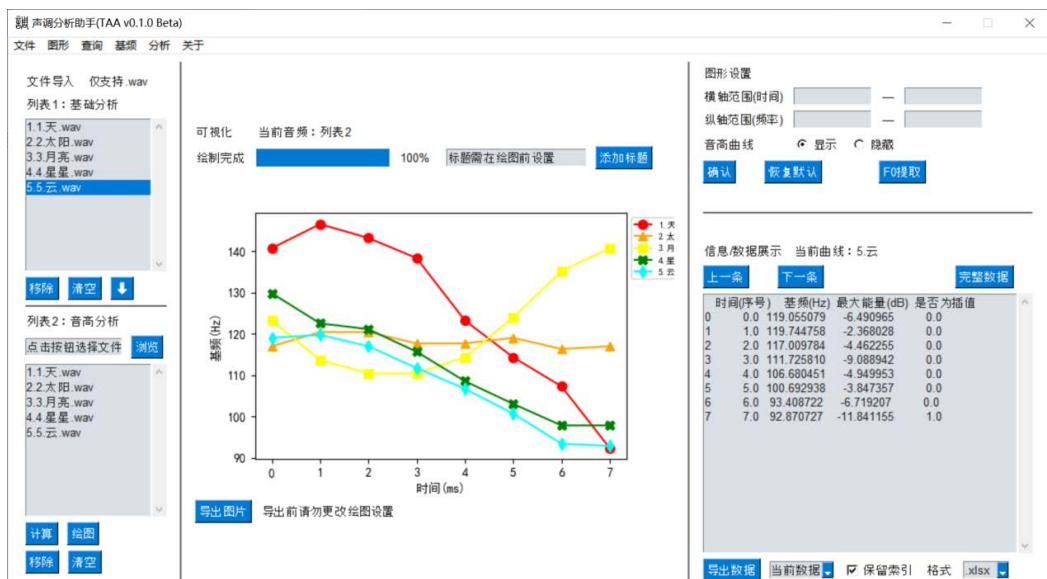


Figure 10. Example picture of multi-curve plot function
图 10. 多曲线绘制功能示例图

列表 2 中的图像即可绘制原始基频图像，也可绘制 T 值转换后的图像，还可绘制两类数据的二次多项式拟合图像，只需在顶部菜单栏“分析”中开启相应功能即可完成绘制。

4.2. 基于大规模语音样本分析功能的实现

调值估计程序的全部流程在软件中是全自动执行的，用户只需将来自同一发音人的单音节词的语音样本批量导入列表 2 中，在顶部菜单栏设置好各类参数，点击左下方“计算”按钮即可自动运行，使用右下角的“导出数据”功能批量导出数据。

各类参数的设置均在顶部菜单栏中，大样本分析流程各环节中涉及到的参数，如基频提取算法、基频点时间间隔、基频探测范围、自相关函数窗口大小、pYIN 算法的两个概率参数、确定声调段的能量阈值、首尾端的舍弃比例、确定声调负载段的能量阈值、缺失值的填充方法等都是可设置的状态。所有的参数均有默认值，导入语音样本后可以一键计算。

自动计算时不会有任何可视化效果，一是没有必要，二是影响性能。页面中间上方实时显示执行进度，右下方文本框实时显示当前处理的音频文件产生的数据。如图 11 所示。



Figure 11. Example picture of large sample analysis execution status
图 11. 大样本分析执行状态示例图

分析完成后，点击右下方“导出数据”按钮，并在其右侧下拉框中选择“全部数据”即可批量导出到指定文件夹，数据按照估计调值分类存储，数据表为 3.5 中展示的形式。

5. 结语

基于大规模语音样本的分析方法能够帮助研究者快速处理使用大容量词表获取的单音词语音样本。笔者使用实现了这种方法的语音分析软件批量分析云南金平县的瑶族优勉语的近千条单音词样本，同时也使用了基于小样本的分析方法确定声调格局。研究表明二者的数据具有可比性，且能够形成印证。在程序得出所有样本估计调值以及聚类结果的基础上，用统计方法梳理和分析各调类的调值估计结果，从中发现各调类的发音稳定性存在差异，并且挖掘出到多条音位变体和演变趋势的规律。优勉语对于笔者而言属于完全陌生的语言，基于大样本的自动化分析程序为快速将研究推向深入提供了较大的帮助，相

关研究成果不就将以学位论文的形式公开发布。

软件也基于 Python 强大的数据科学和语音处理相关工具包实现了各类基础语音分析功能，能够适应诸多语音分析任务。这表明基于 Python 开发语音实验工具的路径是可行的，能够根据实际需求定制分析工具，为实验语音学方法的探索提供了新的技术路径。

基金项目

云南师范大学 2020 年度研究生创新基金一般项目(蓝靛瑶话声调实验研究, 项目编号: ysdjys2020017)。

参考文献

- [1] 石锋. 语音学探微[M]. 北京: 北京大学出版社, 1990.
- [2] Matthias, M. and Dixon, S. (2014) pYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions. 2014 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 4-9 May 2014, 659-663.
- [3] De Cheveigné, A. and Kawahara, H. (2002) YIN, a Fundamental Frequency Estimator for Speech and Music. *The Journal of the Acoustical Society of America*, **111**, 1917-1930. <https://doi.org/10.1121/1.1458024>
- [4] 李倩, 史濛辉, 陈轶亚. 声调研究中的一种新统计方法——“增长曲线分析”法在汉语方言研究中的运用[J]. *中国语文*, 2020(5): 591-608+640.
- [5] 刘俐李. 江淮方言声调实验研究和折度分析[M]. 成都: 巴蜀书社, 2007.
- [6] McFee, B., Raffel, C., Liang, D., Ellis, D.P.W., McVicar, M., Battenberg, E. and Nieto, O. (2015) Librosa: Audio and Music Signal Analysis in Python. *Proceedings of the 14th Python in Science Conference*, 18-25. <https://doi.org/10.25080/Majora-7b98e3ed-003>