一种新的语音和噪声活动检测算法及其在手机双麦克风消噪系统中的应用

章雒霏* 张 铭 李 晨
(南京师范大学物理与科学技术学院 南京 210000)

摘 要：针对现有双通道语音活动检测(Voice Activity Detection, VAD)算法依赖于固定阈值难以在多种噪声环境下准确地检测语音和噪声，应用于手机消噪系统会造成语音失真或噪声消除不好等问题，该文提出一种基于神经网络的 VAD 算法，该算法以分频带能量差和归一化互通道相关为特征，采用神经网络对语音和噪声进行分类。在此基础上，将神经网络 VAD 与基于互通道信号功率比值的 VAD 相结合，提出一种新的适用于手机消噪系统的语音和噪声活动检测算法分别对语音和噪声进行检测，并以此进行噪声抑制处理，减少了消噪系统因 VAD 误判而造成的性能下降。实验结果表明，该处理方法在抑制背景噪声和减少语音失真等方面优于现有的消噪算法，对于方向性语音干扰也有很好的抑制效果。

关键词：语音活动检测；语音增强；神经网络
中图分类号：TN912.35 **文献标识码**：A **文章编号**：1009-5896(2016)08-2020-07
DOI: 10.11999/JEIT151302

A New Voice and Noise Activity Detection Algorithm and Its Application to Dual Microphone Noise Suppression System for Handset

ZHANG Luofei ZHANG Ming LI Chen
(*School of Physics and Technology*, *Nanjing Normal University*, *Nanjing* 210000, *China*)

Abstract: Existing dual microphone Voice Activity Detection (VAD) algorithms use normally a fixed threshold. The fixed threshold can not provide an accurate VAD under various noise environments. In such case, it causes voice quality degradation, particularly in handset applications. This paper proposes a new VAD algorithm based on Neural Network (NN). Both sub-band power level difference and inter-microphone cross correlation are used as features. Then the NN based VAD is combined with the method of inter-microphone signal power ratio to get a new voice and noise activity detection algorithm. Furthermore, the algorithm is used into noise suppression in handset to avoid performance degradation caused by VAD misjudgment. Experimental results show that the proposed method provides better noise suppression performance and lower speech distortion compared to the existing method.

Key words: Voice Activity Detection (VAD); Speech enhancement; Neural Network (NN)

1 引言

说话人处于噪声环境中时，远端接听者往往会听到难以忍受的噪声[1]，为了解决这个问题，现有手机集成了语音增强模块来提高语音质量。传统的单通道语音增强算法[2–6]无法很好地处理非稳态噪声，而多通道算法[1,7–13]在利用语音与噪声性质差异的同时也结合了两者的空间差异性，使得算法在非稳态噪声环境下性能得到很大改善。考虑到尺寸、功耗和计算复杂度等问题，手机主要使用的是双麦克风语音增强系统。

语音活动检测(Voice Activity Detection, VAD)可以从带噪语音信号中确定出语音的起始和结束位置，准确的 VAD 可以帮助消噪算法对噪声进行有效抑制同时尽可能地减少语音信号的失真。目前，各种单通道或者双通道的 VAD 算法已广泛地应用于手机消噪系统中。其中，基于双麦克风能量差(Power Level Differences, PLD)[1]及其改进的算法[10–14]具有较好的检测结果且复杂度低易于实现，因此得到了广泛的关注和研究。通话时，手机底部的主麦克风接收到语音信号能量远大于手机顶端的次麦克风接收能量，而噪声信号的能量基本相同。基于这样

收稿日期：2015-11-23；改回日期：2016-04-12；网络出版：2016-05-31
*通信作者：章雒霏 lincover@126.com
基金项目：江苏省自然科学基金，江苏省声频技术工程重点实验室基金项目(BE2014139)
Foundation Items: Program of Natural Science Research of Jiangsu Higher Education Institutions of China, Program of Science and Technology of Jiangsu (BE2014139)

的特性，PLD 算法通过对双麦克风信号的能量差设定阈值来区分语音和噪声，但其算法性能会受到麦克风增益，噪声种类和信噪比等因素的影响，在此基础上，文献[10]提出了基于双麦克风后验信噪比差异的 VAD 算法减少了麦克风增益的影响，文献[14]提出了基于PLD比率(PLD Ratio, PLDR)的算法提高了 PLD 算法的准确率。虽然上述算法在稳态及非稳态噪声环境中取得了一定效果，但难以同时保证语音和噪声检测的准确性，应用于手机消噪系统会造成语音失真，降低可懂度。

针对上述问题，本文提出了一种新的基于神经网络的 VAD 算法，该算法以分频带能量差和归一化互通道相关作为特征，采用神经网络对语音和噪声进行分类，不依赖于固定阈值，较现有的基于 PLD 的算法准确性更高。在此基础上，本文将神经网络 VAD 与基于互通道信号功率比值的 VAD 相结合，提出一种新的适用于手机消噪系统的语音和噪声活动检测算法，该算法分别对语音和噪声进行检测，减少了消噪算法因 VAD 的误判而造成的性能下降，与现有的双麦克风消噪算法相比，本算法能够更有效地抑制噪声，减少语音失真。

本文第 2 节描述神经网络 VAD 的原理；第 3 节介绍结合神经网络 VAD 提出的语音和噪声检测算法及其在手机消噪系统中的应用；第 4 节给出实验结果和分析；第 5 节进行总结。

2　基于神经网络的语音活动检测算法

5 dB Babble 噪声环境下，双麦克风接收到的带噪语音信号功率如图 1 所示。

频域上双通道接收到的纯净语音信号的能量差几乎都在 10 dB 左右[1]，而背景噪声存在时语音信号的某些频带受到噪声的污染能量差下降(如图 1 中 1.0~1.5 kHz 之间)，但部分频带仍然保持着 10 dB 左右的能量差(如图 1 中 1.5~2.5 kHz 之间)。这些频带的能量差可以作为表征目标语音存在的特征，为了更好地利用这些频带的信息，本算法对频域进行划分，计算子带互通道能量差(sub-band power level difference)作为神经网络的特征，计算过程如式(1)。首先将时域信号转化到频域，得到两个通道在频域的信号：

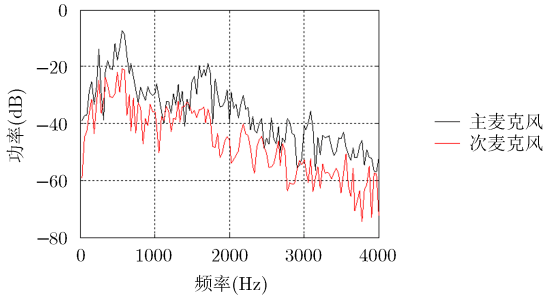

图 1　双麦克风接收的带噪语音信号功率

$$X_i(k,n)=S_i(k,n)+N_i(k,n),\ i=1,2 \tag{1}$$

其中，$X_i(k,n)$，$S_i(k,n)$和 $N_i(k,n)$分别代表频域带噪、纯净语音和噪声信号。k代表频率点，n代表语音帧标号，i为主、次麦克风的标号。通过计算得到主、次麦克风中信号的功率谱密度，如式(2)：

$$P_{X_i}(k,n)=P_{S_i}(k,n)+P_{N_i}(k,n),\quad i=1,2 \tag{2}$$

对每个子带(本算法按照 MEL 频带划分)计算互通道能量差的均值如式(3)所示。

$$S_p(b,n)=\frac{1}{u_h(b)-u_l(b)}\sum_{k=u_l(b)}^{u_h(b)}10\lg_{10}\frac{P_{X_1}(k,n)}{P_{X_2}(k,n)} \tag{3}$$

其中，$u_h(b)$和$u_l(b)$分别为第 b 个子带的上下边界。

因为目标语音距主麦克风较次麦克风近，主麦克风早于次麦克接受到语音信号，而背景噪声到达麦克风的距离基本相等，时延较语音小，所以双通道时延也是区分语音和噪声的一个重要的特征，在本算法中，使用归一化的互通道相关函数来作为表征时延的特征，计算式为

$$T(n,\tau)=\frac{\sum_{l=1}^{L}\left(x_{1,n}(l)-\bar{x}_{1,n}\right)\left(x_{2,n}(l-\tau)-\bar{x}_{2,n}\right)}{\sqrt{\sum_{l=1}^{L}\left(x_{1,n}(l)-\bar{x}_{1,n}\right)^2}+\sqrt{\sum_{l=1}^{L}\left(x_{2,n}(l)-\bar{x}_{2,n}\right)^2}} \tag{4}$$

其中，L代表时域信号的长度，τ为延时，$\bar{x}_{1,n}$和$\bar{x}_{2,n}$分别为主、次麦克风接受的带噪信号的均值。

反向传播(Back Propagation, BP)神经网络是使用最为广泛的神经网络，在训练阶段，通过调整神经元之间连接的权值，BP 神经网络可以完成输入和输出之间复杂的映射关系。本文使用的是 3 层的 BP 神经网络。其中输入层为提取的两个特征矢量，即分频带能量差和归一化互通道相关函数，输出层为对应的语音活动检测的标签(1：语音；0：噪声)。

3　手机双麦克风语音增强系统

双麦克风语音增强系统框图如图 2 所示，滤波器 1 将次麦克风信号作为参考，主麦克风信号作为输入，通过 VAD 检测信噪比较高的语音段控制滤波器调整参数将目标语音从次麦克风中滤除得到噪声信号。滤波器 2 将主麦克风信号作为参考，滤波器 1 输出噪声信号作为输入，通过噪声活动检测 NAD (Noise Activity Detection)在噪声段控制滤波器调整参数将噪声信号从主麦克风的带噪语音信号中滤除得到增强语音信号。

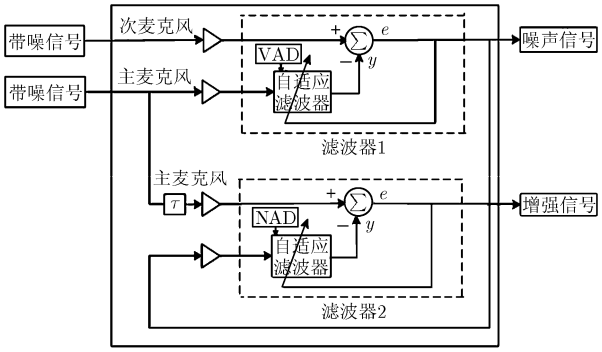

图 2 手机双麦克风语音增强系统框图

实际上，滤波器 1 和滤波器 2 的参数分别模拟了语音和噪声信号在两个麦克风之间的传递函数，为了避免在信噪比较低的语音和噪声混合部分对滤波器参数进行调整造成滤波器参数与传递函数的失配，本文中，我们结合神经网络 VAD 提出一种新的语音和噪声活动检测算法，该算法通过 VAD 检测信噪比较高的语音段落控制滤波器 1 的参数调整，同时利用 NAD 检测噪声段落控制滤波器 2 的参数调整。

3.1 语音活动检测(VAD)

现有的PLD算法通过设定固定阈值δ来区分语音和噪声。但是互通道功率比值的大小会因信噪比和噪声种类的改变而改变，固定的阈值无法得到准确结果。针对这一问题，本算法做了改进，采用不同的平滑参数α计算两个通道信号的功率。

$$P_{i_s}(t)=\alpha_s P_{i_s}(t-1)+(1-\alpha_s)\,x_i^2(t),\ i=1,2 \tag{5}$$

$$P_{i_f}(t)=\alpha_f P_{i_f}(t-1)+(1-\alpha_f)\,x_i^2(t),\ i=1,2 \tag{6}$$

其中，$P_{i_s}(t)$为长时间平滑计算得到的功率，平滑参数为$\alpha_s=0.999$，$P_{i_f}(t)$为短时间平滑得到的功率，平滑参数为$\alpha_f=0.9$，采用带噪语音信号前 32 个采样点的平均作为初值。将长平滑和短平滑分别计算得到的两个通道的功率相比得到互通道功率的比值。

$$P_s(t)=P_{1s}(t)/P_{2s}(t) \tag{7}$$

$$P_f(t)=P_{1f}(t)/P_{2f}(t) \tag{8}$$

其中，$P_s(t)$和$P_f(t)$分别为长、短时间平滑互通道功率比。因为语音信号是高度非平稳信号，短时间平滑得到的信号功率会比长时间平滑大得多；而噪声较语音信号平缓，所以短时间和长时间平滑的信号功率差别较语音小得多，$P_s(t)$和$P_f(t)$的结果如图 3 所示。

从图3中可以看出，语音存在的部分，短平滑计算的互通道功率比$P_f(t)$比长平滑计算的比值$P_s(t)$大得多，可以通过比较$P_f(t)$与$P_s(t)$的大小来确定语音信号存在且信噪比较高的时域采样点，但是通过调整判断阈值不能够完全地区分语音和噪声，如图4所示(VAD等于1表示语音信号)，当设定$P_f(t)>2P_s(t)$的采样点为语音时，部分噪声被误判为语音，而提高阈值为$P_f(t)>6P_s(t)$时，虽然误判为语音的噪声减少了，但是语音检测的准确性也下降了。

基于神经网络的 VAD 可以准确地判断出语音存在的部分，将神经网络 VAD 结果和基于长和短时平滑计算的功率比值确定的语音存在且信噪比较高的部分相结合可以去除误判为语音的噪声采样点，5 dB babble 噪声环境下的结果如图 5 所示。

3.2 噪声活动检测 NAD

将滤波器 1 输出的噪声信号与主麦克风中的带噪语音信号进行比较，因语音部分能量较大，当噪声信号与语音信号的能量相比时，比值会非常小，我们可以对噪声与带噪信号能量的比值设定阈值来确定噪声段，计算过程如式(9)和式(10)：

$$P_{\mathrm{Ns}}(t)=\alpha_s P_{\mathrm{Ns}}(t-1)+(1-\alpha_s)n^2(t) \tag{9}$$

$$P_{\mathrm{Nf}}(t)=\alpha_f P_{\mathrm{Nf}}(t-1)+(1-\alpha_f)n^2(t) \tag{10}$$

其中，$P_{\mathrm{Ns}}(t)$为长时间平滑计算得到的噪声功率，$P_{\mathrm{Nf}}(t)$为短时间平滑计算得到的噪声功率，采用带噪语音信号前 32 个采样点的平均作为初值，$n(t)$为滤波器 1 输出的噪声信号。与 3.1 节一样，分别将$P_{\mathrm{Ns}}(t)$和$P_{\mathrm{Nf}}(t)$与长、短平滑计算得到的主麦克风中带噪信

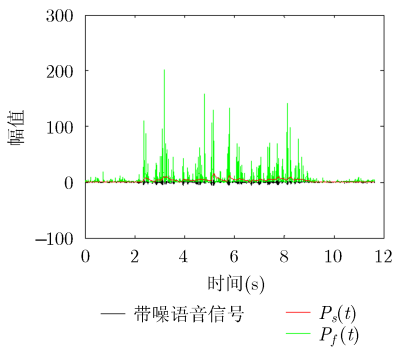

图 3 不同平滑参数计算的互通道能量的比值

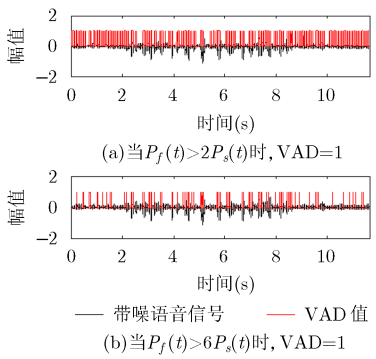

图 4 5 dB Babble 噪声下利用$P_f(t)$和$P_s(t)$判断语音信号

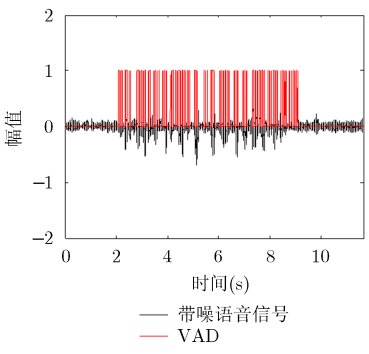

图 5 VAD 的结果

号功率 $P_{1s}(t)$ 和 $P_{1f}(t)$ 进行比较，得到噪声和带噪语音信号的功率比值。

$$P_{n\_s}(t) = P_{Ns}(t) \big/ P_{1s}(t) \tag{11}$$

$$P_{n\_f}(t) = P_{Nf}(t) \big/ P_{1f}(t) \tag{12}$$

其中，$P_{n\_s}(t)$ 与 $P_{n\_f}(t)$ 分别为长平滑和短平滑计算得到的噪声与主麦克风中带噪语音的功率比值，当语音存在的时候，噪声与语音的比值会接近于零，而噪声段的比值则较大且短平滑的值远远大于长平滑的比值，为了在噪声段增加长时与短时平滑功率比值的差距，我们对 $P_{n\_s}(t)$ 再次进行平滑：

$$P_{n\_ss}(t) = \alpha_{ss} P_{n\_ss}(t-1) + (1-\alpha_{ss}) P_{n\_s}(t) \tag{13}$$

其中，$P_{n\_ss}(t)$ 为对 $P_{n\_s}(t)$ 进行再次平滑得到的功率比值，这里的平滑系数 α_{ss} 根据神经网络 VAD 的结果进行调整，在语音段 α_{ss} 为 1 保持 $P_{n\_ss}(t)$ 不变，在噪声段 α_{ss} 为 0.999 迭代平滑计算 $P_{n\_ss}(t)$，经过再次平滑后的 $P_{n\_ss}(t)$ 在噪声段更为平缓，与 $P_{n\_f}(t)$ 的差距更大，将 $P_{n\_f}(t)$ 与 $P_{n\_ss}(t)$ 进行比较更有利于我们准确地判断出噪声采样点。

4　实验结果

实验使用手机长度为 13 cm，在一个 $7.91\times 7.31\times 4.85$ m^3 的房间中进行测试，房间的混响为 0.3 s，使用 B&K HATS 仿真头的人工嘴来播放目标语音信号，通过 ACTS 的 8 个喇叭噪声放音系统来模拟真实的噪声环境，人工头放置在圆点，8 个喇叭以一个环形位于人工头的四周，距离人工头大约为 2 m。信号的采样率为 8 kHz，帧长 L=256，帧移 M=128。实验选取 100 段语音，其中 80 段用于神经网络的训练，剩余 20 段用于验证神经网络的结果。选取 6 种常见的噪声环境，Babble, Car, Restaurant, Office, Street 和方向性的语音干扰，信噪比分别为 5 dB, 10 dB 和 15 dB。神经网络采用 MATLAB 2014a 的神经网络工具箱。隐藏层为 30 个神经元，输入层到隐藏层采用 tansig 作为激活函数，隐藏层到输出层采用 purline 作为激活函数，最大迭代次数为 2000 次，学习步长为 0.01，学习函数为 traingdx。采用 24 个 MEL 频带计算子带互通道能量差，同时，选取时延从-10 到+10 每隔 1 个采样点计算归一化互通道相关。一共 45 个值作为神经网络的输入，输出层为对应的语音活动检测的标签(1：语音；0：噪声)。

首先对神经网络 VAD 算法的准确性进行验证，将该算法与基于 PLD 比率(PLDR)[14]的 VAD 算法进行比较。分别用 3 个性能指标来衡量语音活动检测的准确性，P_{sh} 为检测正确的语音信号帧/语音信号总帧数，P_{nh} 为检测正确的噪声信号帧/非语音信号总帧数，P_{gh} 为总的准确率。

表 1　10 dB 信噪比噪声环境下，PLDR 和本文算法的语音活动检测结果

噪声种类	P_{sh}		P_{nh}		P_{gh}	
	本文算法	PLDR	本文算法	PLDR	本文算法	PLDR
Babble	95.26	91.77	90.95	87.41	93.95	90.44
Car	94.61	92.13	93.05	90.95	94.13	91.77
Restaurant	94.72	87.16	93.01	90.66	94.20	88.23
Office	93.87	90.72	93.38	89.59	93.72	90.38
Street	94.16	89.60	93.09	90.81	93.83	90.01
45° 干扰	92.82	89.22	92.80	81.53	92.76	86.87
135° 干扰	94.63	91.31	92.18	82.73	93.88	88.72
225° 干扰	94.88	89.13	90.09	80.04	93.42	86.25
315° 干扰	94.58	84.32	90.74	75.11	93.41	81.51

从表 1 中可以看出，本文算法无论是在语音帧、噪声帧还是总的准确率方面都要优于 PLDR 算法。干扰人声也是手机通话中非常常见的一类噪声，但是，由于干扰人声是高度非平稳信号且具有方向性，现有的 VAD 算法无法很好地处理这类噪声。我们选取 4 个不同方位的语音干扰比较两个算法的性能。如表 1 所示，本文提出的算法利用了目标语音和干扰人声的空间差异来区分两者获得了准确的结果。而 PLDR 算法在干扰人声的噪声环境下性能有了很大的下降。

为了测试神经网络 VAD 在不同信噪比下的性能，我们分别选取 5 dB, 10 dB, 15 dB 的信噪比进行验证，结果如表 2 所示。从表 2 中可以看到，本文算法不依赖于固定的阈值，即使在 5 dB 这样的低信噪比下依旧可以取得很好的 VAD 结果，非常适合于手机的应用。

本文采用 ACTS 音频评价系统中的对数谱距离(Logistic Spectral Distance, LSD)，客观质量评估(Perceptual Evaluation of Speech Quality, PESQ[15])和信噪比(SNR)分别对本文提出的语音增强算法和文献[1]提出的基于 PLD 的手机双麦克风语音增强算法的性能进行了衡量。

信噪比衡量了语音增强算法的噪声抑制效果。从表 3 中可以看出，本文提出的消噪算法相较于 PLD 算法有了很大的提升，特别是在 5 dB 信噪比的条件下，本文算法输出的信噪比均能够达到 15 dB 左右。为了验证算法对于方向性干扰人声的抑制效果，我们选取了 45° 方位入射的干扰人声，因为 45° 方位的干扰人声与目标语音的入射方位非常接近，传统的消噪算法很难对其进行有效的抑制，从结果中可以看出，本文算法对于 45° 方位的干扰人声也有很好的效果，而 PLD 算法的性能则大大地下降。

表 2 不同信噪比环境下，本文算法的语音活动检测结果

噪声种类	信噪比(dB)								
	5			10			15		
	P_{sh}	P_{nh}	P_{gh}	P_{sh}	P_{nh}	P_{gh}	P_{sh}	P_{nh}	P_{gh}
Babble	93.65	89.47	92.38	95.26	90.95	93.95	96.27	90.70	94.57
Car	91.63	94.04	92.36	94.61	93.05	94.13	96.76	91.48	95.15
Restaurant	92.84	93.34	92.99	94.72	93.01	94.20	96.27	91.98	94.96
Office	91.59	94.16	92.38	93.87	93.38	93.72	96.13	91.86	94.82
Street	91.41	93.99	92.20	94.16	93.09	93.83	96.31	91.90	94.96

表 3 在不同噪声和信噪比条件下经过语音增强处理之后的输出信噪比(dB)

噪声种类	输出信噪比(dB)					
	5		10		15	
	本文算法	PLD	本文算法	PLD	本文算法	PLD
Babble	14.01	9.02	18.88	13.43	23.19	17.82
Car	19.22	13.21	24.86	17.42	29.35	21.22
Restaurant	16.54	10.43	21.43	14.74	26.56	19.01
Office	19.32	13.91	23.37	17.51	28.23	21.14
Street	18.75	12.92	22.65	16.93	26.98	20.55
45°干扰	15.98	6.94	18.79	11.55	22.12	16.10

语音的可懂度在手机的通信中非常的重要，消噪算法会带来一定程度的语音失真，LSD 指标主要用来衡量增强语音的失真度，LSD 值越大说明语音信号的失真越严重，越小表明语音信号失真越小，质量越接近于原始语音。表 4 给出本文算法与 PLD 算法增强处理后的 LSD 对比结果

从表 4 中可以看出，本文提出的消噪算法相较于 PLD 算法对语音信号的损失更小，说明经过本文算法处理的语音失真更小，语音质量更接近于原始语音信号，对于方向性的语音干扰也得到了较好的结果。

本文还采用 PESQ 来测试语音增强算法对语音客观质量的影响，PESQ 的值越高说明语音质量越高。从表 5 中可以看出，与 PLD 的算法相比，本文提出的消噪算法的输出语音具有更好的语音质量，非正式的主观听觉测试与上述结果一致。

5 总结

本文提出了一种新的基于神经网络的 VAD 算法，结合两个表征目标语音空间特性的特征，即分频带能量差和互通道相关函数作为神经网络的输入训练神经网络进行语音活动检测。再将基于双通道功率比值的 VAD 结果与神经网络 VAD 的结果相结合，提出一种新的适用于手机消噪系统的语音和噪声检测算法，该算法分别对语音和噪声进行检测，减少了消噪系统因 VAD 的误判而造成的性能下降。实验结果表明，与现有的基于 PLD 的消噪算法相比，无论是 VAD 的准确率还是语音增强的效果均有了提升，避免了消噪算法对于语音信号的损害，提高了语音的可懂度，保证了手机通话的质量。

表 4 本文算法与 PLD 算法增强处理后的 LSD 对比结果

噪声种类	输入信噪比(dB)					
	5		10		15	
	本文算法	PLD	本文算法	PLD	本文算法	PLD
Babble	3.77	4.15	2.58	2.88	1.73	1.88
Car	2.27	2.37	1.59	1.70	1.42	1.65
Restaurant	2.70	3.62	1.76	2.48	1.41	1.61
Office	2.38	2.43	1.75	1.92	1.39	1.45
Street	2.62	2.73	1.82	1.93	1.39	1.56
45°干扰	3.19	10.49	2.11	9.34	1.71	8.09

表 5 不同信噪比和噪声条件下经过语音增强处理之后的 PESQ

噪声种类	输入信噪比(dB)					
	5		10		15	
	本文算法	PLD	本文算法	PLD	本文算法	PLD
Babble	2.76	2.45	3.02	2.80	3.31	3.23
Car	3.44	3.21	3.57	3.43	3.74	3.60
Restaurant	3.08	2.69	3.38	3.11	3.61	3.39
Office	3.29	2.98	3.45	3.27	3.75	3.51
Street	2.52	2.16	3.13	2.88	3.76	3.43
45°干扰	2.69	1.96	2.97	2.53	3.49	3.09

参 考 文 献

[1] JEUB M, HERGLOTZ C, NELKE C M, *et al*. Noise reduction for dual-microphone mobile phones exploiting power level differences[C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, Kyoto, 2012: 1693–1696. doi: 10.1109/ICASSP.2012.6288223.

[2] XU Y, DU J, and DAI L R. A Regression approach to speech enhancement based on deep neural networks[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2015, 23(1): 7–19. doi: 10.1109/TASLP.2014.2364452.

[3] XU Y, DU J, and DAI L R. An experimental study on speech enhancement based on deep neural networks[J]. *IEEE Signal Processing Letters*, 2014, 21(1): 65–68. doi: 10.1109/LSP.2013.2291240.

[4] WANG Y X, NARAYANAN A, and WANG D L. On training targets for supervised speech separation[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2014, 22(12): 1849–1859. doi: 10.1109/TASLP.2014.2352935.

[5] 王明合, 张二华, 唐振明, 等. 基于Fisher 线性判别分析的语音信号端点检测方法[J]. 电子与信息学报, 2015, 37(6): 1343–1349. doi: 10.11999/JEIT141122.

WANG Minghe, ZHANG Erhua, TANG Zhenmin, *et al*. Voice activity detection based on Fisher linear discriminant analysis[J]. *Journal of Electronics & Information Technology*, 2015, 37(6): 1343–1349. doi: 10.11999/JEIT141122.

[6] 郭海燕, 李枭雄, 李拟珺. 基于基频状态和帧间相关性的单通道语音分离算法[J]. 东南大学学报(自然科学版), 2014, 44(6): 1100–1104.

GUO Haiyan, LI Xiaoxiong, and LI Nijun. Single-channel speech separation based on pitch state and interframe correlation[J]. *Journal of Southeast University* (*Natural Science Edition*), 2014, 44(6): 1100–1104.

[7] NELKE C, BEAUGEANT C, and VARY P. Dual microphone noise PSD estimation for mobile phones in hands-free position exploiting the coherence and speech presence probability[C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, Vancouver, 2013: 7279–7283. doi: 10.1109/ ICASSP.2013.6639076.

[8] YOUSEFIAN N, RAHMANI M, and AKBARI A. Power level difference as a criterion for speech enhancement[C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, Taipei, 2009: 4653–4656. doi: dx.doi.org/10.1109/ICASSP.2009.4960668.

[9] YOUSEFIAN N, AKBARI A, and RAHMANI M. Using

power level difference for near field dual-microphone speech enhancement[J]. *Applied Acoustics*, 2009, 70(11/12): 1412–1421.

[10] FU Z H, FAN F, and HUANG J D. Dual-microphone noise reduction for mobile phone application[C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, Vancouver, 2013: 7239–7243. doi: 10.1109/ICASSP.2013.6639068.

[11] MEYER-BAESE U. Digital Signal Processing with Field Programmable Gate Arrays[M]. Third Edition, Berlin Heidelberg: Springer, 2007: 298–305.

[12] RUBIO J E, ISHIZUKA K, SAWADA H, *et al*. Two-microphone voice activity detection based on the homogeneity of the direction of arrival estimates[C]. IEEE International Conference on Acoustics, Speech, and Signal Processing, Honolulu, 2007: 385–388. doi: 10.1109/ICASSP.2007.366930.

[13] ZHAO H C, LI L G, and LI L H, *et al*. Dual-microphone adaptive noise canceller with a voice activity detector[C]. IEEE Region 10 Symposium, Kuala Lumpur, 2014: 551–554. doi: 10.1109/TENCONSpring.2014.6863095.

[14] CHOI J H and CHANG J H. Dual-microphone voice activity detection technique based on two-step power level difference ratio[J] *IEEE Transactions on Audio, Speech and Language Processing*, 2014. 22(6): 1069–1081.

[15] HU Y, and LOIZHOU P C. Evaluation of objective quality measures for speech enhancement[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, 16(1): 229–238.

章雒霏： 女，1990 年生，博士生，研究方向为信号处理、语音增强、语音识别、语音定位.

张 铭： 男，1963 年生，博士生导师，特聘教授，研究方向为信号处理、语音增强、语音识别.

李 晨： 女，1980 年生，博士，研究方向为信号处理、语音增强、语音识别、语音定位.