

采用压缩感知的麦克风阵列远场声源方位估计

曾帆, 黄惠祥, 童峰

(厦门大学 水声通信与海洋信息技术教育部重点实验室, 福建 厦门 361100)

摘要: 传统 CS-DOA 算法采用事先测量获得的声传播冲激响应组成稀疏恢复方程中的混合矩阵, 远场条件下测量冲激响应误差增大, 导致性能下降。针对该问题, 将房间冲激响应(RIP)分解为时延部分和混响部分, 并在稀疏恢复方程中将混响部分移至方向矢量项, 表明: 通过频点叠加的方式, 在远场条件下可直接利用阵列空间关系构造冲激响应组成混合矩阵, 进行远场压缩感知声源方位估计。实验结果表明: 与相位变换加权的可控响应功率和延时-累加定位算法相比, 在远场条件下该算法具有更高的方位估计分辨率。

关键词: 压缩感知; 麦克风阵列; 方位估计; 远场

本文引用格式: 曾帆, 黄惠祥, 童峰. 采用压缩感知的麦克风阵列远场声源方位估计[J]. 兵器装备工程学报, 2018, 39(5): 134-138.

Citation format: ZENG Fan, HUANG Huixiang, TONG Feng. Far Field Compressed Sensing Microphone Array DOA Estimation[J]. Journal of Ordnance Equipment Engineering, 2018, 39(5): 134-138.

中图分类号: TN912.3

文献标识码: A

文章编号: 2096-2304(2018)05-0134-05

Far Field Compressed Sensing Microphone Array DOA Estimation

ZENG Fan, HUANG Huixiang, TONG Feng

(Key Laboratory of Underwater Acoustic Communication and Marine Information Technology
of the Ministry of Education, Xiamen University, Xiamen 361100, China)

Abstract: The traditional CS-DOA algorithm needs to measure the acoustic propagation impulse response in advance to act as mixture matrix of sparse recovery equation, which leads to substantial performance degradation under far field as the error of response measurement increases. In this paper, the room impulse response is firstly decomposed into the delay part and the reverberation part, then the latter is moved to the orientation vector in the sparse recovery equation. By the way of frequency domain accumulation, we can use the array spatial relationship to directly construct the impulse response to make up mixture matrix, thus enable far field compressed sensing DOA. The experimental results showed that, compared with steered response power-phase transform and Delayed-sum algorithm, the proposed algorithm has a higher DOA resolution performance under the far field condition.

Key words: compressed sensing; microphone array; DOA; far field

收稿日期: 2017-12-15; 修回日期: 2018-01-10

基金项目: 国家自然科学基金(11574258); 福建省高校产学研合作项目(2015H6019); 福建省中青年教育科研项目(JAS170012)

作者简介: 曾帆(1992—), 女, 硕士研究生, 主要从事阵列信号处理研究。

通讯作者: 童峰(1973—), 男, 教授, 博士生导师, 主要从事声信号处理、阵列信号处理研究。

基于麦克风阵列的声源方位估计技术在音/视频会议、语音增强、数字助听器、狙击手定位^[1]、车辆/直升机定位^[2]等领域应用十分广泛,而狙击手、车辆/直升机定位等远场应用场景给声源方位估计带来更大的挑战。

传统麦克风阵列声源定位技术主要分为3类:基于最大输出功率的可控波束形成的声源定位技术^[3]、基于高分辨率谱估计的声源定位技术^[4]、基于时延估计的声源定位技术^[5]。

广义互相关方法(GCC)是传统的时延估计方法,该方法易于实现,但在低信噪比或者混响环境中性能急剧降低。基于高分辨率谱估计的声源定位方法一般用于处理窄带信号,对于宽带自然语音信号应用受限。相位变换加权的可控响应功率(SRP-PHAT)和延时-累加(DS)声源方位估计均属于波束形成的声源定位技术,其中SRP-PHAT结合了可控响应功率和相位变换加权的优点^[6,7],DS原理简单,运算量小,但是当混响和噪声较强时,传统的声源定位算法性能降低,无法正确估计声源位置。

当信号是稀疏或者可压缩时,可线性投影到低维空间,并通过一定的重构算法对信号进行CS重建^[8]。考虑到声源方向向量具有稀疏性,众多学者在压缩感知DOA估计理论和实际应用方面展开了广泛研究。文献[9]提出了一种时域压缩采样的DOA估计算法。文献[10]提出了基于压缩感知的单快拍海底声源DOA估计。文献[11]将基于CS的DOA估计应用于声矢量阵列。文献[12]将压缩感知DOA估计应用于卫星干扰源定位。文献[13]将压缩感知理论用于麦克风阵列声源定位。

文献[13]建立了一种室内混响环境中麦克风阵列接收信号的稀疏表示模型,以声源可能存在位置的冲激响应作为特征构建字典,利用正交匹配追踪(OMP)算法重建信号,最后通过搜索矢量最大元素对应的空间位置确定声源位置。但该文仅进行仿真实验,仿真结果表明在高混响低信噪比环境下,CS-OMP定位精度明显优于SRP-PHAT。同时,该方法需要事先测量各声源可能存在方位的冲激响应,实际推广应用。文献[14]在混响和噪声背景下,通过对延迟相关矩阵进行自适应特征分解,从其零空间中辨识出冲激响应。

在上述工作基础上,本研究首先将稀疏恢复方程混合矩阵中的房间冲激响应(RIP, room impulse response)^[15]分解为时延部分和混响部分,并将其中混响部分移动至方向向量中,表明:通过频域叠加抑制混响部分影响,在远场条件下可直接利用阵列阵元间时延关系构造观测矩阵,并进行了室内远场条件下声源方位估计实测实验,对SRP-PHAT、DS、CS-OMP、FCS-OMP在不同信噪比下的方位估计性能进行对比和评估。

1 算法

在远场环境中,麦克风阵列接收到的信号可用声源信号和声传播冲激响应的卷积结果与环境噪声之和表示,即 M

元线性麦克风阵列,阵元间距为 d ,第 m 个阵元接收到的信号 $x_m(n)$ 为

$$x_m(n) = h_m(r_s, n) * s(n) + w_m(n) \quad m = 1, 2, \dots, M \quad (1)$$

式中: n 为离散时间序列; r_s 为声源位置; $s(n)$ 为声源信号; $w_m(n)$ 为环境噪声信号; $h_m(r_s, n)$ 为 r_s 到第 m 个阵元的房间冲激响应RIP。在 $x_m(n)$ 中 $h_m(r_s, n)$ 需事先测量,给实际应用带来不便。

通过将RIP分解为时延部分和混响部分,则式(1)可写为

$$x_m(n) = hd_m(r_s, n) * hr_m(r_s, n) * s(n) + w_m(n) \quad (2)$$

式中: $hd_m(r_s, n)$ 为时延部分, $hr_m(r_s, n)$ 为混响部分;以线性麦克风阵列为例,即

$$hd_m(r_s, n) = \delta[n - \frac{(m-1)d \cos \theta}{c}] \quad m = 1, 2, \dots, M \quad (3)$$

式中: θ 为信号入射方向与阵列夹角; c 为声速; f_s 为采样率^[16-17]。则阵元接收信号 $x_m(n)$ 经加窗和离散傅里叶变换后,若冲激响应长度远小于窗函数长度,可得到接收信号的频域表示,将频域模型用矢量表示为:

$$X(k) = H(r_s, k) S(k) + W(k) = Hd(r_s, k) Hr(r_s, k) S(k) + W(k) \quad (4)$$

式中 $X(k) = \{X_1(k), X_2(k), \dots, X_M(k)\}^T$ 为阵元接收信号的频域矢量, $H(r_s, k) = \{H_1(r_s, k), H_2(r_s, k), \dots, H_M(r_s, k)\}^T$ 为 r_s 处的频域传输响应矢量, $Hd(r_s, k) = \{Hd_1(r_s, k), Hd_2(r_s, k), \dots, Hd_M(r_s, k)\}^T$ 为 r_s 处RIP中时延部分的频域矢量, $Hr(r_s, k) = \{Hr_1(r_s, k), Hr_2(r_s, k), \dots, Hr_M(r_s, k)\}^T$ 为 r_s 处RIP中混响部分的频域矢量, $W(k) = \{W_1(k), W_2(k), \dots, W_M(k)\}^T$ 为环境噪声信号的频域矢量。

以声源可能存在位置的RIP作为特征构建字典 $D(k)$,则式(4)可改写为

$$X(k) = D(k) \bar{S}(k) + W(k) = Dd(k) [Hr \bar{S}(k)] + W(k) \quad (5)$$

式中 $\bar{S}(k) = \{\bar{S}_1(k), \bar{S}_2(k), \dots, \bar{S}_I(k)\}^T$ 为以声源可能存在位置RIP的时延部分作为特征构建字典 $Dd(k)$ 下扩展的频域声源信号矢量。由于目标声源个数远小于声源可能存在的位置个数 I 。在实际应用中,冲激响应长度远大于窗函数长度,所以式(5)只能近似相等,可改写为:

$$X(k) \approx Dd(k) S'(k) + W(k) \quad (6)$$

式中:

$$S'(k) = \{Hr_1(k) \bar{S}_1(k), Hr_2(k) \bar{S}_2(k), \dots, Hr_I(k) \bar{S}_I(k)\}^T$$

由于矢量 $S'(k)$ 由矢量 $\bar{S}(k)$ 及 $Hr(k)$ 组合而成, $\bar{S}(k)$ 在空域下具有稀疏性,混响部分 $Hr(k)$ 则具有大量反射造成的随机频率选择性衰落特性,因而无法确保 $\bar{S}(k)$ 的空间稀疏性。考虑到这个问题,本文利用多个频点进行叠加以随机抑制频率选择性衰落特性对 $\bar{S}(k)$ 稀疏性的影响,从而保证在叠加意义上 $\bar{S}(k)$ 具有空间稀疏性,采用稀疏恢复方法求解方程(6)。

在频点叠加中,取 K 个采样点作为 FFT 长度,则可获得 $K/2$ 个频点的接收信号矢量 $X(k)$ 。同时,考虑到语音信号大多集中于低频部分,能量较大,可将每帧信号的各频点能量按照从大到小的顺序排列,设置阈值 η ($0 < \eta \leq 100$),取前 $\eta\%$ 个频点,通过 OMP 算法重构获得 $\eta\%K/2$ 个 $\bar{S}(k)$ 并将其相加得

$$\bar{S} = \sum_{k \in \Omega} \bar{S}(k) \quad (7)$$

Ω 为 $\eta\%K/2$ 个频点的集合 $\bar{S} = \{\bar{S}_1, \bar{S}_2, \dots, \bar{S}_I\}^T$ 。1 维矢量 \bar{S} 中最大元素所对应的空间位置即为声源的估计位置,即:

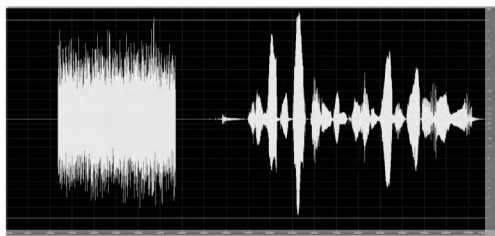
$$\hat{r}_s = r_{\arg \max_i \bar{S}_i} \quad (8)$$

2 实验

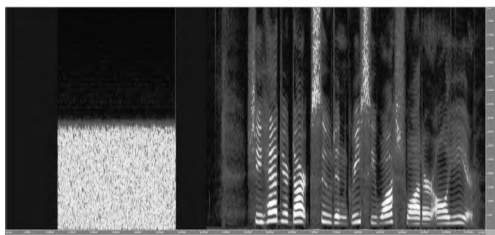
2.1 实验设置

实验在大小为 $30 \text{ m} \times 20 \text{ m} \times 6 \text{ m}$ 的实验室大厅进行,房间混响时间 T_{60} 用赛宾公式^[18] 近似估算约为 2.4 s 。麦克风阵列为阵元间距 13.5 cm 的 4 元均匀线阵,语音信号为 TIMIT 标准语音信号^[19],由 Marshall 蓝牙音箱播放,采样率为 16 kHz 。阵列与声源处于同一水平线且距离声源处 10 m ,将麦阵正前方 $0^\circ \sim 180^\circ$ 区域离散为 13 个空间位置即每隔 15° 作为声源可能存在的位置。

声源信号分别采用频率范围为 $0 \sim 4 \text{ kHz}$ 的白噪声和频率范围为 $300 \sim 3400 \text{ Hz}$ 的 TIMIT 标准语音信号,信号时域/时频图见图 1。原始阵列接收到的白噪声和语音信号信噪比分别为 11.5 dB 和 10.5 dB ,通过对原始阵列接收信号叠加实录噪声构造不同信噪比下的阵列接收信号,用于比较 SRP-PHAT、DS、CS-OMP 及本文 FCS-OMP 方法在不同信噪比下的声源方位估计性能。其中,通过实测获得声源各可能方位到麦克风阵列的信道冲激响应以用于 CS-OMP 方法构造稀疏恢复方程。各算法实验参数设置如表 1 所示。



(a) 信号时域图



(b) 信号时频图

图 1 声源信号时域/时频图

表 1 实验参数

采样率	16 kHz
信号处理帧长	128
加窗	汉明窗
OMP 算法频点阈值	30

2.2 实验结果与分析

实验结果如图 2、图 3、图 4 所示。图 2、图 3 分别是当声源为白噪声和语音信号时不同信噪比下, FCS-OMP、CS-OMP、DS、SRP-PHAT 将声源分别位于 13 个方向的指向性结果按照不同方向进行调整对齐累加后的定位指向性结果,图 4 是当语音信号 $\text{SNR} = 2.5 \text{ dB}$ 时, FCS-OMP 与 CS-OMP 在不同频点数下的性能比较。

图 2(a) 是在原始信噪比条件下各算法对远场白噪声源的方位估计结果,图 2(b)、图 2(c) 是信噪比降低条件下,各算法对远场白噪声源的方位估计结果。从图 2(a) 可以看出, FCS-OMP 和 CS-OMP 曲线尖锐程度几乎重叠,且明显高于 SRP-PHAT 和 DS。随着信噪比降低, FCS-OMP 曲线的尖锐程度逐渐高于 CS-OMP,且仍高于 SRP-PHAT 和 DS。

图 3 给出了各算法对远场语音声源在不同信噪比条件下的方位估计结果。由图 3(a)、图 3(b) 可见,当语音信号信噪比为 10.5 dB 和 4.9 dB 时, FCS-OMP 的曲线尖锐程度明显高于 SRP-PHAT 和 DS,但低于 CS-OMP,但在图 3(c) 中,当 $\text{SNR} = 2.5 \text{ dB}$ 时, FCS-OMP 曲线的尖锐程度高于其他 3 种算法,这是由于在远场低信噪比条件下测量获得的冲激响应质量下降,使算法性能降低,但本文利用阵列空间关系构造冲激响应能很好地解决这个问题,使算法在低信噪比环境下仍保持较好的性能。

图 4 给出了不同频点数下,语音信号信噪比为 2.5 dB 时 FCS-OMP 与 CS-OMP 性能的曲线图。从图 4(a) 可以看出,当频点阈值 $\eta = 10$ 时, FCS-OMP 的曲线尖锐程度明显高于 CS-OMP,而且 CS-OMP 几乎不能估计出声源的方位,这是由于此时频点数较低,实测的冲激响应与真实值之间误差很大,导致算法性能降低;而当 $\eta = 20$ 时,由图 4(b) 可知,此时 FCS-OMP 曲线的尖锐程度小于 CS-OMP,这可能是由于此时的频点数使得实测的冲激响应更接近于真实值;当 $\eta = 30$ 时, FCS-OMP 曲线尖锐程度又高于 CS-OMP,这是由于随着频点数的增加,空间位置信息获取越多, FCS-OMP 构造的冲激响应越接近真实值,使得 FCS-OMP 算法的性能越好,而 CS-OMP 算法实测的冲激响应与真实值之间误差较大。

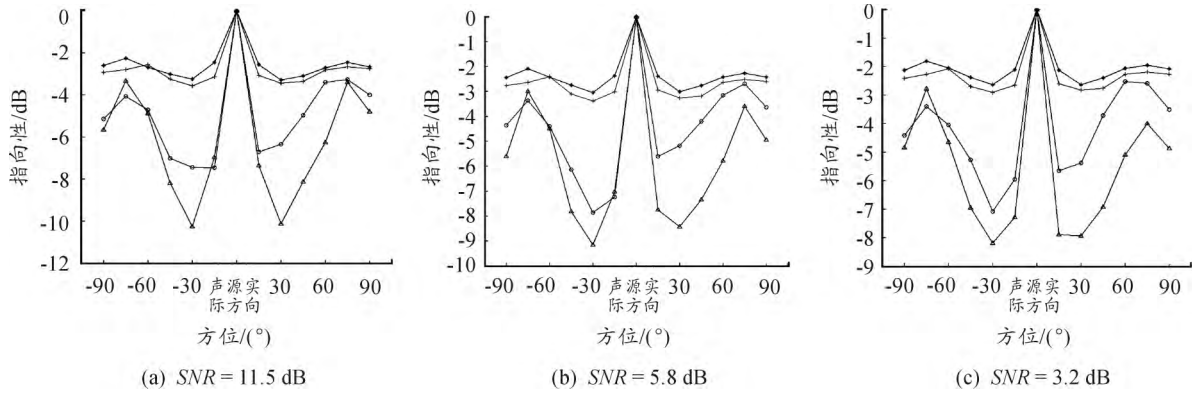


图2 不同信噪比条件下白噪声源远场方位估计结果

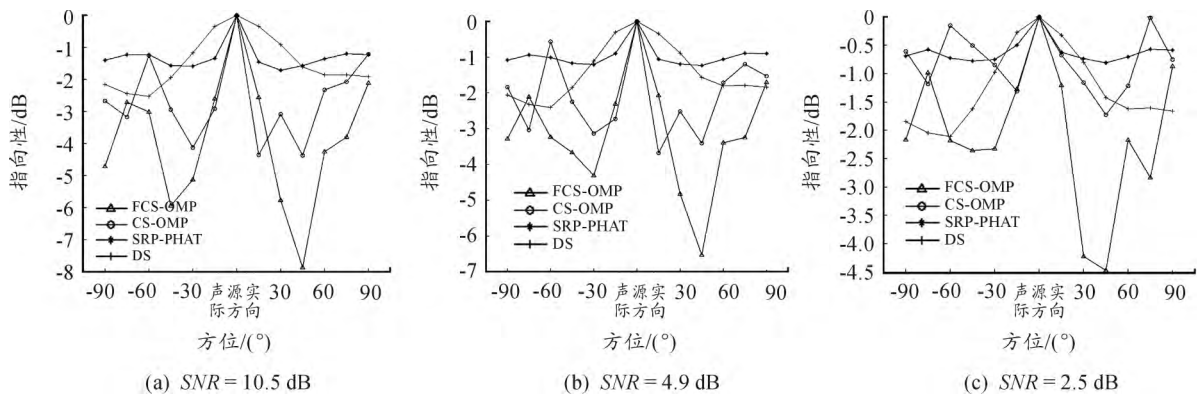


图3 不同信噪比条件下语音声源远场方位估计结果

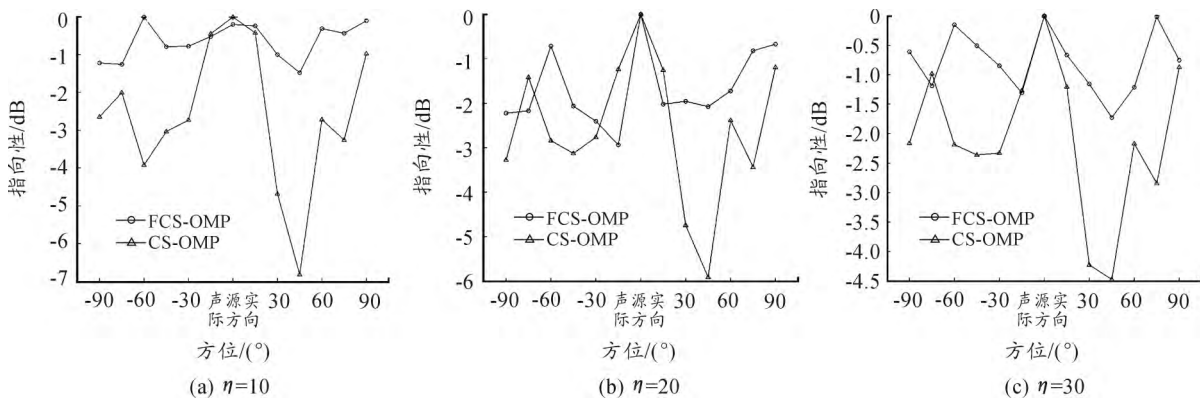


图4 不同频点数下 FCS-OMP 与 CS-OMP 的性能

3 结论

将压缩感知麦克风阵列声源定位方法应用于远场场景,并通过 RIP 分解与重组表明:基于频点叠加可直接利用阵列空间关系构造声源各可能方位对应的冲激响应进行稀疏恢复。实验结果表明,与 SRP-PHAT、DS 算法比较,本研究提出的 FCS-OMP 算法在远场低信噪比环境中更有更好的方位估计

性能。与传统 CS-OMP 算法相比,FCS-OMP 算法采用构造冲激响应替代实测,应用更为方便,可保证远场条件下的估计性能。

参考文献:

[1] 蒋婷,刘建平,张一闻.基于多麦克风阵列的枪声定位算法研究[J].计算机应用与软件,2012,29(13):229

- 231.
- [2] 左佑,于胜云,黄浩,等.低空目标光纤麦克风阵列无源测向技术[J].电子信息对抗技术,2013,28(3):18-21.
- [3] WAX M,KAILATH T. Optimum localization of multiple sources by passive arrays [J]. Acoustics Speech & Signal Processing IEEE Transaction on ,1983,31(5):1210-1217.
- [4] GUSTAFSSON T,RAO B D,TRIVEDI M. Source Localization in Reverberant Environments: Modeling And Statistical Analysis [J]. IEEE Transactions on Speech and Audio Processing 2003,11(6):791-803.
- [5] HUANG L,WU S J,ZHANG L R. A Novel MUSIC Algorithm for Direction-of-Arrive Estimation without the Estimate of Covariance Matrix and Its Eigende Composition [C]//Proceedings of IEEE International Conference on Vehicular Technology, Intercontinental Hotels Dallas,Dallas Texas 2005,1:16-19.
- [6] DIBIASE T H. A high-accuracy,low-latency technique for talker localization environments using microphone arrays [D]. Providence,Rhode Island,USA:Brown University 2000.
- [7] ZHAO Xiaoyan,TANG Jie,ZHOU Lin,et al. Accelerated steered response power method for sound source localization via clustering search [J]. Science China Physics, Mechanics and Astronomy 2013,56(7):1329-1338.
- [8] CANDÈS E J,WAKIN M B. An introduction to compressive sampling [J]. IEEE Signal Processing Magazine,2008,25(2):21-30.
- [9] GURBUZ L C,CEVHER V,MCCLELLAN J H. Bearing estimation via spatial sparsity using compressive sensing [J]. IEEE Transaction on Aerospace and Electronic Systems, 2012,48(2):1358-1369.
- [10] XENAKI A,GERSTOFT P,MOSEGAARD K. Compressive beam-forming [J]. The Journal of the Acoustical Society of America 2014,136(1):260-271.
- [11] 付金山,李秀坤.声矢量阵 DOA 估计的稀疏分解理论研究 [J]. 哈尔滨工程大学学报,2013,34(3):280-286.
- [12] 赵宏伟,刘波,刘恒.用于卫星干扰源定位的压缩感知 DOA 估计方法 [J]. 火力与指挥控制,2016,41(10):25-28.
- [13] 赵小燕,周琳,吴镇扬.基于压缩感知的麦克风阵列声源定位算法 [J]. 东南大学学报(自然科学版),2015,45(2):203-207.
- [14] 张奕,殷福亮.混响和有色噪声环境下的顽健时延估计方法 [J]. 通信学报,2008,29(5):6-12.
- [15] RÉMI MIGNOT,GILLES CHARDON,LAURENT DAUDET. Low frequency interpolation of room impulse responses using compressed sensing [J]. IEEE/ACM Transactions on Audio,Speech,and Language Processing,2014,22(1):205-216.
- [16] 李芳兰,周跃海,童峰,等.采用可调波束形成器的 GSC 麦克风阵列语音增强方法 [J]. 厦门大学学报(自然科学版) 2013,52(2):186-189.
- [17] 陈磊,江伟华,童峰,等.一种可跟踪移动声源方向的麦克风阵列语音增强算法 [J]. 厦门大学学报(自然科学版) 2015,54(4):551-555.
- [18] 张武威.关于室内混响时间的计算问题 [J]. 电声技术,2005(3):17-20.
- [19] GAROFOLO J,LAMEL L,FISHER W,et al. TIMIT Acoustic-Phonetic Continuous Speech (MS-WAV version) [J]. Journal of the Acoustical Society of America,1993,88(88):210-221.

(责任编辑 杨继森)