ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 33,No. 5,Sep. 2018,pp. 769－778
DOI:10. 16337/j. 1004-9037. 2018. 05. 001

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

骨导麦克风语音盲增强技术研究现状及展望

张雄伟　郑昌艳　曹铁勇　杨吉斌　邢益搏

(陆军工程大学指挥控制工程学院,南京,210007)

摘　要:骨导麦克风是一种非声传感器,由于其语音传输通道天然屏蔽了周围环境噪声的影响,因而具有很强的抗噪性能,已在多种强噪声环境的语音通信中发挥重要作用。由于人体传导的低通性能以及传感器工艺水平的限制等,骨导语音听起来比较沉闷、不够清晰,增强骨导语音对进一步改善强噪声环境下的语音通信质量以及骨导产品的推广具有重要意义。骨导麦克风语音盲增强在语音增强阶段仅拥有骨导语音信息,相比于融合带噪气导语音的增强,这种直接的增强方式具有更广泛的应用前景。本文在分析骨导语音特点的基础上,梳理总结了无监督频谱扩展法、均衡法和谱包络转换法等3种骨导麦克风语音盲增强方法,并展望了骨导麦克风语音盲增强研究的发展方向。

关键词:骨导麦克风;语音盲增强;强背景噪声;非声传感器

中图分类号:TN912. 3　　**文献标志码**:A

Blind Enhancement of Bone-Conducted Microphone Speech: Review and Prospects

Zhang Xiongwei, Zheng Changyan, Cao Tieyong, Yang Jibin, Xing Yibo

(College of Command and Control Engineering, Army Engineering University of PLA, Nanjing, 210007, China)

Abstract: As one kind of non-acoustic sensors, bone-conducted microphone has excellent anti-noise characteristics because its speech transmission channel naturally shields the influence of ambient noise. Recently, bone-conducted microphone has gradually played an important role in speech communication systems in various strong noise environments. However, due to the low-pass characteristic of human body and the limitations of sensor technology, bone-conducted speech sounds muffled and unclear. It is of great significance to improve the bone-conducted speech for more efficient speech communication in strong noise environments and wider application prospects of bone-conducted products. The blind enhancement algorithm means only the bone-conducted speech information can be acquired during the speech enhancement stage. This kind of direct enhancement algorithm is more applicable than the fusion technique with air-conducted speech. Here, the characteristics of bone conducted speech are firstly analyzed, then the existing methods including unsupervised bandwidth extension, equalization and spectral envelope transformation are introduced. Finally, we share some views of future research prospects.

Key words: bone-conducted microphone; speech blind enhancement; strong background noise; non-acoustic sensor

---

基金项目:国家自然科学基金(61471394)资助项目。
收稿日期:2018-05-01;**修订日期**:2018-07-10

引 言

语音是人与人之间最自然和最便捷的交流方式，也是人机交互的一种重要方式。随着便携式通信设备的普及，随时随地的语音交流已成为可能，人们在享受语音交流便利的同时，也受困于各种各样的环境噪声带来的干扰与不适。为保证良好的语音通信质量，语音增强技术应运而生。语音增强的目的就是在保证语音可懂度的前提下，去除各类噪声干扰，提升人耳的听觉感受和人机交互质量。虽然现代的语音增强技术已取得了重大进展，但是在复杂强噪声环境下，现有的语音增强算法性能会大幅下降。因此，如何保证强噪声背景下良好的语音通信质量仍是研究者们亟待解决的问题。

骨导麦克风是一种非声传感器，人说话时声带振动会传递到喉头和头骨，这种麦克风正是通过采集这种振动信号并转换为电信号来获得语音(以下称为骨导语音)。与传统的空气传导麦克风语音(以下称为气导语音)不同，背景噪声很难对这类非声传感器产生影响，所以骨导语音从声源处就屏蔽了噪声，因此非常适用于强噪声环境下的语音通信。目前，许多国家在坦克、武装直升机等军事装备上都配备了基于骨导的通讯系统，美国的“未来战士”单兵作战系统中骨导耳机是其重要通信工具。在民用方面，美国 iASUS 公司针对赛车、摩托赛车等极限运动，研发了多款喉头麦克风、骨导耳机等设备，日本的松下、索尼等公司也研发出多种骨导通讯产品，并被应用到消防、特勤、矿山开采、公共交通、紧急救援等行业中。

虽然骨导语音能够有效抵抗环境噪声的干扰，但由于声音传输路径的变化和传感器工艺水平的限制，骨导语音与气导语音的声学特性存在一定差异。声学特性的差异使得骨导麦克风带给使用者的舒适度体验并不理想，这是阻碍其进一步推广应用的重要原因。另外，骨导语音中也会混入一些物理噪声，例如传感器与紧贴的皮肤产生的摩擦噪声、极限运动时强力的风力摩擦噪声、人咀嚼或牙齿相碰时引入的噪声等，这些噪声也降低了骨导语音的通信质量。因此，开展对骨导语音增强算法的研究，对进一步改善强噪声环境下的语音通信质量，进一步扩大骨导麦克风的应用范围，具有重要的理论意义和实用价值。

近年来，诸多国内外学者开展了与骨导语音相关的语音增强算法的研究。针对骨导语音增强问题，国防科技大学[1-2]、解放军理工大学[3]将骨导语音特点和传统的气导语音增强算法相结合，展开了深入的研究。基于多传感器融合的语音增强算法目前研究最多，例如，文献[4]开创性地提出了使用概率最佳滤波器(Probabilistic optimum filter, POF)映射带噪气导和骨导的混合语音来估计干净的声学语音特征，并将这种增强后的特征用于语音识别系统；文献[5]首先利用预先定义的均衡滤波器扩展骨导语音频谱，然后利用带噪的气导语音与均衡后骨导语音信息，基于最佳幅度和相位估计算法估计出纯净的语音特征；哈尔滨工业大学[6]、华南理工大学[7]均针对多传感器语音的融合进行了深入研究。这些融合性的增强算法在增强阶段必须同时具有骨导与气导语音信息，但在强噪声环境下，带噪气导语音可能完全不可用，并且一些传感器并未同时配置有骨导麦克风和气导麦克风，因此这类算法存在较大的应用局限性。

骨导语音盲增强(Blind enhancement)，也称为盲恢复[8](Blind restoration)，是指在增强阶段直接从已有的骨导语音中推断出类纯净气导语音信号，而不需要气导语音信息作为辅助。相比于融合性的增强算法，这种直接的增强算法由于拥有的信息少，因此增强难度更大。需要指出的是，现阶段的骨导语音盲增强研究中，主要关注如何从骨导语音中恢复出类纯净气导声学特征，而不考虑骨导语音中混入噪声的问题。目前，日本高等科学技术研究所、日本奈良科学技术研究所、印度国际信息技术研究所等，一直在骨导语音盲增强算法的研究上进行不懈努力。其中，日本高等科学技术研究所主要专注于基于线性预测系数特征转换的骨导语音盲增强算法，该研究所的工作包括从理论上分析骨导语

音盲增强的可行性、不同的语音分解合成模型在骨导语音盲增强中的优劣、线性预测系数特征的选取与改进以及特征转换模型的研究等;日本奈良科学技术研究所采用语音转换的技术,针对非声喃语(Non-audible murmur, NAM)麦克风语音盲增强算法展开了深入的研究;印度国际信息技术研究所针对喉头麦克风语音盲增强算法、语音盲增强的低速率编码等问题展开了一系列研究。国内高校中,西北工业大学[9]最早研究了骨导语音的重构技术,哈尔滨工业大学[6]针对头骨麦克风中传感器不同位置的骨导语音盲增强进行了相关研究。

总的来说,国内外针对骨导语音盲增强的算法研究较少,其中一个重要原因是骨导语音信息损失严重,传统的语音增强技术难以解决这种信息严重丢失的问题。随着人工智能深度学习技术的发展,从大量数据中准确学习数据的先验知识成为可能,这也为骨导语音盲增强算法提供了发展契机,已有研究人员[10, 11]开始尝试使用深度学习技术来解决这个问题。

本文在分析骨导语音信号的基础上,总结了现有的骨导语音盲增强算法,并指出了骨导语音盲增强算法今后可能的发展方向。

1 骨导语音信号分析

1.1 骨导语音的产生

语音的产生涉及到人体一系列器官和肌肉的运动,语音的产生过程中,肺部、胸腔等产生气流,气流经过声带控制形成了语音的原始激励信号,这种激励信号经过声道的调制,形成不同音调和内容的语音。

骨导语音与气导语音信号均由同一发声源产生,两者最大的不同在于声音传输路径的改变,如图 1 所示。气导语音是激励信号经过声道调制后,再经过口腔、鼻腔等辐射最终形成的语音。骨导语音则可看成激励信号经过人体内部骨头、组织等路径传输形成的语音。需指出的是,骨导语音传输路径因传感器放置的位置而改变,图 1 的传输路径为头骨顶部。

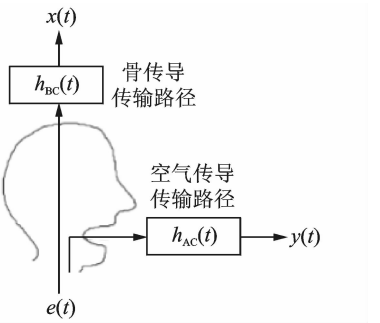

图 1 头骨麦克风语音与气导语音传输路径示意图

Fig. 1 Transmission channels for bone-conducted speech and air-conducted speech

将纯净气导语音 $y(t)$、骨导语音 $x(t)$的传输路径函数分别定义为 $h_{AC}(t)$、$h_{BC}(t)$,语音激励信号为 $e(t)$,骨导语音信号可以看成由纯净气导语音信号经过一个传输通道变换函数 $h(t)$得到

$$x(t)=e(t)*h_{BC}(t)=y(t)*h_{AC}^{-1}(t)*h_{BC}(t)=y(t)*h(t) \tag{1}$$

研究表明,传输通道变换函数 $h(t)$是非常复杂的非线性函数,不仅与骨导传感器的性能和传感器放置的位置有关,还与说话内容的发声音节、说话人的身体特性相关。

1.2 骨导麦克风的分类

骨导语音的特性与其非声传感器采集的身体振动部位有关,这里介绍几类常见的骨导麦克风产品。

(1)喉头麦克风

喉头麦克风是最早的骨导麦克风产品,它通过拾取说话人的脖颈部位的振动形成电声信号,不仅可用于强噪声环境下的语音获取,还可应用于因疾病而失声的患者,或者气管切开术后有暂时性语音丧失的患者。

(2)头戴式麦克风

头戴式麦克风通过拾取说话人头顶、额头、面颊、太阳穴、下颚、鼻翼等部位的振动形成电声信号。通常情况下,该类麦克风配置在头盔或者面罩中,也有针对特殊需求,将麦克风放置在头发或者发饰中。

(3)入耳式麦克风

入耳式麦克风,或者称为耳塞式麦克风,通过采集内耳的振动形成语音信号。通常入耳式麦克风采用一体化设计,不仅可采集语音信号,抵抗环境噪声干扰实现语音通信,同时也可兼具耳机的功能,尤其适合于听力受损者使用。由于可以利用耳机功能获取外部噪声的先验信息,因此这类麦克风可利用融合性的算法实现骨导语音增强。

(4)非声喃语麦克风

Nakajima 等[12]发明了 NAM 麦克风,这种麦克风通常放置在耳后位置。与抵抗环境噪声的骨导麦克风不同,NAM 麦克风主要是针对公开场合中语音通信的私密性需求研发,用于获取说话者的低声耳语。即使说话者轻声说话时,NAM 麦克风也可通过拾取耳后骨骼和组织的振动获取语音信号。

多项研究表明,骨导语音的清晰度与骨导麦克风的位置以及个体的身体特性有关。有研究表明[13-14],头戴式麦克风中,额头部位拾取的信号明显优于其他部位,其次为太阳穴和面颊,喉头麦克风语音的能量明显高于头骨麦克风语音,但是语音清晰度低于头骨麦克风语音。

1.3 骨导语音与气导语音的特征差异

骨导语音与气导语音存在明显的声学特征差异,这些差异使得骨导语音听起来较为沉闷、不够清晰。本节通过对比气导语音与喉头麦克风语音语谱图,具体阐述骨导语音与气导语音之间的特征差异。图 2(a,b)与图 2(c,d)分别为两个说话者使用同样的麦克风,录制的相对应的气导语音和喉头麦克风语音的语谱图。

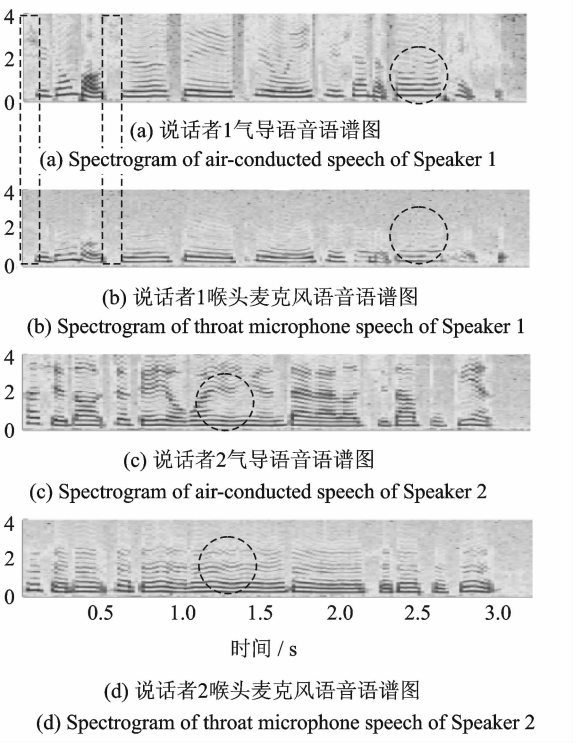

图 2 气导语音与喉头麦克风语音语谱图

Fig. 2 Spectrograms of throat microphone speech and air-conducted speech

与气导语音相比,骨导语音有如下典型的特征差异:

(1)高频衰减严重

人体传导具有低通性,语音信号到达骨导麦克风位置时,类似于经过了一个低通滤波器,高频成份显著衰减。高频成份衰减的程度与骨导麦克风的位置密切相关,例如当气导语音信号在 6 kHz 仍有能量时,额头位置采集的信号高频截止频率约为 3.5 kHz,喉头位置采集的信号高频截止频率约为 2 kHz。从图 2(b)与图 2(d)中可明显看出,喉头麦克风语音 2 kHz 以上的能量几乎已完全衰减。

(2)辅音音节损失

由于振动产生的电声信号不再经过或者不再全部经过口腔、鼻腔、唇等声音“调音”区域,与这些区域相关的摩擦音、爆破音、清音等辅音音节丢失严重。例如,喉头部位采集的语音,基本上已完全无辅音迹象,额头、鼻翼等采集部位,虽然能够采集到辅音,但由于能量小,仍存在辅音丢失的现象。从图 2(a,b)中的矩形窗可看出,喉头麦克风语音的辅音音节已丢失。

(3)中低频谐波能量改变

骨导语音在中低频段能保持良好的谐波结构,但是谐波的能量会发生明显改变,并且与不同人的身体传导特性密切相关。从图 2(a,b)圆形窗中可看出,喉头麦克风语音低频的谐波能量明显弱于气导语

音,而图 2(c,d)中,喉头麦克风语音的谐波能量却明显高于气导语音,这说明由于不同人的身体传导特性不尽相同,会导致骨导语音的中低频谐波能量的变化不同。

2 骨导语音盲增强方法

从 1.3 节对骨导语音特性分析中可以看出,骨导语音存在的最大问题是高频信息的严重丢失。与多传感器融合性的增强算法不同,骨导语音盲增强在增强阶段仅有骨导语音信息,因此,盲增强的难点在于基于有限的中低频带信息推断并恢复出高频信息。按照增强方法构建思路的不同,本节介绍 3 种典型的骨导语音盲增强方法。

2.1 无监督频谱扩展法

无监督频谱扩展法认为骨导语音与气导语音具有一致的共振峰结构,或者语音的低频与高频具有一致的谐波结构。利用这种结构特性,可直接对骨导语音的低频频谱进行扩展,得到高频的共振峰或者谐波结构,从而实现语音增强。

无监督频谱扩展法的框图如图 3 所示。文献[15]认为骨导语音在高频仍然有相同的共振峰结构,只是共振峰的振幅衰减严重,通过对 LP 系数极点的规整,实现了骨导语音的高频共振峰振幅的加强。文献[16]认为语音低频与语音高频之间具有一致的谐波结构,通过对骨导语音激励信号的处理,获得了高频的奇次谐波结构。

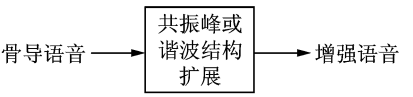

图 3 无监督频谱扩展法

Fig. 3 Unsupervised spectrum extension method

无监督频谱扩展法根据语音的结构性特点,改善了骨导语音的高频成份,在一定程度上可提升人耳的听觉感受。由于这种方法不需要先验知识,具有算法简单、运算量小的特点,因此在现有的骨导电子设备中已得到应用。但是,该方法基于语音结构一致性的假设,并不完全符合现实的语音特性情况,例如文献[15]的方法仅适用于频带较宽的骨导语音,如额头振动采集的骨导语音信号。频带较窄的骨导语音信号,如喉头麦克风语音,基本上不存在高频共振峰结构,因此难以采用共振峰结构扩展的方法实现增强,而文献[16]的方法无法恢复不具有谐波结构的摩擦音、爆破音等音素。

2.2 均衡法

2.1 节分析了骨导语音的产生过程,指出了骨导语音信号可以看成由纯净气导语音信号经过一个传输通道变换函数 $h(t)$ 得到。均衡法的思想是找到传输通道变换函数 $h(t)$ 的逆变换函数 $g(t)$,从骨导语音信号中恢复出气导语音信号。根据图 1,我们可推断出气导语音的恢复过程如图 4 所示。其中,$x(t)$ 表示骨导语音信号,$y(t)$ 表示气导语音信号,$h_{BC}(t)$ 为骨导语音传输通道函数,$h_{AC}(t)$ 为气导语音传输通道函数,$g(t)$ 表示由骨导语音到气导语音传输通道的变换函数,对骨导语音的盲增强可表示为

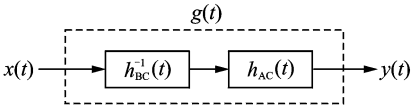

图 4 从骨导语音恢复气导语音

Fig. 4 Recovering air-conducted speech from bone-conducted speech

$$y(t)=x(t)*h_{BC}^{-1}(t)*h_{AC}(t)=x(t)*g(t) \tag{2}$$

均衡法首先由 Shimamura[16] 提出,通过建模 $g(t)$ 并构造出逆滤波器实现骨导语音增强。实际实现中,可根据提取的不同的语音特征建立不同的 $g(t)$ 模型。将时域信号变换到 z 域,则式(2)变换为

$$Y(Z)=X(Z)\cdot G(Z) \tag{3}$$

可得

$$G(Z)=\frac{Y(Z)}{X(Z)} \tag{4}$$

式中 $G(Z)$ 称为均衡滤波器，可看成骨导语音与纯净气导语音信号特征之间的传递特性的建模。这种均衡滤波器的思想，在多传感器语音融合的语音增强算法中，得到了广泛应用。例如，文献[4,5]分别利用带噪气导语音与骨导语音的线性预测倒谱系数（Linear prediction cepstrum coefficient，LPCC）和功率谱密度计算均衡滤波器。在骨导语音盲增强中，由于在增强阶段缺乏带噪气导语音信号，因此需要利用先验的骨导与纯净气导语音数据，训练得到均衡滤波器的系数，再将待增强的骨导语音通过构造好的均衡滤波器得到增强的语音信号。基于均衡法的骨导语音盲增强框架如图 5 所示。

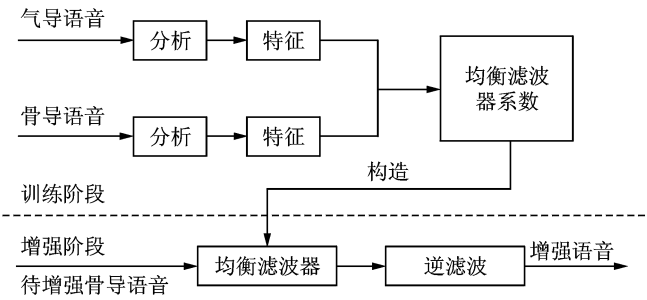

图 5 基于均衡法的骨导语音盲增强框架

Fig. 5 Framework of blind enhancement for bone-conducted speech based on the equalization method

现有的骨导语音盲增强算法，均是选择幅度谱作为均衡滤波器系数的训练对象。设气导语音信号与骨导语音信号的幅度谱分别表示为 $S_{AC}(f)$ 与 $S_{BC}(f)$，文献[17]选取了对应的骨导与气导语音长时幅度谱得到估计的均衡滤波器系数，即 $\hat{G}(f)=\dfrac{\mathrm{Long}(S_{AC}(f))}{\mathrm{Long}(S_{BC}(f))}$，Long 指长时谱；文献[18]在文献[17]基础上进行了改进，选取对应的骨导与气导语音短时幅度谱得到均衡滤波器，即 $\hat{G}(f)=\dfrac{\mathrm{Short}(S_{AC}(f))}{\mathrm{Short}(S_{BC}(f))}$，short 指短时谱，并对滤波器系数进行了平滑处理。

当 $\hat{G}(f)$ 估计较为准确时，这种乘性的特征变换不仅可以保持原有信号的结构特点，并且能有效调节骨导语音与气导语音不匹配的时频点能量。但是现有的骨导语音均衡法尚存在两方面的不足：①对于语音的高频成份恢复较难，因为当骨导语音高频能量几乎为 0 时，$\hat{G}(f)$ 在高频的响应即使再大，也很难起到能量提升作用；②文献中设计的均衡滤波器 $\hat{G}(f)$ 是固定不变的，事实上，$G(f)$ 会随着骨导语音的变化而变化。因此，设计时变的均衡滤波器，是提升基于均衡法的骨导语音盲增强算法效果的重要途径。

2.3 谱包络转换法

目前，大多数的骨导语音盲增强采用基于谱包络转换的方法。谱包络转换法基于语音的源-滤波器模型，该模型将语音视为声音激励信号（源）经过声道（滤波器）调制得到，这种声道特点通常由谱包络特征表示。由于骨导语音与气导语音来源于同一声源，激励信号近似相同，那么只需转换谱包络特征即可得到类气导语音信号，实现语音增强。

令 $x(t)$，$y(t)$ 分别为骨导和纯净气导语音信号，则语音信号可表示为

$$x(t)=e_{BC}(t)*s_{BC}(t) \tag{5}$$

$$y(t)=e_{AC}(t)*s_{AC}(t) \tag{6}$$

式中，$e_{BC}(t)$，$e_{AC}(t)$ 分别表示骨导语音与气导语音的激励，$s_{BC}(t)$，$s_{AC}(t)$ 分别表示骨导语音与气导语音的谱包络，$*$ 表示卷积操作。

基于谱包络转换法的骨导语音盲增强算法通常包括训练阶段和增强阶段，其典型框架如图 6 所示。在训练阶段，骨导语音与气导语音数据经过分析合成模型，抽取出激励特征和谱包络特征，通过训练构建骨导语音到气导语音的谱包络特征之间的转换模型 $f(x)$，$f(x)$ 为复杂非线性函数。在增强阶段，首

先提取待增强语音的激励特征 $e_{BC}(t)$和谱包络特征 $s_{BC}(t)$,然后可利用训练好的模型从骨导语音谱包络特征中估计出类气导语音谱包络特征 $\hat{s}_{AC}(t)$

$$\hat{s}_{AC}(t)=f(s_{BC}(t)) \tag{7}$$

由于骨导与气导语音的激励信号近似相同,可直接将骨导语音激励信号作为估计的类气导语音激励信号

$$\hat{e}_{AC}(t)=e_{BC}(t) \tag{8}$$

根据估计出的谱包络和骨导语音原始的激励特征合成出增强的语音。

$$\hat{y}(t)=\hat{e}_{AC}(t)*\hat{s}_{AC}(t) \tag{9}$$

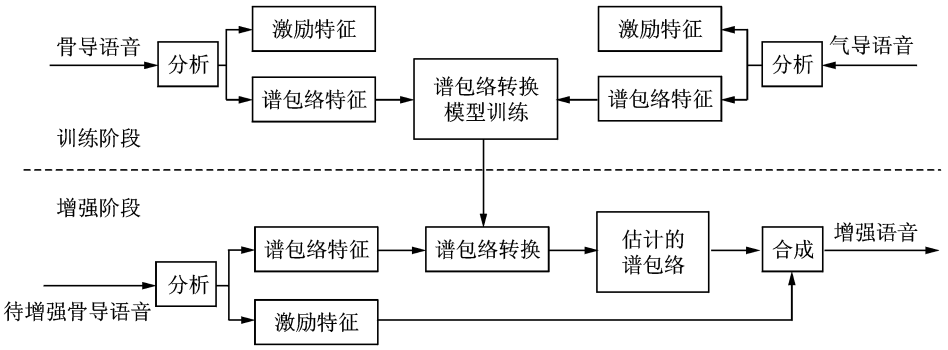

图 6 典型谱包络转换法框架

Fig. 6 Typical framework for spectral envelope transformation based methods

谱包络转换法中的关键是如何选择分析合成模型、谱包络特征以及谱包络转换模型。下面按照以下 3 个关键问题进行介绍。

(1)分析合成模型

最初的骨导语音盲增强采用调制传递函数(Modulation transfer function, MTF)[19]将语音信号分解为不同频带下的激励信号和能量包络信号。文献[20]证明了采用基于线性预测系数的 LPC 模型,相比于 MTF 可获取更好的分解合成效果。文献[21]针对 NAM 语音,利用 STRAIGHT 分解合成模型,将语音分解为平滑后的 STRAIGHT 谱、基音频率和非周期分量,实验证明,该模型可获得更高的语音质量,缺点是计算量较大。

(2)谱包络特征

文献[20]探讨了基于 LP 分解合成的可行性,基于文献[20]的研究,开始出现大量的有关 LP 系数表示谱包络特征的研究,文献[22]直接提取 LPC 特征作为谱包络特征,文献[23]认为加权的 LPCC 系数更优于 LPC 系数,文献[24]经过实验,认为 LSF 参数具有更好的量化特性和插值特性鲁棒性,文献[10,25]在此基础上均采用 LSF 参数特征,文献[26]利用广义梅尔倒谱系数(Mel generalized cepstral coefficients, MGCC)以获得更好的人耳听觉感知,文献[21,27]利用 STRAIGHT 谱得到的多帧联合的 MFCC 参数作为谱包络特征。

(3)谱包络转换模型

谱包络转换模型的选取是谱包络转换法最为关键的问题。与语音转换、频谱扩展中谱转换技术的发展十分类似,谱包络转换模型也是由浅层、线性的模型向深层、复杂非线性模型发展。文献[28]建立了骨导语音与气导语音加权的 LPCC 系数的码本,采用码本映射的方法实现加权 LPCC 特征的转换,在码本映射的方法中,由于矢量量化的原因,会得到估计不连续的声学特征。文献[25,29]采用 GMM 模

型实现 LSF、MFCC 谱包络转换，相比于码本映射，GMM 模型可以实现特征的连续估计，因此采用 GMM 模型映射的特征，可获得更小的谱失真。文献[23]首次采用简单前馈神经网络模型建模骨导语音与气导语音加权的 LPCC 特征之间的转换关系，文献[8]在文献[23]基础上进一步采用更适用于语音数据建模的递归神经网络，文献[10]则利用深度神经网络建模骨导与气导语音的 LSF 参数间的关系。

在典型的谱包络转换方法框架下，有研究[29-30]认为，传统的源-滤波器语音分解合成模型（如 LPC 模型）假设了激励特征与谱包络特征相互独立，实际情况中，两者的独立性并不能完全得到满足。因此，在建立骨导语音与气导语音谱包络特征之间的转换关系基础上，激励特征之间转换关系的建立是值得研究的方向。相比于谱包络特征，激励特征映射建模更为困难，因为谱包络特征对应为声道特征，声道形状的变化具有缓变性，因此声道特征具有短时平稳性，而激励特征的本质是高度非线性的，其在一个基音周期内都可能快速变换[30]。目前有关骨导语音与气导语音激励特征建模的文献较少，文献[30]将语音分析窗口固定在声门闭合时刻周围的区域，计算该区域的希尔伯特包络作为激励特征，并采用神经网络映射这种激励特征，文献[29]将激励信号的频谱包络差作为频谱倾斜矢量（Spectral tilt vector）进行建模，并提出了基于 GMM 模型的音素相关频谱倾斜矢量映射方案。

除了利用基于源-滤波模型的分解合成方法，文献[11]采用基于信号模型的分解合成方法，将语音信号分为高维幅度谱与相位，并利用深度学习技术，建立骨导与纯净气导语音高维幅度谱之间的对应关系。

3 总结与展望

本文针对骨导语音盲增强技术进行了综述，首先对骨导语音的产生及其语音特点进行了分析，在此基础上，梳理总结了现有的 3 类骨导语音盲增强算法。由于骨导语音盲增强算法在语音增强阶段拥有的信息少，因此如何有效地从先验知识中学习数据的特点，推断出缺失的信息，是骨导语音盲增强面临的重要问题。由于骨导语音与说话人的身体特性、传感器放置位置及性能密切相关，实现普适性的骨导语音盲增强算法具有相当大的难度。本文认为未来针对骨导语音盲增强算法可能集中在以下几个方面：

（1）高维谱特征的转换

现有的骨导语音盲增强算法，大多集中在低维的谱包络特征转换基础上，难以进一步提升语音增强的质量。目前基于深度学习的高维特征转换方法已在语音去噪、语音转换、语音频谱扩展中得到应用，文献[11]已开始尝试利用深度学习的方法实现骨导语音高维谱特征的转换。利用深度学习实现高维特征转换，或将成为提升骨导语音盲增强效果的重要研究方向。

（2）不基于特定说话人

骨导语音的特性与说话人的身体传导特性密切相关，这对骨导语音盲增强算法的通用性提出了巨大的挑战。现有的盲增强算法均集中在特定说话人骨导语音增强上，在现实中，这种基于特定说话人的增强算法需要“定制”使用。研究不基于特定说话人的骨导语音盲增强技术，将有助于推动骨导通信设备的广泛使用。

（3）与语音编码结合

骨导语音增强的落脚点是实现基于骨导麦克风的语音通信，而在语音通信中，语音编码是必须考虑的问题，语音增强中过大的计算量以及过多的资源占用，在现实的语音通信中并不可行，研究如何实现骨导语音增强算法与语音编码的有效结合是骨导语音增强算法实用化进程的必经之路。

（4）人体语音传导理论研究

目前针对骨导语音特性的理论研究，大都集中在骨导传感器位置对骨导语音质量的影响上，骨导语音特性与人体传导特性相关性的理论研究较少，导致增强算法缺乏有效的理论指导，严重受制于先验数

据的数量与质量,算法的泛化性能、通用性能难以取得突破。通过探索人体的语音传导特性,可以为骨导语音增强算法的发展提供有力支撑。

参考文献:

[1] 肖新华.面向骨传导的语音消噪算法及硬件实现技术研究[D].长沙:国防科学技术大学,2009.
Xiao Xinhua. Research on the hardware technique of algorithm in noise reduction for boneknocker[D]. Changsha: National University of Defense Technology, 2009.

[2] 姚利俊.骨传导语音增强算法及其可重构硬件结构研究与实现[D].长沙:国防科学技术大学,2013.
Yao Lijun. Research and implementation on reconfigurable hardware architecture of speech enhancement algorithm for boneconduction[D]. Changsha: National University of Defense Technology, 2013.

[3] 乔林.面向骨导传感器语音的单通道语音增强算法研究[D].南京:解放军理工大学,2017.
Qiao Lin. Research on single channel speech enhancement algorithm for bone-conducted Speech [D]. Nanjing: PLA University of Science and Technology, 2017.

[4] Graciarena M, Franco H, Sonmez K, et al. Combining standard and throat microphones for robust speech recognition[J]. Signal Processing Letters IEEE, 2003,10(3):72-74.

[5] Dekens T, Verhelst W. Body conducted speech enhancement by equalization and signal fusion[J]. IEEE Transactions on Audio Speech & Language Processing, 2013,21(12):2481-2492.

[6] 李敏杰.骨导和气导结合的语音增强系统搭建[D].哈尔滨:哈尔滨工业大学,2016.
Li Minjie. Bone conduction combined with air conduction to structure speech enhancement system [D]. Harbin: Harbin Institute of Technology, 2016.

[7] 朱颖莉. 基于多传感器的语音增强技术研究[D].广州:华南理工大学,2013.
Zhu Yingli. Research on speech enhancement technology based on multi-sensors[D]. Guangzhou: South China University of Technology, 2013.

[8] Thang T, Masashi U, Masato A. A blind restoration model for bone-conducted speech based on a linear prediction scheme [C]//International Symposium on Nonlinear Theory and its Applications. [S. l.]: [s. n.], 2007.

[9] 李静.基于骨导信号的语音重构技术[D].西安:西北工业大学,2004.
Li Jing. Speech reconstruction technology based on bone conduction signals [D]. Xi'an: Northwestern Polytechnical University, 2004.

[10] Huang B, Gong Y, Sun J, et al. A wearable bone-conducted speech enhancement system for strong background noises[C]// International Conference on Electronic Packaging Technology. Harbin, China: [s. n.], 2017:1682-1684.

[11] Watanabe D, Sugiura Y, Shimamura T. Speech enhancement for bone-conducted speech based on low-order cepstrum restoration[C]//Intelligent Signal Processing and Communication Systems (ISPACS). Xiamen, China:[s. n.], 2017.

[12] Nakajima Y, Kashioka H, Shikano K, et al. Non-audible murmur recognition input interface using stethoscopic microphone attached to the skin[C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hong Kong, China:[s. n.], 2003.

[13] Tran P, Letowski T, Mcbride M. The effect of bone conduction microphone placement on intensity and spectrum of transmitted speech items[J]. Journal of the Acoustical Society of America, 2013,133(6):3900-3908.

[14] Pollard K, Tran P, Letowski T. The effect of vocal and demographic traits on speech intelligibility over bone conduction[J]. Journal of the Acoustical Society of America, 2015,137(4):2060-2069.

[15] Rahman M, Shimamura T. Intelligibility enhancement of bone conducted speech by an analysis-synthesis method[J]. Midwest Symposium on Circuits & Systems, 2011,47(10):1-4.

[16] Bouserhal R, Falk T, Voix J. In-ear microphone speech quality enhancement via adaptive filtering and artificial bandwidth extension[J]. Journal of the Acoustical Society of America, 2017,141(3):1321.

[17] Shimamura T, Tamiya T. A reconstruction filter for bone-conducted speech[C]//Midwest Symposium on Circuits and Systems Covington. USA:[s. n.], 2005:1847-1850.

[18] Kondo K, Fujita T, Nakagawa K. On equalization of bone conducted speech for improved speech quality [C]//IEEE International Symposium on Signal Processing and Information Technology. Vancouver, Canada: IEEE, 2007:426-431.

[19] Unoki M. A study on a bone conducted speech restoration with the modulation transfer function[J]. Phyciol Acoust, 2005, 3(35):191-196.

[20] Thang T. A study on restoration of bone-conducted speech with MTF-Based and LP-Based models, Special Issue on Nonlinear Circuits and Signal Processing[J]. Journal of Signal Processing, 2006, 86(5): 1421-1425.

[21] Toda T, Nakagiri M, Shikano K. Statistical voice conversion techniques for body-conducted unvoiced speech enhancement [J]. IEEE Transactions on Audio Speech & Language Processing, 2012, 20(9): 2505-2517.

[22] Tat V, Unoki M, Akagi M. A study on the LP-based blind model in restoring bone conducted speech[J]. IEICE Technical Report, 2008, 107: 17-22.

[23] Shahina A, Yegnanarayana B. Mapping speech spectra from throat microphone to close-speaking microphone: A neural network approach[J]. EURASIP Journal on Advances in Signal Processing, 2007(1): 1-10.

[24] Phungng T N, Unoki M, Akagi M. A study on restoration of bone-conducted speech in noisy environments with LP-based model and Gaussian mixture model[J]. Journal of Signal Processing, 2012, 16(5): 409-417.

[25] Turan M, Erzin E. Source and filter estimation for throat-microphone speech enhancement[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2016, 24(2): 265-275.

[26] Vijayan K. Comparative study of spectral mapping techniques for enhancement of throat microphone speech[C]//Twentieth National Communications Conference (NCC) Kanpur, India: [s. n.], 2014.

[27] Toda T. Statistical approaches to enhancement of body-conducted speech detected with non-audible murmur microphone [C]//International Conference on Complex Medical Engineering. Kobe, Japan: [s. n.], 2012: 623-628.

[28] Murty K, Khurana S S, Itankar Y, et al. Efficient representation of throat microphone speech[C]//Conference of the International Speech Communication Association, INTERSPEECH. Brisbane, Australia: [s. n.], 2008: 2610-2613.

[29] Turan T. A new statistical excitation mapping for enhancement of throat microphone recordings[C]//Conference of the International Speech Communication Association, INTERSPEECH. Lyon, France: [s. n.], 2013.

[30] Medabalimi A, Mallidi S, Yegnanarayana B. Speaker-dependent mapping of source and system features for enhancement of throat microphone speech[C]//Conference of the International Speech Communication Association, INTERSPEECH. Makuhari, Chiba, Japan: [s. n.], 2010: 985-988.

作者简介：

张雄伟(1965-)，男，教授，博士生导师，研究方向：语音与图像处理、多媒体信息处理，E-mail：xwZhang9898@163.com。

郑昌艳(1990-)，女，博士研究生，研究方向：语音处理、深度学习，E-mail：echoaimaomao@163.com。

曹铁勇(1971-)，男，教授，研究方向：智能信息处理、图像处理，E-mail：cty_ice@sina.com。

杨吉斌(1978-)，男，副教授，研究方向：语音信号处理、深度学习，E-mail：yjbice@sina.com。

邢益搏(1994-)，男，硕士研究生，研究方向：语音增强、智能信息处理，E-mail：18252059100@163.com。

（编辑：张　彤）