

高阶最优 LPC 根值筛选的共振峰估计算法研究

龙华, 苏树盟

(昆明理工大学信息工程与自动化学院, 云南 昆明 650031)

摘要: 为解决线性预测 (LP) 共振峰检测误差较大的问题, 针对高阶 LP 提取共振峰中伪根难以去除及极点交互带来的频谱混叠, 提出一种基于高阶最优 LP 系数根值筛选的共振峰估计算法。考虑提高 LP 的阶次, 获取含有更高线性峰值拟合精度的线性系统根值。采用语音数字共振模型约束共振峰的根幅值范围, 滤除非共振峰频率值对应的伪根。结合功率加权来预加重信号的主要频谱成分, 提高频谱峰值频率区分度。实验结果表明, 所提算法消除了高阶邻近极点间的相互干扰, 去除了非共振峰谐波对应的高阶伪根, 与其他算法相比, 所提算法定位语音信号共振峰频率的误差更低, 更能稳定精确地检测共振峰。

关键词: 线性预测; 数字共振; 功率加权; 共振峰

中图分类号: TN912.32

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2022113

Research on formant estimation algorithm for high order optimal LPC root value screening

LONG Hua, SU Shumeng

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650031, China

Abstract: It is difficult to remove false roots and spectrum aliasing caused by pole interaction in the formant extraction of high-order linear prediction (LP). In order to solve the problem of large error of LP formant detection, a formant estimation algorithm based on high-order LP coefficient root value screening was proposed. The root value of linear system with higher precision linear fitting peak was obtained by LP of higher order. The speech digital resonance model was used to constrain the amplitude range of the root of the formant and filter the pseudo-root corresponding to the formant frequency. Finally, combined with power weighting, the main spectral components of the signal were pre-weighted to improve the spectrum peak frequency discrimination. Experimental results show that the proposed algorithm eliminates the interference between high-order adjacent poles and removes the high-order pseudo-roots corresponding to the harmonics of non-formants. Compared with the other algorithms, the proposed algorithm has a lower error in locating the formant frequency of speech signals and can detect formants more stably and accurately.

Keywords: linear prediction, digital resonance, power weighting, formant

0 引言

共振峰检测是降低高质量语音比特率任务的前提, 在多种语音信号处理场景中^[1-3], 共振峰频率的精确估计是必要的。实现共振峰频率精确估计的

方法包括倒谱法、线性预测 (LP, linear prediction) 分析法^[4]、经验模态分解 (EMD, empirical mode decomposition) 希尔伯特黄变换 (HHF, Hilbert-Huang transformation) 非线性模型法^[5]和深度学习法^[6]。其中, 倒谱法实现简单但是缺乏系统所

收稿日期: 2022-01-06; 修回日期: 2022-04-30

通信作者: 苏树盟, 20202104018@stu.kust.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61761025)

Foundation Item: The National Natural Science Foundation of China (No.61761025)

需的准确性^[7]；HHF 模型针对语音的非平稳特性以高时间分辨率可靠地提取语音信号共振峰，但 EMD 过程中算法复杂度较高，影响了检测的速度，降低了检测的有效性；深度学习法输入语音片段进行模型训练，基于特定数据集训练可以得到较好的结果，但受数据集影响较大，导致泛化能力较低。LP 分析技术公式化语音信号的本质模型，在声道建模中运用广泛^[8]，一个 p 阶的全极点滤波器通常在 $8 \leq p \leq 20$ 的范围内适用于语音^[9]，短时语音的线性预测系数 (LPC, linear prediction coefficient) 对应的系统频谱包含与共振峰频率相关的声道共振频率值，LPC 提取共振峰的普遍问题是不能从语音中正确、稳定地提取共振峰频率^[10]，并且存在去除共振峰的伪根困难以及线性预测极点交互两大问题。传统线性预测共振峰估计 (TLP-FE, traditional linear prediction-formant estimation) 方法^[8]在保留共振峰和 LP 响应幅度谱基本形状的前提下，减小 p 的取值，牺牲信号拟合的精度以去除大部分和激励有关的频谱特征 (基音谐波与共振峰谐波)，减少伪根数，降低线性预测极点交互，带来的问题是限制了共振峰提取的精度。在传统方法的基础上，文献[11]依据相同音素、不同语音帧间共振峰的连续性构建相邻帧间的峰值向量完备组合矩阵，通过求取满足相邻帧的峰值组合中欧几里得距离最小的谱峰根值的方法来防止共振峰的混叠，采用变阶 LPC 的方式保证 LPC 共振峰提取的精确度，这种方法在共振峰提取上过滤了基音及共振峰谐波，但是高阶次、高精度地求取共振峰仅在少部分语音帧中得以体现。

此外，LPC 加权方案在共振峰提取中获得了较好的性能，近些年提出的加权方案包括稳定加权线性预测 (SWLP, stabilized weighted LP)^[12]、扩展加权线性预测 (EWLP, extension weighted LP)^[13]、功率加权线性预测 (PWLP, power weighted LP)^[14]及数据驱动加权线性预测 (DWLP, data-driven weighted LP)^[15]。文献[14]对语音信号离散傅里叶频率集合中的共振频率部分进行分离，加重共振频率对应谱幅值，提出一种功率加权 LPC 的共振峰估计 (PWLP-FE, power weighted LP-formant estimation) 算法，此算法强调语音信号的主要频谱成分，对消除极点交互带来的峰值混叠具有较好的效果。目前，已有的 LPC 共振峰检测算法均存在 LP 伪根难以去除的问题，PWLP-FE 算法改变了原始语音的

语音波形消除极点交互，但仍然无法有效地去除高阶 LP 伪根，在共振峰频率精确估计的性能上存在较大误差。

为解决现有 LPC 共振峰估计存在的 LPC 伪根干扰、LPC 极点交互的问题及降低共振峰检测的误差，本文首先增大 p 的取值，提高线性预测系统频谱对语音信号的拟合程度，在不同阶次下分析语音信号共振峰频率的计算精度，并且引入语音数字共振模型^[8,10]来约束不同阶次线性预测共振峰根幅值范围；然后通过匹配阶次的根幅值来筛选线性系统根值的方法来减少伪根数；最后结合功率加权来消除极点交互，解决谐波产生的干扰，提出一种高阶最优 LPC 根值筛选的共振峰估计 (HOLP-FE, high order optimality LP-formant estimation) 算法。实验结果显示，相比于最近提出的 PWLP-FE 算法，本文提出的 HOLP-FE 算法可以有效消除高阶次下的极点交互及滤除混叠频率，能更加精确地提取出语音信号的共振峰频率。

1 LPC 共振峰检测

1.1 功率加权 LPC

功率加权是对原始语音在频域上进行幅值加权修正，使语音信号的主要频谱成分更具区分度。信号能量大 (谱峰) 的区域比信号能量小 (谱谷) 的区域线性预测谱和信号谱匹配更好^[16]，语音能量大的区域比语音能量小的区域对语音频率的加权更多，进而语音谱峰处对应的频率取值的区分度更加清晰。

功率加权 LPC 是对功率加权修正的信号进行 LP 分析并通过语音信号估计线性时变系统模型参数的过程。不同的语音离散时间模型对应着不同的线性时变系统，在线性时变系统模型 $h(n)$ 中输入激励信号值 $u(n)$ 后，输出的语音采样信号值为

$$s(n) = u(n) * h(n) \quad (1)$$

在语音离散时间模型中，模型当前的输出是模型当前的所有输入激励信号以及模型过去的输出语音采样信号的线性组合，输入与输出之间的时域关系表示为

$$s(n) = \sum_{i=1}^p a_i s(n-i) + G \sum_{l=0}^q b_l u(n-l) \quad (2)$$

其中， n 表示信号采样点数， p 表示时变线性系统的极点阶数， q 表示时变线性系统的零点阶数， a_i 表

示模型极点系数， b_l 表示模型零点系数， G 表示模型增益因子。

在时变线性系统模型中，模型 $h(n)$ 的传递函数为

$$H(z) = \frac{S(z)}{U(z)} = G \frac{1 + \sum_{l=1}^q b_l z^{-l}}{1 - \sum_{i=1}^p a_i z^{-i}} \quad (3)$$

其中， $H(z)$ 为语音信号的系统函数， $S(z)$ 为语音采样信号 $s(n)$ 的频域表达式， $U(z)$ 为激励信号 $u(n)$ 的频域表达式。语音采样清音及鼻音浊音在声道传输函数中同时包含零点和极点^[17-18]，语音采样非鼻音浊音在声道传输函数中模型简化只包含极点，相应的时变线性系统模型参数中的零点系数 b_l 全为 0，声道模型为自回归 (AR, auto-regressive) 全极点模型，并且当时变线性系统数字滤波器的极点阶数 p 足够大时，可以采用全极点模型来拟合所有的语音。

对采样语音信号做线性分析构建 LP 方程，求解 LPC 方程，得到语音信号的线性预测系数 a_i ^[16]，同时 a_i 也对应 AR 模型方程的实数根及共轭复根。

1.2 共振峰检测约束及语音信号修正

相较于文献[11]提出的最小化相邻帧间峰值向量完备组合矩阵欧几里得距离，以及通过极点交互因子搜索频谱极点角局部极大值^[19]求谱峰值，从语音数字共振的整体频率贡献角度引入数字共振模型约束，定位谱峰峰值的搜索由局部改进至整体，有效避免了局部搜索步距界定繁杂的计算及峰值带宽与局部搜索边界失配造成的局部峰值定位失误。

LPC 共振峰检测依据 LPC 原理确定匹配阶次的 AR 模型极点系数，以极点系数频域特性确定对应的峰值频率及峰值带宽。共振峰检测的数字共振模型是浊音声道传输函数的稳态形式^[8,10]，描述的是与共振峰频率相对应的数字共振器的频域贡献。将 AR 模型传递函数写成关于共振峰频率 F_j 以及谱峰带宽 W_j 稳态数字共振级联形式为

$$H(z) = G \prod_{j=1}^{\frac{p}{2}} \frac{1 - 2e^{-\frac{W_j}{F_s}} \cos\left(\frac{2\pi F_j}{F_s}\right) + e^{-\frac{2W_j}{F_s}}}{1 - 2e^{-\frac{W_j}{F_s}} \cos\left(\frac{2\pi F_j}{F_s}\right) z^{-1} + e^{-\frac{2W_j}{F_s}} z^{-2}} \quad (4)$$

其中， $\frac{p}{2}$ 始终向下取整， F_s 为语音信号采样频率值，可根据线性预测系数值 a_i 采用 LPC 内插法或 LPC

求根法等方法确定语音信号谱峰频率值 F_j 以及谱峰带宽 W_j 。

式(4)中，当 p 取 8 阶或 9 阶时， j 的取值为 $\{1,2,3,4\}$ ，对应 4 个谱峰共振器，可从这 4 个谱峰频率值 F_j 中找到满足 $150 \text{ Hz} < F_j < \frac{F_s}{2}$ 且 $W_j < 700 \text{ Hz}$ 的 3 个共振峰 F_1 、 F_2 和 F_3 ，并用这 3 个频率值作为数字共振约束参数。

为提高语音信号谱峰区分度^[14]，结合功率加权原理将原始语音信号 $s(n)$ 修正为 $s'(n)$ ，用于共振峰检测的修正信号改变语音原始波形；结合功率加权和 LPC 做共振峰检测，修正的信号 $s'(n)$ 为

$$s'(n) = \text{IDFT} \left\{ \text{DFT}[s(n)] \left| \text{DFT}[s(n)] \right|^2 \right\} \quad (5)$$

修正的信号 $s'(n)$ 的波形不同于 $s(n)$ ，将 $s'(n)$ 信号作为 LP 的输入可得到更具频率区分特性的 LPC 根值。

2 高阶根值筛选的 LPC 共振峰估计算法

高阶根值筛选的 LPC 共振峰估计算法流程如图 1 所示。在对语音信号做共振峰估计时，需要对语音信号做预处理端点检测等前期处理，预处理对语音信号进行预加重，削减低频信息以降低基频对共振峰检测的干扰，增强高频信息以增加高频谱线的幅值区分度，端点检测有效地隔离无语段，进而对有话段做进一步的 LP 分析。

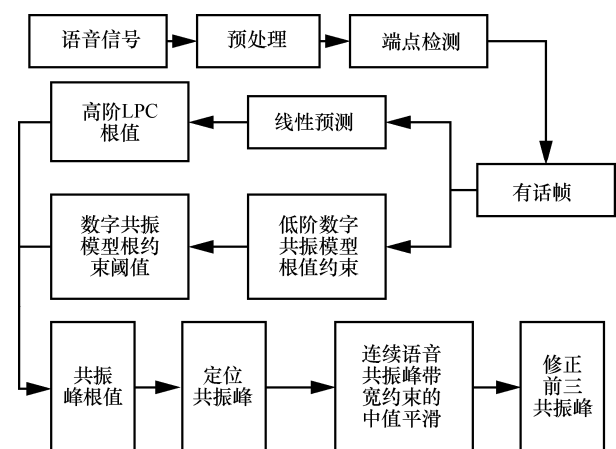


图 1 高阶根值筛选的 LPC 共振峰估计算法流程

本节首先对有话段做线性预测分析及低阶数字共振模型根值约束分析，求解高阶 LPC 根值及数字共振模型根约束阈值，然后以约束阈值得到约束

条件下的共振峰根值，最后定位、平滑修正连续语音信号共振峰。

2.1 求解高阶 LPC 峰值频率

低阶线性预测仅保留了语音信号 LP 响应幅度谱的基本形状，提高线性预测阶次增大了信号的拟合度，消除低阶线性拟合带来了共振峰拟合精度的牺牲。

信号线性预测分析中，LPC 对应 AR 模型的极点系数 a_i ，其与时变线性系统传输函数根值 z_i 间的关系^[16,20]为

$$\begin{cases} H(z) = \frac{G}{1 - \sum_{i=1}^p a_i z^{-i}} \\ H(z) = \frac{G}{\prod_{i=1}^p (1 - z_i z^{-1})} \end{cases} \quad (6)$$

由式(6)，依据 a_i 求解共轭复根 z_i

$$z_i = \alpha_i + j\beta_i \quad (7)$$

其中， α_i 与 β_i 分别表示共轭复根 z_i 的实部与虚部。

共轭复根 z_i 对应的频率 f_i 及带宽 w_i 在频率域 f 中的位置关系为

$$\begin{cases} z^{-i} = e^{-\frac{j2\pi f i}{F_s}} \\ z_i = e^{-\frac{w_i \pi}{F_s}} e^{\frac{j2\pi f_i}{F_s}} = e^{-\frac{\pi}{F_s}(w_i - j2f_i)} \end{cases} \quad (8)$$

联立式(6)和式(8)，得到 LP 模型的频率域 $H(f)$

$$H(f) = \frac{G}{\prod_{i=1}^p \left(1 - e^{-\frac{\pi}{F_s}(w_i - j2f_i)} e^{\frac{j2\pi f}{F_s}} \right)} = \frac{G}{1 - \sum_{i=1}^p a_i e^{-\frac{j2\pi f i}{F_s}}} \quad (9)$$

在 $H(f)$ 中，存在关于 f_i 的式(10)成立，即功率加权后的谱峰为共振峰值点。

$$\begin{cases} \left. \frac{d(|H(f)|^2)}{df} \right|_{f=f_i} = 0 \\ \left. \frac{d^2(|H(f)|^2)}{df^2} \right|_{f=f_i} < 0 \end{cases} \quad (10)$$

由式(7)求得单位圆内的根幅值 γ_i 为

$$\gamma_i = \sqrt{\alpha_i^2 + \beta_i^2} \quad (11)$$

联立式(7)、式(8)和式(11)可得

$$\begin{cases} f_i = \frac{F_s \arctan\left(\frac{\beta_i}{\alpha_i}\right)}{2\pi} \\ w_i = \frac{-2\ln(\gamma_i)}{2\pi} = \frac{-\ln(\gamma_i)}{\pi} \end{cases} \quad (12)$$

利用式(12)得到高阶线性预测的系数 a_i 对应的谱峰频率值 f_i 及谱峰带宽值 w_i ，在谱峰频率值 f_i 中包含共振峰频率值、共振峰谐波频率值及基音谐波频率值，需要在所有谱峰频率值中加入约束来滤除谐波，定位出前 3 个共振峰频率值。

2.2 线性预测根值的数字共振模型约束

AR 模型在取较低的阶次时，有效地隔离了大部分和激励有关的频谱特征，滤除了频谱中的谐波分量，避免了极点交互作用，获得了低拟合精度的 LPC 根值，求解得到满足式(4)的 3 个谱峰参考频率 F_1 、 F_2 和 F_3 。

采用数字共振模型确定的参考频率作为约束对 z_i 进行筛选，定义根幅值阈值变量为 γ 且 $0 < \gamma < 1$ ，前三共振峰为 f_k 且 $k = \{1, 2, 3\}$ 。结合式(11)，满足 $\gamma_i > \gamma$ 的 z_i 的实部与虚部分别为

$$\alpha_i(\gamma), \beta_i(\gamma) = \arg \{ \gamma_i - \gamma > 0 \}, i \in \mathbb{Z}^+, 0 < m \leq p \quad (13)$$

由式(12)可知，关于根幅值阈值变量 γ 的 LPC 谱根值频率 $f_i(\gamma)$ 为

$$f_i(\gamma) = \frac{F_s \arctan\left(\frac{\beta_i(\gamma)}{\alpha_i(\gamma)}\right)}{2\pi} \quad (14)$$

引入数字共振频率约束，从高阶拟合的 $f_i(\gamma)$ 中定位关于根幅值阈值变量 γ 的 3 个共振峰频率 $f_k(\gamma)$ 为

$$f_k(\gamma) = \arg \min_{f_i(\gamma)} |f_i(\gamma) - F_k|, k = \{1, 2, 3\} \quad (15)$$

在高阶次下，求满足 3 个峰值在数字共振约束下总体误差最小的根幅值筛选阈值 γ' 。此时， γ' 与线性预测的阶次相关。

$$\gamma' = \arg \min_{\gamma} \left\{ \sum_{k=1}^3 |f_k(\gamma) - F_k| \right\} \quad (16)$$

2.3 高阶根值筛选

结合式(11)以及式(16)筛选高阶 LPC 对应的 z_i ，得到满足 $\gamma_i > \gamma'$ 的实部与虚部分别为

$$\alpha_i', \beta_i' = \arg\{\gamma_i - \gamma' > 0\}, i = (0, \dots, m), i \in \mathbb{Z}^+, 0 < m \leq p \quad (17)$$

由式(12)得到高阶次下 3 个共振峰频率 f_i' 为

$$f_i' = \frac{F_s \arctan\left(\frac{\beta_i}{\alpha_i}\right)}{2\pi} \quad (18)$$

区分峰值频率对应第 k 共振峰 f_k

$$f_k = \arg \min_{f_i} |f_i - F_k|, k = \{1, 2, 3\} \quad (19)$$

求解连续共 M 帧的语音共振峰及带宽，第 $m(0 \leq m \leq M - 1)$ 帧语音信号 $s_m(n)$ 的共振峰频率为 $f_k(m)$ ，联立式(11)、式(12)、式(17)，求得第 m 帧语音信号 $s_m(n)$ 的第 k 共振峰对应的带宽 $w_k(m)$ 。同一个音素对应的共振峰频率值是连续的，采用 $w_k(m)$ 作为判决对共振峰频率进行中值平滑求解，满足式(20)或式(21)的 $f_k(m - 1)$ 和 $f_k(m + 1)$ 。

$$\begin{cases} f_k(m) - f_k(m - 1) > w_k(m) \\ f_k(m + 1) - f_k(m) < w_k(m) \end{cases} \quad (20)$$

$$\begin{cases} f_k(m) - f_k(m - 1) < w_k(m) \\ f_k(m + 1) - f_k(m) > w_k(m) \end{cases} \quad (21)$$

进行中值平滑处理修正 $f_k(m)$ ，如式(22)所示，得到连续语音每一帧最终的共振峰频率。

$$f_k(m) = \frac{f_k(m + 1) + f_k(m - 1)}{2} \quad (22)$$

3 实验及其分析

为了抑制基频谱线幅度值，降低基频对共振峰检测的干扰，并且抵消声门脉冲引起的高端频谱幅度下跌，对语音信号进行预加重处理来提升信号高频，强调声道部分的频谱，得到频率特性平坦、共振峰频率特性接近的信号。实验对共振峰频率已知的连续合成单元音语音进行共振峰估计，对语音端点检测的有语段的音素进行逐帧估计。

3.1 实验条件

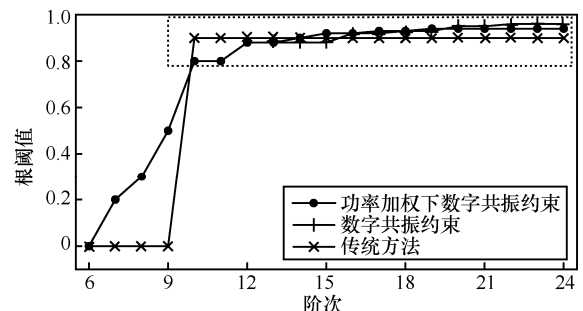
本节指定反映声道信息及声带振动信息的共振峰频率和基音频率参数合成参考语音^[21]，使用共振峰频率确定的合成语音来验证共振峰提取算法的

准确性。信号选取[a]、[i]、[u]这 3 个元音的连续合成信号，指定 3 个元音的基音频率均为 140 Hz，3 个元音的共振峰带宽相同，前 3 个共振峰对应带宽分别为 150 Hz、200 Hz、250 Hz，前 3 个共振峰频率值元音[a]分别为 906 Hz、1 359 Hz、2 609 Hz，元音[i]分别为 328 Hz、2 421 Hz、2 890 Hz，元音[u]分别为 406 Hz、765 Hz、2 312 Hz。其中，音频采样频率为 8 000 Hz，帧长为 25 ms，帧移为 1 ms，共计 161 帧。

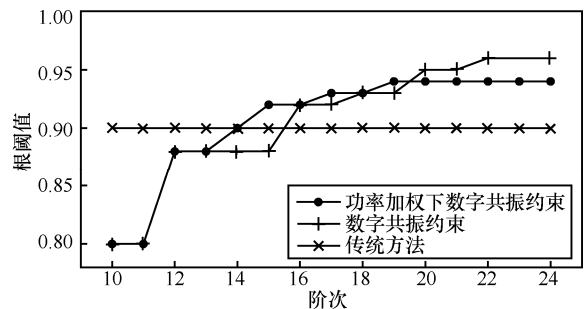
本文实验选取合成元音[a]的语音信号分析数字共振根值约束下的根判定阈值、共振峰根值分布，选取合成元音[a]对应的多帧平均值分析共振峰频率预测误差，以整体连续包含 3 个元音的合成语音分析算法稳健性及误差分布。

3.2 数字共振根值约束下的根判定阈值

通过实验计算 AR 模型取不同阶次时共振峰根幅值的范围，分析本文引入的数字共振根幅值约束下的根幅值阈值与阶次的匹配性，如图 2 所示。



(a) 6~24阶根阈值与阶次关系



(b) 10~24阶根阈值与阶次关系

图 2 根阈值与阶次关系

图 2(b)是图 2(a)中 10~24 阶根阈值与阶次关系的放大图。如图 2(a)所示，采用传统方法求取 3 个共振峰频率时，阶次低于 10 时设置阈值为 0，高于 10 时设置阈值为 0.9，粗略区分共振峰峰值对应的根以及非共振峰峰值对应的根，传统方法伪根滤除时，误差较大。如图 2(b)所示，在高阶次时共振峰根幅值与伪根根幅值区分度较小，线性预测阶次增

大, 峰值拟合精度增大, 根幅值筛选阈值也有所增大, 或高于 0.9 或低于 0.9, 故不能以固定的传统标准来衡量。由图 2(b)可知, 由于功率加权突出了频谱峰值的幅值, 同阶次时引入功率加权与不引入功率加权下数字共振约束根幅值阈值大小不同, 小于 19 阶时, 引入功率加权下数字共振根幅值阈值较大, 大于 19 阶时则相反; 不引入功率加权需要更加精确的阈值来筛选根值。

3.3 共振峰根值分布比较

线性预测多项式的共轭复根对应单位圆上所有的零极点, 靠近单位圆上的极点是声道系统函数频率响应谱峰峰值对应的根, 单位圆上根幅值大于根幅值阈值的极点是共振峰峰值对应的根, 在图 3 中分别为“所有根”“峰值根”“最终根”。是“峰值根”但不是“最终根”的根值是由声道谐振作用、声门脉冲形状、传输效应等多因素产生的, 即伪根。对不同阶次 PWLP-FE^[14]所提功率加权 LPC 共振峰检测以及本文提出的 LPC 共振峰检测 HOLP-FE 的根值分布进行分析, 探究 2 种方法共振峰提取对伪根识别的效果。

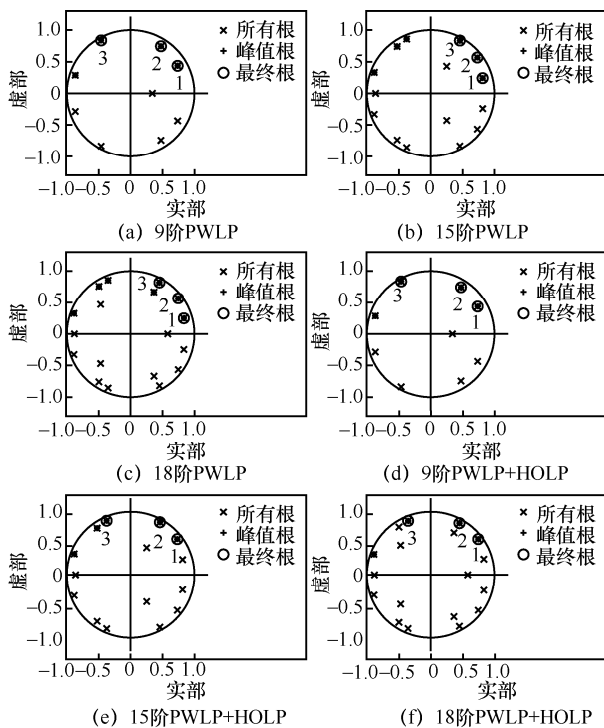


图 3 不同方法 LPC 共振峰检测根值分布

对不同方法 LPC 共振峰检测根值分布进行分析可以看出, 线性预测的根值数目与线性预测的阶数相同; 对根进行筛选, 满足共振峰条件的根小于或等于 LPC 阶数的一半, 说明提高线性预测阶次产

生更多难以辨认的伪根。在此基础上, 比较 PWLP-FE 和 PWLP-FE+HOLP-FE 的共振峰在单位圆上的分布, 如图 3(a)所示, PWLP-FE 在 9 阶 LPC 时能够正确地筛选出共振峰对应的根值点; 如图 3(b)和图 3(c)所示, PWLP-FE 在 15 阶及 18 阶线性预测时将第一共振峰以及第二共振峰对应的根值误判为第二共振峰以及第三共振峰对应的根值, 并且第一共振峰根值点筛选错误; 如图 3(d)、图 3(e)和图 3(f)所示, PWLP-FE+HOLP-FE 对线性预测根值的筛选更精确, 将满足条件的根锁定在更小的范围, 并且在所讨论 LPC 的不同阶次上均能够正确地找出共振峰对应的根值点, 解决了 PWLP-FE 在高阶时不能较好地辨认伪根的问题。

3.4 共振峰频率及频谱包络比较

本节计算单独一帧合成语音的共振峰频率, 通过估计图 4 中单帧语音在不同阶次线性预测时的前三共振峰, 比较 TLP-FE、PWLP-FE、HOLP-FE 以及 PWLP-FE+HOLP-FE 的前三共振峰分布, 分析不同阶次下这 4 种方法检测共振峰受峰值混叠的影响情况。

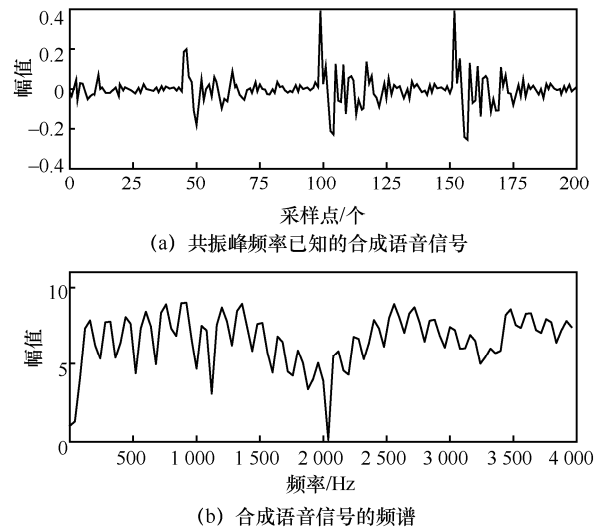


图 4 语音波形及频谱

图 5 对 4 种方法的 9 阶 LPC 频谱包络及共振峰峰值分布进行了分析。从图 5 中可以看出合成信号语音频谱的幅度趋势, 所比较的 4 种方法均较平滑且有 4 个明显峰值的 LPC 谱包络线, 在共振峰峰值附近不存在相互混叠的峰值。这 4 种方法均能正确地找出 LPC 谱包络的峰值点, 区别在于 PWLP-FE 对信号进行了加权, 改变了原本信号的 LPC 谱包络结构, 谱峰值中心点相较于 HOLP-FE 以及 TLP-FE 产生偏差。

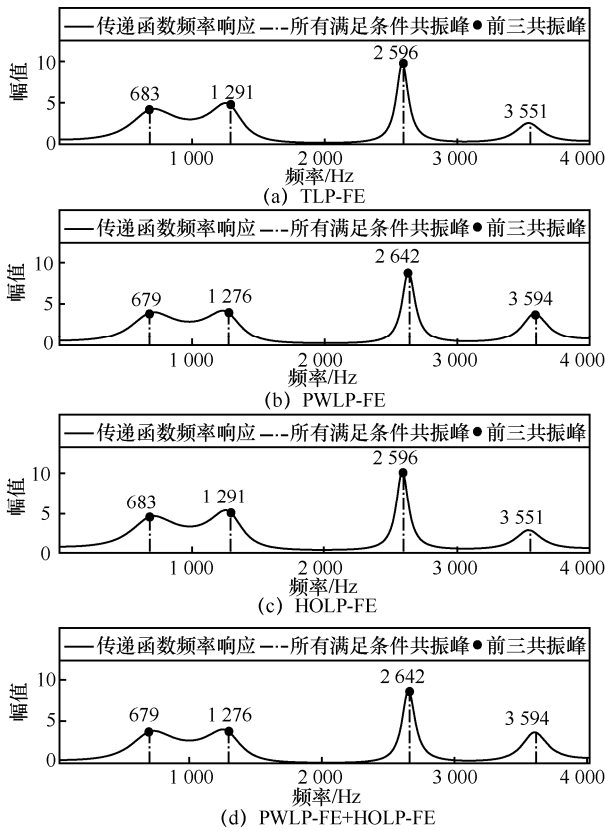


图 5 9 阶频谱包络及共振峰峰值分布

图 6 对 4 种方法的 15 阶 LPC 频谱包络及共振峰峰值分布进行了分析。从图 6 中可以看出，15 阶比 9 阶 LPC 包含更多的频谱细节，在传递函数幅频响应中，第一个尖峰峰值前及第三个尖峰峰值后出现了平缓的峰值，产生了混叠程度较低、数量较少的混叠峰值。图 6(a)为 TLP-FE 的共振峰检测方法，此检测方法将频率为 332 Hz 的基音谐波误判为第一共振峰，将频率为 799 Hz 的第一共振峰及频率为 1 324 Hz 的第二共振峰误判为第二共振峰及第三共振峰；图 6(b)为 PWLP-FE 的共振峰检测方法，此检测方法将频率为 366 Hz 的基音谐波误判为第一共振峰，将频率为 839 Hz 的第一共振峰及频率为 1 360 Hz 的第二共振峰误判为第二共振峰及第三共振峰；图 6(c)和图 6(d)分别为 HOLP-FE 和 PWLP-FE+HOLP-FE 的共振峰检测方法，此 2 种检测方法滤除 332 Hz 及 366 Hz 的基音谐波，实现了 3 个共振峰的正确判定。

图 7 对 4 种方法的 18 阶 LPC 频谱包络及共振峰峰值分布进行了分析。相较 15 阶谱包络，最大的不同在于 18 阶 LPC 包络中包含更多的幅频响应频谱细节，存在更多并且混合程度更大的混叠峰值。4 种方法在 18 阶时对 3 个共振峰判定情况与

15 阶 LPC 频谱包络分析近似，但值得注意的是，HOLP-FE 以及 PWLP-FE+HOLP-FE 有效地滤除了另外 2 种方法所不能滤除的混叠峰值。

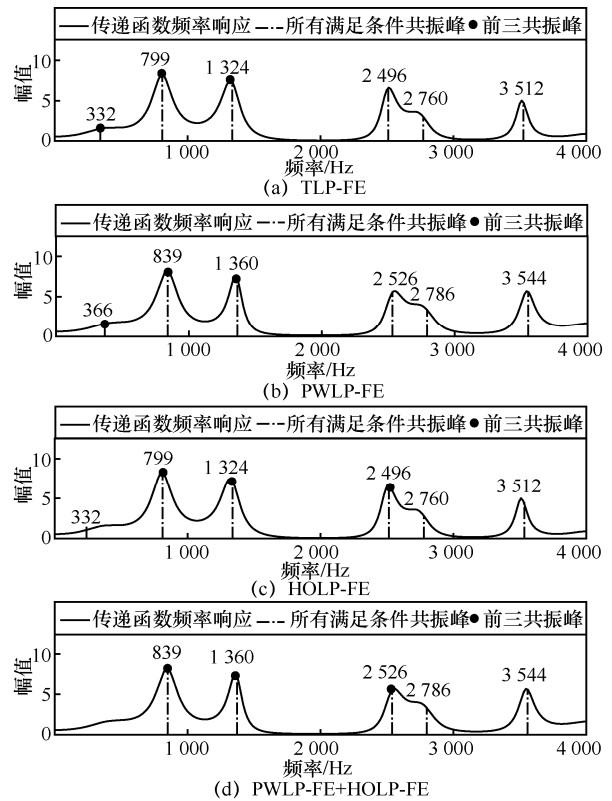


图 6 15 阶频谱包络及共振峰峰值分布

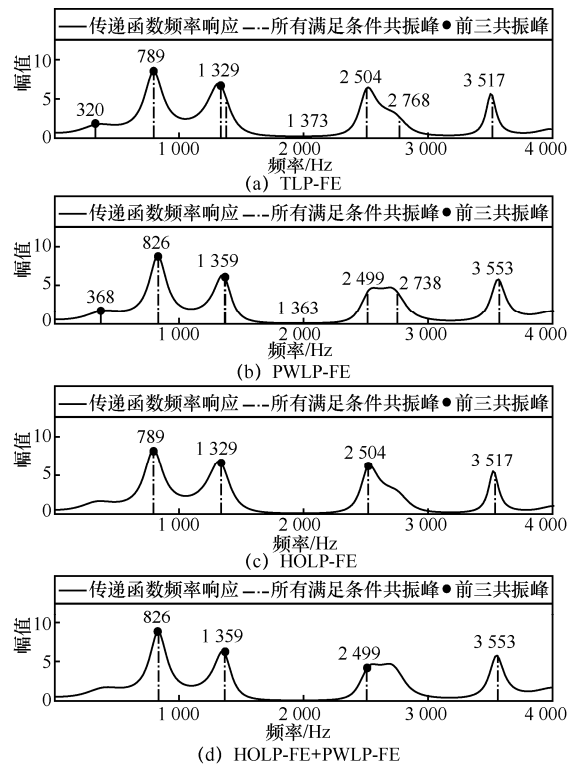


图 7 18 阶频谱包络及共振峰峰值分布

3.5 共振峰频率预测误差分析

实验计算不同方法共振峰频率检测相对误差,一方面,通过同阶次下不同方法相对误差值的大小来分析不同方法估计共振峰的精度;另一方面,通过相同方法不同阶次下相对误差值的大小分析不同信号拟合度下共振峰提取的精度。

定义共振峰预测频率 \hat{F} 相对于合成参考频率 F 的相对误差 ξ 为

$$\xi = \frac{\hat{F} - F}{F} \times 100\% \quad (23)$$

表 1、表 2 及表 3 中的数据为 TLP-FE、PWLP-FE、HOLP-FE 以及 PWLP-FE+HOLP-FE 共振峰检测方法检测合成语音信号的 3 个共振峰 \hat{F}_1 、 \hat{F}_2 、 \hat{F}_3 的频率值及总体值 \hat{F} 相对合成信号共振峰频率值 F 的相对误差,数据表示的相对误差值直观地体现了 HOLP-FE 相对其他方法在共振峰检测精确度上的优势,分析共振峰频率相对合成信号共振峰参考频率的误差可知,相对误差值越小,共振峰检测精确度越高。

当 $p=9$ 时,代表取较低的线性预测阶次。对于 \hat{F}_1 , TLP-FE 和 HOLP-FE 的误差值相等, PWLP-FE 和 PWLP-FE+HOLP-FE 的误差值相等; 4 种方法的

误差值相对较大,处在 -26% 到 -24% 之间,但是 HOLP-FE 的误差绝对值最小,比 PWLP-FE 的误差绝对值小 0.5% 。

对于 \hat{F}_2 , TLP-FE 和 HOLP-FE 的误差值相等, PWLP-FE 和 PWLP-FE+HOLP-FE 的误差值相等; 4 种方法的误差值相对较小,处在 -7% 到 -5% 之间,并且 HOLP-FE 的误差绝对值最小,比 PWLP-FE 的误差绝对值小 1.1% 。

对于 \hat{F}_3 , TLP-FE 和 HOLP-FE 的误差值相等, PWLP-FE 和 PWLP-FE+HOLP-FE 的误差值相等; 4 种方法的误差值相对最小,处在 -1% 到 2% 之间,并且 HOLP-FE 的误差绝对值最小,比 PWLP-FE 的误差绝对值小 0.8% 。

对于 \hat{F} , TLP-FE 和 HOLP-FE 的误差值相等, PWLP-FE 和 PWLP-FE+HOLP-FE 的误差值相等; 4 种方法的误差值相对较小,处在 -7% 到 -5% 之间,并且此时 PWLP-FE+HOLP-FE 的误差绝对值最小, HOLP-FE 与 PWLP-FE 的结合不会拉低功率加权方法性能的下限。

当 $p=15$ 时,代表取较高的线性预测阶次。对于 \hat{F}_1 , PWLP-FE 比 TLP-FE 的误差绝对值小 0.4% , 二者的误差值较大,处在 -64% 到 -61% 之间;

表 1 9 阶共振峰频率预测误差分析

共振峰	F / Hz	TLP-FE	PWLP-FE	HOLP-FE	PWLP-FE+HOLP-FE
\hat{F}_1	906	-24.6%	-25.1%	-24.6%	-25.1%
\hat{F}_2	1 359	-5.0%	-6.1%	-5.0%	-6.1%
\hat{F}_3	2 609	-0.4%	1.2%	-0.4%	1.2%
\hat{F}	4 874	-6.2%	-5.6%	-6.2%	-5.6%

表 2 15 阶共振峰频率预测误差分析

共振峰	F / Hz	TLP-FE	PWLP-FE	HOLP-FE	PWLP-FE+HOLP-FE
\hat{F}_1	906	-63.3%	-62.9%	-11.8%	-7.4%
\hat{F}_2	1 359	-41.2%	-38.2%	-2.6%	-0.1%
\hat{F}_3	2 609	-49.2%	-47.8%	-4.3%	-3.2%
\hat{F}	4 874	-49.6%	-47.9%	-5.2%	-3.1%

表 3 18 阶共振峰频率预测误差分析

共振峰	F / Hz	TLP-FE	PWLP-FE	HOLP-FE	PWLP-FE+HOLP-FE
\hat{F}_1	906	-64.6%	-59.3%	-12.9%	-8.8%
\hat{F}_2	1 359	-41.9%	-39.2%	-2.2%	0
\hat{F}_3	2 609	-49.1%	-47.9%	-4.0%	-4.2%
\hat{F}	4 874	-50.0%	-47.6%	-5.1%	-3.9%

PWLP-FE+HOLP-FE 比 HOLP-FE 的误差绝对值小 4.4%，二者的误差值较小，处在-12%到-7%之间；4 种方法的误差值最小的是 PWLP-FE+HOLP-FE，误差值为-7.4%。

对于 \hat{F}_2 ，PWLP-FE 比 TLP-FE 的误差绝对值小 3.0%，二者的误差值较大，处在-42%到-38%之间；PWLP-FE+HOLP-FE 比 HOLP-FE 的误差绝对值小 2.5%，二者的误差值较小，处在-3%到 0 之间；4 种方法的误差值最小的是 PWLP-FE+HOLP-FE，误差值为-0.1%。

对于 \hat{F}_3 ，PWLP-FE 比 TLP-FE 的误差绝对值小 1.4%，二者的误差值较大，处在-50%到-47%之间；PWLP-FE+HOLP-FE 比 HOLP-FE 的误差绝对值小 1.1%，二者的误差值较小，处在-5%到-3%之间；4 种方法的误差值最小的是 PWLP-FE+HOLP-FE，误差值为-3.2%。

对于 \hat{F}_4 ，PWLP-FE 比 TLP-FE 的误差绝对值小 1.7%，二者的误差值较大，处在-50%到-47%之间；PWLP-FE+HOLP-FE 比 HOLP-FE 的误差绝对值小 2.1%，二者的误差值较小，处在-6%到-3%之间；4 种方法的误差值最小的是 PWLP-FE+HOLP-FE，误差值为-3.1%。

对于 TLP-FE 和 PWLP-FE 来说，虽然前者共振峰提取改进的优势是每个共振峰检测的相对误差绝对值总是小于前者，但是在高度线性拟合下后者的相对误差仍然保持在较高的水平，已经不能再应用于共振峰检测。

对于 HOLP-FE 和 PWLP-FE+HOLP-FE 来说，二者的相对误差均保持在较低的水平，并且 PWLP-FE+HOLP-FE 的相对误差总是小于 HOLP-FE。

当 $p=18$ 时，代表取更高的线性预测阶次。不同方法各个共振峰提取的精度与 $p=15$ 时大致相近，仍然处在较低水平。

结合表 1~表 3 比较相同方法各个共振峰在不同阶次下的提取精度。对于 TLP-FE 和 PWLP-FE 来说，共振峰的提取只在取较低阶次时能够适用于共振峰检测，在高阶次时预测误差较大而不再适用。对于 HOLP-FE 和 PWLP-FE+HOLP-FE 来说，在取较低阶次时，其各个共振峰提取的精度可与 TLP-FE 和 PWLP-FE 持平；在取较高阶次时，其各个共振峰提取的相对误差较取较低阶次时均有所降低，并且保持在较低水平；在取更高阶次时，其

各个共振峰提取的相对误差与取较高阶次时相接近，处在较低水平。

总体上，相较于 TLP-FE 和 PWLP-FE，HOLP-FE 和 PWLP-FE+HOLP-FE 在 $p=15$ 和 $p=18$ 时各个共振峰的提取相对误差均明显减小，这表明 HOLP-FE 通过提高线性预测阶次实现了共振峰提取精度的提升。

3.6 算法稳健性分析

本节对多帧合成连续单元音语音进行共振峰检测，通过实验计算不同方法在不同阶次下共振峰检测的相对误差绝对值的平均值，如图 8 所示，分析不同阶次下不同方法在多帧语音下共振峰提取的稳健性。

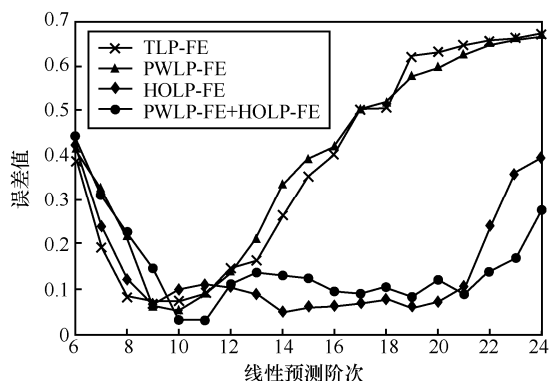


图 8 线性预测不同阶数连续语音共振峰预测误差

对比实验的 4 种方法，当线性预测阶次取 9 阶到 12 阶时，4 种方法预测误差值相差不大且均处在较低水平。9 阶时，TLP-FE、PWLP-FE、HOLP-FE 的误差值几乎相同并且小于 PWLP-FE+HOLP-FE；10 阶到 12 阶时，PWLP-FE+HOLP-FE 有较好的稳健性，线性预测误差处在较低水平。当线性预测阶次取 13 阶到 21 阶时，TLP-FE 和 PWLP-FE 的误差值急剧增大，算法不稳健，HOLP-FE 和 PWLP-FE 的误差值始终在 0.2 以下，说明多帧语音下共振峰提取算法是稳健的。当线性预测阶次取 6 阶到 8 阶和 22 阶到 24 阶时，4 种方法都不能有效提取共振峰，这时，共振峰提取的线性预测阶次超出线性预测阶次的限制范围，算法不具备稳健性。

3.7 连续多帧语音不同阶次下模型误差分布分析

通过实验对多帧合成连续单元音语音进行共振峰检测，计算各帧语音相对误差的绝对值，绘制不同阶次下各种方法共振峰检测概率分布，如图 9 所示，比较连续语音下 TLP-FE、PWLP-FE 及 PWLP-FE+HOLP-FE 模型共振峰提取的整体性能。

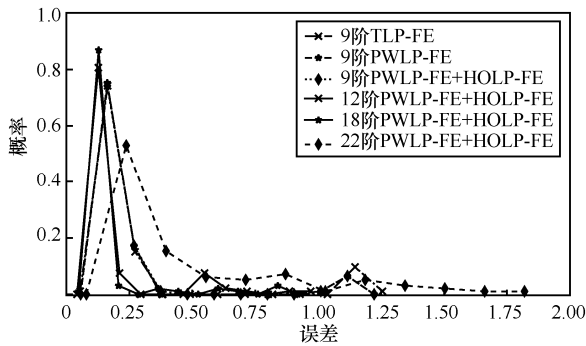


图 9 不同阶次下各种方法共振峰检测概率分布

由图 9 可知, 3 种模型在 9 阶时语音共振峰预测误差概率分布相重合, 即在 9 阶下 3 种模型的性能是相当的。PWLP-FE+HOLP-FE 通过提高线性预测阶次去除伪根、消除极点交互带来的峰值混叠来提升整体共振峰提取的性能, 图 9 中, “12 阶 PWLP-FE+HOLP-FE” 曲线峰值点概率分布值高于 “9 阶 PWLP-FE” 且 “12 阶 PWLP-FE+HOLP-FE” 曲线峰值点共振峰预测误差值低于 “9 阶 PWLP-FE”, 这表明 PWLP-FE+HOLP-FE 在 12 阶时的模型性能优于 PWLP-FE; “18 阶 PWLP-FE+HOLP-FE” 曲线比 “12 阶 PWLP-FE+HOLP-FE” 曲线更加 “高瘦”, 这表明 PWLP-FE+HOLP-FE 在 18 阶时的预测误差分布更小且更集中, 对应模型性能更优; “22 阶 PWLP-FE+HOLP-FE” 曲线峰值分布概率减小、峰值误差值更大且误差区间有较长的 “拖尾”, 性能大幅度下降, 这表明本文算法的总体性能不是随阶次提升而线性增长的, 在阶次稳健范围内取较高阶次是实现模型整体最优性能的关键。

4 结束语

本文对基于 LPC 共振峰检测的算法做出改进, 研究提高线性预测阶次对提取共振峰的影响, 针对提高线性预测阶次带来的多伪根以及多极点交互的问题, 最小化含语音数字共振模型约束共振峰提取误差, 分析线性预测阶次与根幅值筛选阈值的关系。采用数字共振约束下的根幅值反馈的方式获取匹配高阶次的低误差率筛选阈值来去除伪根, 并且结合功率加权突出频谱峰值的幅值, 消除共振峰提取过程中的极点交互, 实现精准有效的共振峰提取。实验表明, 在高阶次时, 所提算法相较于最近提出的功率加权 LPC 共振峰检测算法能更精确稳定地提取出语音信号的共振峰频率。

参考文献:

- [1] VANITHA L M, SUDHA S. Noise diminution and formant extraction on vowels for hearing aid users[J]. *Multimedia Tools and Applications*, 2020, 79(5/6): 3729-3741.
- [2] LIU Z T, REHMAN A, WU M, et al. Speech emotion recognition based on formant characteristics feature extraction and phoneme type convergence[J]. *Information Sciences*, 2021, 563: 309-325.
- [3] 曹冲, 解焱陆, 张劲松. 不同共振峰分布下元音对声调感知的影响[J]. *清华大学学报(自然科学版)*, 2018, 58(4): 352-356.
CAO C, XIE Y L, ZHANG J S. Influence on tone perception from vowels with different formant distributions[J]. *Journal of Tsinghua University (Science and Technology)*, 2018, 58(4): 352-356.
- [4] MCCANDLESS S. An algorithm for automatic formant extraction using linear prediction spectra[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1974, 22(2): 135-141.
- [5] 黄海, 陈祥献. 基于 Hilbert-Huang 变换的语音信号共振峰频率估计[J]. *浙江大学学报(工学版)*, 2006, 40(11): 1926-1930.
HUANG H, CHENG X X. Speech formant frequency estimation based on Hilbert-Huang transform[J]. *Journal of Zhejiang University*, 2006, 40(11): 1926-1930.
- [6] DISSEN Y, GOLDBERGER J, KESHET J. Formant estimation and tracking: a deep learning approach[J]. *The Journal of the Acoustical Society of America*, 2019, 145(2): 642-653.
- [7] 赵涛涛, 杨鸿武. 结合 EMD 和加权 Mel 倒谱的语音共振峰提取算法[J]. *计算机工程与应用*, 2015, 51(9): 207-212.
ZHAO T T, YANG H W. Formant extraction algorithm of speech signal by combining EMD and WMCEP[J]. *Computer Engineering and Applications*, 2015, 51(9): 207-212.
- [8] RABINER L R, SCHAFFER R W. 数字语音处理理论与应用[M]. 刘加, 张卫强, 何亮, 译. 北京: 电子工业出版社, 2011.
RABINER L R, SCHAFFER R W. *Theory and applications of digital speech processing*[M]. Translated by LIU J, ZHANG W Q, HE L. Beijing: Publishing House of Electronics Industry, 2011.
- [9] TREMAIN E T. The government standard linear predictive coding algorithm: LPC10[J]. *Speech Technol*, 1982, 1(1): 40-49.
- [10] YAN Z Y, ZHAO H M. Formant estimation algorithm based on digital waveguide models[C]//*Proceedings of 2010 2nd International Conference on Information Engineering and Computer Science*. Piscataway: IEEE Press, 2010: 1-4.
- [11] MESSAOUD Z B, HAMIDA A B. Combining formant frequency based on variable order LPC coding with acoustic features for TIMIT phone recognition[J]. *International Journal of Speech Technology*, 2011, 14(4): 393-403.
- [12] MAGI C, POHJALAINEN J, BÄCKSTRÖM T, et al. Stabilised weighted linear prediction[J]. *Speech Communication*, 2009, 51(5): 401-411.
- [13] KERONEN S, POHJALAINEN J, ALKU P, et al. Noise robust feature extraction based on extended weighted linear prediction in LVCSR [C]//*Proceedings of the 12th Annual Conference of the International Speech Communication Association*. Saarland: DBLP, 2011: 1-5.

- [14] FRÉIN R D. Power-weighted LPC formant estimation[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2021, 68(6): 2207-2211.
- [15] SUDHARSHAN R, RAMALINGAM C S. A data-driven weighted LP method for formant estimation[C]//Proceedings of 2020 IEEE 4th Conference on Information & Communication Technology. Piscataway: IEEE Press, 2020: 1-6.
- [16] XU L, LIU H J, ZHANG S L, et al. Speech feature extraction based on linear prediction residual[C]//Proceedings of 2020 IEEE 5th International Conference on Signal and Image Processing. Piscataway: IEEE Press, 2020: 768-772.
- [17] DIGGLE P J, WHITTLE P. Prediction and regulation by linear least-square methods[J]. Biometrics, 1984, 40(3): 871-877.
- [18] YOKOTA K, ISHIKAWA S, KOBA Y, et al. Inverse analysis of vocal sound source using an analytical model of the vocal tract[J]. Applied Acoustics, 2019, 150(7): 89-103.
- [19] XU K N, HU W, WANG Y H. An improved singer's formant extraction method based on LPC algorithm[C]//Proceedings of 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). Piscataway: IEEE Press, 2017: 1-5.
- [20] ZAPATA J L G, DÍAZ MARTÍN J C, VILDA P G. Fast formant estimation by complex analysis of LPC coefficients[C]//Proceedings of 2004 12th European Signal Processing Conference. Piscataway: IEEE Press, 2004: 737-740.
- [21] ZHANG H J, YANG Y. Fundamental frequency adjustment and formant transition based emotional speech synthesis[C]//Proceedings of 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery. Piscataway: IEEE Press, 2012: 1797-1801.

[作者简介]



龙华 (1963-), 女, 回族, 云南大理人, 博士, 昆明理工大学教授, 主要研究方向为无线网络及音频信号处理。



苏树盟 (1996-), 男, 云南保山人, 昆明理工大学硕士生, 主要研究方向为音频信号处理、语音识别。

《通信学报》第十届编辑委员会

顾 问： 邬江兴 刘韵洁 方滨兴 于 全 郑建华 费爱国
何 友 尹 浩 陆建华 陆 军 姚富强 沈学民
王怀民 王金龙 崔铁军

主任委员：张 平

副主任委员：张延川 马建峰 杨 震 沈连丰 陶小峰 刘华鲁

委 员：

丁 群 王汝言 王良民 龙 军 卢建民 田 辉 田有亮
田俊峰 朱洪波 仲 红 任保全 刘西蒙 许文俊 李 伊
李少谦 李凤华 李玉峰 李建东 李陶深 杨 亮 吴 怡
吴 巍 吴启晖 吴晓平 沙学军 沈玉龙 宋令阳 宋铁成
张士兵 张云勇 张玉清 张钦宇 张朝阳 陈 巍 陈山枝
陈后金 范九伦 林金朝 欧阳缮 易东山 金 石 周一青
周武旻 周 亮 桂 冠 贾 焰 夏银水 袁东风 钱志鸿
倪国新 徐立中 郭 庆 郭 磊 郭渊博 黄 韬 黄建伟
黄梦醒 崔琪楣 梁永生 隆克平 普园媛 裴庆祺 谭晓衡

Shuguang Cui (美国) Yi Qian (美国) Shiping He (美国)

Jiangzhou Wang (英国) Wen Tong (加拿大)

收录声明

本刊对发表的文章,拥有出版电子版、网络版版权,并拥有和其他网站交换信息的权利。本刊支付的稿酬中已经包含上述费用。

Journal on Communications has the copyright to publish electronic edition, online edition of the published articles, and has the right to exchange information with other sites. The expenses have been included in the fee paid by editorial department.

道德声明

本刊发表的论文是作者独立取得的原创性研究成果,无一稿多投;论文内容不涉及国家机密;未曾以任何形式用任何文种在国内外公开发表过;论文内容不侵犯他人著作权和其他权利。若发生一稿多投、侵权、泄密等问题,论文作者将承担全部责任。

The authors of *Journal on Communications* guarantee that their submitted articles are original and contain nothing confidential. The said article is only submitted to *Journal on Communications*. The said article has not been published before and has not been submitted elsewhere for print or electronic publication consideration. The said article is no way whatever a violation or an infringement of any existing copyright or license from the third party. Otherwise, the authors of the said article shall take the blame for the violation or infringement of the related copyright and the leakage of secrets.

通信学报

Journal on Communications



发行代号：
国内2-676
国外M395

2022年6月25日出版 定价：98.00元

ISSN 1000-436X



9 771000 436229