

Research on Noise Suppression Algorithm Combining Geometric Phase Information

Shuo Zhang, Xue Li, Chen Chen, Ruirui Han

School of Opto-Electronic Information, Yantai University, Yantai Shandong
Email: 98293409@qq.com

Received: Jul. 4th, 2018; accepted: Jul. 18th, 2018; published: Jul. 25th, 2018

Abstract

In this paper, a noise suppression algorithm is proposed to establish a mathematical relationship between amplitude distortion and prior signal-to-noise ratio and post signal-to-noise ratio (SNR) in speech signal processing, due to the introduction of noise from the estimated pure speech amplitude error. Speech enhancement process can be expressed by gain function, but the gain function will introduce two types of speech distortion: amplifying distortion and attenuating distortion of speech signal amplitude. In fact, when the processed speech contains only attenuation distortion and amplification distortion less than 6 dB, the influence on speech quality is relatively small. It is necessary to limit the amplitude spectrum after enhancement. At the same time, different constraints for speech regions with amplification distortion exceeding 6 dB will help to improve speech quality. In this paper, we propose a fusion of geometric phase information to establish the mathematical relationship between amplitude distortion and prior signal-to-noise ratio and posteriori signal-to-noise ratio. According to the criteria of priori SNR and post signal-to-noise ratio, the restriction of amplify distortion exceeding 6 decibels, attenuation distortion and less than 6 decibels is given at the same time. Experimental results show that the algorithm proposed in this paper is simple and feasible, and it can effectively improve speech quality and improve speech intelligibility.

Keywords

Speech Enhancement, Geometric Phase, Priori Signal-to-Noise Ratio, Noise Reduction

融合几何相位信息的噪声抑制算法研究

张 硕, 李 雪, 陈 晨, 韩蕊蕊

烟台大学光电信息科学技术学院, 山东 烟台
Email: 98293409@qq.com

收稿日期: 2018年7月4日; 录用日期: 2018年7月18日; 发布日期: 2018年7月25日

摘要

本文针对语音信号处理中因估计的纯净语音幅度误差会引入噪声,提出了一种通过融合几何相位信息,使幅度失真与先验信噪比和后验信噪比建立数学关系的噪声抑制算法,具有较高的可行性,提高了语音的可懂度。现有数据表明,当被处理的语音只包含衰减失真和小于6分贝的放大失真时,对语音质量影响较小,需要对增强之后的幅度谱进行简单限制。同时,需要对于放大失真超过6分贝的语音区域采取不同的约束条件,会对提高语音质量有较大的帮助。但目前的研究大多采用估计的纯净语音与原是纯净语音谱幅度进行直接对比,现实中可行性较差。本文主要提出一种融合几何相位信息,把幅度产生的失真与先验信噪比和后验信噪比建立数学关系,然后根据此先验信噪比和后验信噪比的判定条件对幅度放大失真超过6分贝以及衰减失真和小于6分贝的放大失真同时进行不同的限制约束。实验结果表明,在对数谱距离(LSD),分段信噪比(SegSNR),短时清晰度客观测度(STOI)等评测条件下,语音增强效果有明显的提高。总而言之,本文提出的算法简单易行,能有效地提升语音质量,提高语音的可懂度。

关键词

语音增强, 几何相位, 先验信噪比, 噪声抑制

Copyright © 2018 by authors and Hans Publishers Inc.

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

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



Open Access

1. 引言

目前语音增强算法技术已经取得了瞩目的发展,语音增强算法极大地提升了语音质量。其中比较经典的算法有维纳滤波器、谱减法、最小均方误差(MMSE)估计器[1]等有效算法。但在单一控制条件下,传统的噪声抑制算法会对输入语音产生一定的噪声污染,典型的如维纳滤波。当从均方误差角度进行度量时,纯净语音和增强语音幅度谱之间差值的正负比较容易被忽略。所以用传统的语音增强算法处理被破坏的语音时,经常会由以上原因产生失真。

由 Loizou 等人根据增强语音与纯净语音幅度的大小关系,把增强之后的语音失真分为三个部分,分别是幅度衰减部分,幅度放大且放大倍数小于等于 2 倍(即放大上限为 6.02 dB),幅度放大超过 2 倍(放大倍数超过 6.02 dB)三个部分[2]。目前的数据表明,近一半的数据下降在第一区,其特点是衰减失真,而几乎另一半的数据落在第三区,其特点是放大失真超过 6.02 分贝。只有一小部分(12%~18%)的数据在第二区域[3],其特点是低放大失真,小于 6.02 分贝。但是 Loizou 等人的研究是根据估计的纯净语音与原始纯净语音的幅度谱比较,这在现实中是不可能实现的,因为在实际运用时原始纯净语音信号是未知的。

本文基于 loizou 等人对第一区域第二区域第三区域的划分,把第一和第二区划归为一个区域,结合几何相位信息,提出了针对新划分的区域以及第三区域,利用先验信噪比与后验信噪比的判定条件,代替了与原始纯净语音的幅度对比,从而对幅度增益进行约束,具有简便易行的特点。通过 MATLAB 进行大量实验仿真,将本文提出的改进算法与现有算法进行比较,结果显示具有较好的语音增强性能。

2. 语音增强基本理论

在语音增强技术中,假设 t 时刻的带噪语音信号为 $y(t)$,纯净语音信号 $x(t)$ 和噪声 $v(t)$ 互不相关,其时

域表达式如下[1]:

$$y(t) = x(t) + v(t) \quad (1)$$

对上式进行 N 点 DFT 变换后, 与时域语音对应的语音谱表示如下:

$$Y_{m,k} = X_{m,k} + V_{m,k}, k = 0, \dots, N-1 \quad (2)$$

其中, $Y_{m,k}$, $X_{m,k}$ 和 $V_{m,k}$ 分别表示带噪语音频谱、纯净语音频谱和噪声谱, m , k 分别表示帧索引和频率。

经过增强处理过后的纯净语音频谱估计可以由增益因子 $G_{k,m}$ 表示为:

$$\hat{Y}_{m,k} = G_{m,k} \cdot Y_{m,k} \quad (3)$$

其中, $\hat{Y}_{m,k}$ 是估计的纯净语音谱, 增益因子 $G_{m,k}$ 一般来说是先验 SNR 和后验 SNR 的函数。

先验 SNR 的定义为:

$$SNR_{prio} = \frac{E\{|X_{m,k}|^2\}}{E\{|V_{m,k}|^2\}} \quad (4)$$

后验 SNR 的定义为:

$$SNR_{post} = \frac{|Y_{m,k}|^2}{E\{|V_{m,k}|^2\}} \quad (5)$$

其中, $E\{\cdot\}$ 表示数学期望算子。

一般来说, 普通维纳滤波语音增强算法的系统增益因子可以表示仅为先验信噪比参数的函数[4]:

$$G_{m,k} = \frac{SNR_{prio}}{1 + SNR_{prio}} \quad (6)$$

目前已经证明, 平方根维纳滤波相比其他比较复杂的降噪算法, 具有计算量小且易于实现的优点, 且在语音质量和清晰度上仍然能够达到同样的要求, 本文在实验中决定采用平方根维纳滤波。

平方根维纳滤波增益函数 $G_{m,k}$ 为[3]:

$$G_{m,k} = \sqrt{\frac{SNR_{prio}}{1 + SNR_{prio}}} \quad (7)$$

其中 SNR_{prio} 是先验 SNR, 根据以下方程得:

$$SNR_{prio}(m,k) = \alpha \frac{|\hat{Y}_{m-1,k}|^2}{\lambda_{N(m-1,k)}} + (1-\alpha) \cdot \max\left[\frac{|Y_{m,k}|^2}{\lambda_{N(m,k)}} - 1, 0\right] \quad (8)$$

其中, $\lambda_{N(m,k)}$ 是估计的背景噪声功率谱, α 是平滑因子, 通常取值为 0.98。

最后, 通过(2) (3) (7) (8)式, 可获得维纳滤波语音增强系统的输出 $\hat{Y}_{m,k}$, 再将其通过 N 点 IDFT 变换至时域, 即可得到估计的纯净语音信号。

3. 本文改进算法

首先, 为了分析估计的纯净语音信号与输入纯净语音振幅的大小关系对噪声抑制的影响, 根据 loizou 等人的研究, 定义了剩余信噪比[5]的概念:

$$SNR_{ESI(m,k)} = \frac{|X_{m,k}|^2}{(|X_{m,k}| - |\hat{Y}_{m,k}|)^2} = \frac{1}{\left(1 - \frac{|\hat{Y}_{m,k}|}{|X_{m,k}|}\right)^2} \quad (9)$$

由上式 $\hat{Y}_{m,k}$ 与 $X_{m,k}$ 的比值, 可以把语音信号分为三个区间[6]:

区间 I: 在此区域, $|\hat{Y}_{m,k}| \leq |X_{m,k}|$, 表示幅度产生了衰减失真。

区域 II: 在此区域, $|X_{m,k}| < |\hat{Y}_{m,k}| \leq 2|X_{m,k}|$, 表示幅度产生了放大失真且放大上限为 6.02 dB。

区域 III: 在此区域, $2|X_{m,k}| < |\hat{Y}_{m,k}|$, 表示振幅产生超过 6.02 dB 放大。

由目前的研究可知, 对第 I 和第 II 区域可以进行合并, 并进行如下限制约束[7]时语音质量有所改善:

$$|\hat{Y}_M| = \begin{cases} |\hat{Y}_{m,k}|, & \text{当 } |\hat{Y}_{m,k}| \leq 2|X_{m,k}| \text{ 时} \\ 0, & \text{其他} \end{cases} \quad (10)$$

把(2)(3)式代入(10)式区域判定条件可得:

$$G_{m,k} |X_{m,k} + V_{m,k}| \leq 2|X_{m,k}| \quad (11)$$

对方程两边同时平方:

$$G_{m,k}^2 |X_{m,k}^2 + V_{m,k}^2 + 2|X_{m,k}||V_{m,k}|\cos[\theta_x - \theta_v]| \leq 4|X_{m,k}|^2 \quad (12)$$

引入几何相位[8]:

根据图 1 的极坐标形式可得:

$$a_y e^{j\theta_y} = a_x e^{j\theta_x} + a_v e^{j\theta_v} \quad (13)$$

其中, θ_y 表示噪声相位谱, a_y , a_x 和 a_v 分别表示带噪声语音、纯净语音和噪声信号幅度谱。

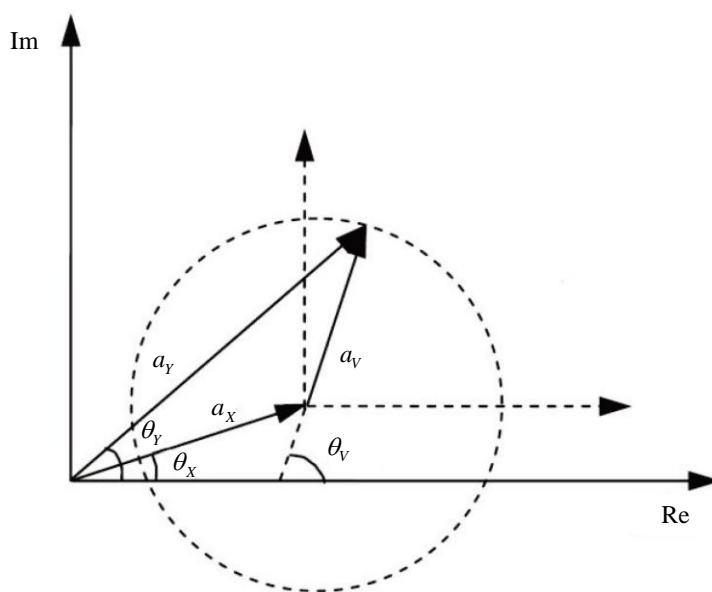


Figure 1. The geometric relationship between pure speech, noise and noisy speech spectrum

图 1. 纯净语音、噪声和带噪声语音谱之间的几何关系

根据三角函数法则很容易得到以下关系:

$$a_y^2 = a_x^2 + a_v^2 + 2a_x a_v \cos(\theta_x - \theta_v) \quad (14)$$

等式两边同时除以 a_v^2 可得:

$$\cos(\theta_x - \theta_v) = \frac{a_v^2/a_v^2 - a_x^2/a_v^2 - 1}{2\sqrt{a_x^2/a_v^2}} \quad (15)$$

其中, a_x^2/a_v^2 表示瞬时的先验信噪比, a_y^2/a_v^2 表示瞬时的后验信噪比, \hat{a}_v^2 由 MS 噪声估计算法可得[9]。

把(15)式代入(12)式, 可得:

$$\frac{SNR_{post}}{4} - 1 \leq SNR_{prio} \quad (16)$$

最后, 综合幅度约束为:

$$Y'_M = \begin{cases} |\hat{Y}_{m,k}|, & \frac{SNR_{post}}{4} - 1 \leq SNR_{prio} \\ \beta |\hat{Y}_{m,k}|, & \frac{SNR_{post}}{4} - 1 > SNR_{prio} \end{cases} \quad (17)$$

其中, β 由多次试验结果取 0.68。

4. 仿真实验结果分析

为了检验改进算法的性能, 本文运用 MATLAB 软件进行实验仿真, 选取多种来自语音库的纯净语音 sp01.wav~sp10.wav, 噪声来自于噪声库的 white, babble, pink, F16 座舱, Destroyengine 噪声, 输入信噪比分别取 5 dB, 10 dB, 15 dB, 20 dB。所有信号的取样频率均为 8 kHz/s, 仿真实验中语音信号帧长 $K = 256$, 重叠率为 50%, 系数 $\beta = 0.68$ 。

首先根据语谱图进行对比分析:

图 2 是纯净语音信号的语谱图, 图 3 是带噪语音信号的语谱图, 图 4 基于幅度判定条件约束的语音增强语谱图, 通过对比, 可以看出此算法虽然能够在一定程度上消除背景噪声, 但产生了语音失真, 并且在可行性上有较大难度。图 5 是基于本文算法的语音增强语谱图, 与其他算法相比, 降低了语音失真和背景噪声的影响, 体现了较好的语音增强性能。

另外, 通过对数谱距离(LSD), 分段信噪比(SegSNR), 短时清晰度客观测度(STOI)来[10] [11]对比分析基于本文提出的算法与幅度判定条件约束语音增强算法的性能:

从表 1 中可以看出在输入信噪比分别为 5 dB, 10 dB, 15 dB, 20 dB 时输入噪声为的 White, Pink, Babble, F16 座舱, Destroyengine 五种噪声环境下, 对数谱距离(LSD)是表示增强语音与原始语音之间接近程度的测试, LSD 测度的数值越小, 说明语音的失真的程度越低。本文的改进判定条件算法 LSD 值小于现有幅度判定条件约束算法, 其中五中不同的噪声在输入 SNR 为 15 dB 和 20 dB 时的改进效果最突出, 能达到大约 1.9 的提升效果, 由表 1 可以明显得出本文提出算法输出的语音与纯净语音更接近, 语音的增强效果更好。

分段信噪比(SegSNR)反映出每一帧信号中估计语音与原始语音之间的差距的大小, 算法的分段信噪比越大, 说明估计的增强语音与纯净语音之间的差距越小, 语音质量也就越好。从表 2 中可以看出在以上五种噪声环境下, 本文改进算法的分段信噪比要优于目前的幅度判定条件约束算法, 在输入 SNR 为 15 dB 时效果最佳, 最高可达到大约 1.4 的改进数值, 表 2 比较直观的说明本文算法具有较好的语音增强性能。

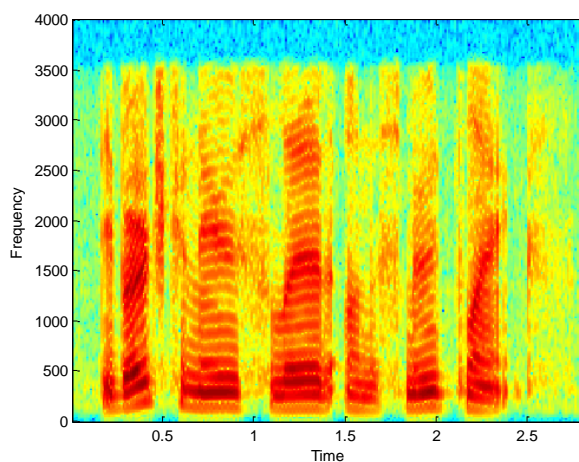


Figure 2. Speech spectrogram of pure speech signal
图 2. 纯净语音信号语谱图

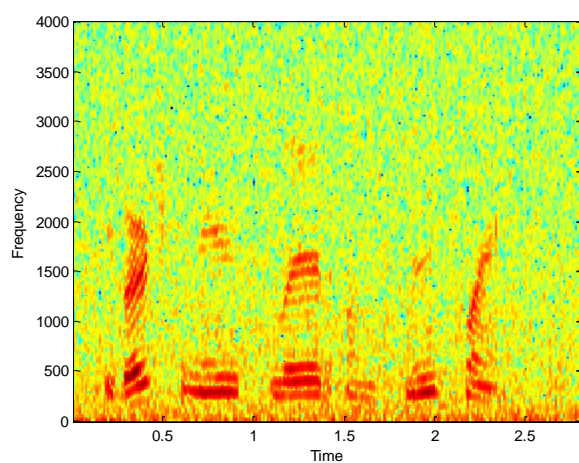


Figure 3. Speech spectrogram with noisy speech
图 3. 带噪语音语谱图

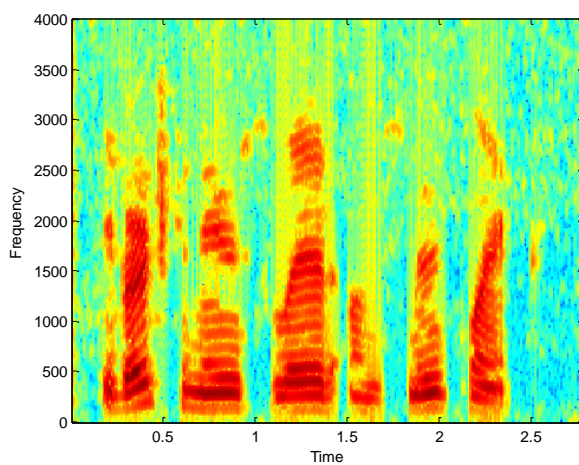


Figure 4. Speech enhancement spectrogram based on amplitude decision condition constraint
图 4. 基于幅度判定条件约束的语音增强语谱图

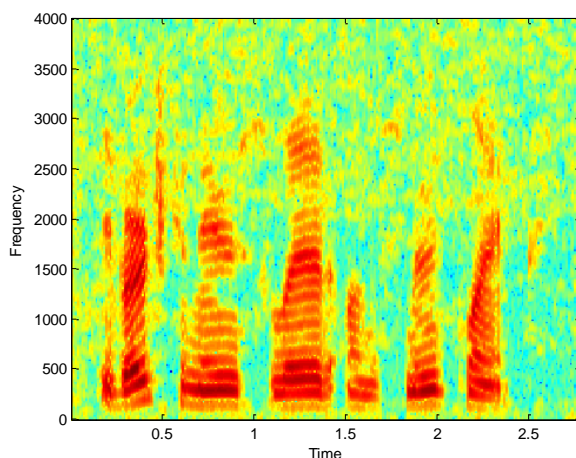


Figure 5. Speech enhancement spectrogram based on improved condition algorithm in this paper

图 5. 基于本文改进判定条件算法的语音增强语谱图

Table 1. LSD data contrast table

表 1. LSD 数据对比表

噪声类型	算法	输入SNR			
		5 dB	10 dB	15 dB	20 dB
White	幅度判定条件约束	4.7875	4.7042	5.3967	5.6444
	本文改进判定条件算法	4.7865	4.2107	3.8719	3.7042
Babble	幅度判定条件约束	4.6656	4.3504	4.6157	4.8617
	本文改进判定条件算法	3.7583	3.3973	3.2914	3.1788
Pink	幅度判定条件约束	4.7461	4.5271	4.9785	5.2380
	本文改进判定条件算法	4.3698	3.7367	3.5196	3.3860
F16	幅度判定条件约束	4.8707	4.4616	5.0221	5.1069
	本文改进判定条件算法	4.3008	3.6750	3.5566	3.3912
Destroyengine	幅度判定条件约束	4.7841	4.5347	4.8888	4.9303
	本文改进判定条件算法	4.2105	3.6366	3.3818	3.2438

短时清晰度客观测度(STOI)是表征语音增强算法在语音可懂度方面性能好坏的测度参数, STOI 测度的数值越大, 说明该算法增强后语音与原始语音之间的偏差越小, 语音可懂度就越好。从表 3 的可知本文改进判定条件算法的输出语音的可懂度有所提高, 增强效果大于之前幅度判定条件约束。因此可以认定为, 本文提出的改进算法在一定噪声环境下, 相比目前的算法不仅在可行性上有很大的进步, 而且在语音增强性能上有提升。

5. 总结与展望

本文主要研究了一种通过融合几何相位信息, 使幅度失真与先验信噪比和后验信噪比建立数学关系的噪声抑制算法。根据 loizou 等人目前所做的研究, 其大多采用估计的纯净语音与原始纯净语音的幅度谱进行比较, 实际运用起来可行性较差, 本文通过引入几何相位信息, 推导出与先验信噪比和后验信噪比有直接联系的约束条件, 实施起来简单易行。实验结果表明, 新算法在 LSD, STOI, SegSNR 评价标准下具有较好的性能, 有效地提升语音质量, 提高语音的可懂度。

Table 2. SegSNR data contrast table
表 2. SegSNR 数据对比表

噪声类型	算法	输入SNR			
		5 dB	10 dB	15 dB	20 dB
White	幅度判定条件约束	8.9086	10.0078	10.3718	10.0604
	本文改进判定条件算法	9.4968	11.4123	12.2416	11.8125
Babble	幅度判定条件约束	7.6045	9.4961	9.9694	9.8618
	本文改进判定条件算法	8.2564	10.6414	11.3207	11.2690
Pink	幅度判定条件约束	8.7192	10.2447	10.7117	9.9308
	本文改进判定条件算法	9.2945	11.4203	12.3134	11.1873
F16	幅度判定条件约束	8.4953	10.2901	10.6655	10.0598
	本文改进判定条件算法	9.0026	11.3389	12.0814	11.4116
Destroyengine	幅度判定条件约束	8.2000	9.8575	10.5018	10.2139
	本文改进判定条件算法	8.2765	10.4498	11.7899	11.6658

Table 3. STOI data contrast table
表 3. STOI 数据对比表

噪声类型	算法	输入SNR			
		5 dB	10 dB	15 dB	20 dB
White	幅度判定条件约束	0.7271	0.7838	0.8290	0.8705
	本文改进判定条件算法	0.7378	0.8142	0.8816	0.9268
Babble	幅度判定条件约束	0.7169	0.8080	0.8484	0.8792
	本文改进判定条件算法	0.7517	0.8376	0.8935	0.9337
Pink	幅度判定条件约束	0.7239	0.7957	0.8346	0.8728
	本文改进判定条件算法	0.7520	0.8316	0.8911	0.9364
F16	幅度判定条件约束	0.7835	0.8048	0.8411	0.8727
	本文改进判定条件算法	0.7533	0.8282	0.8871	0.9316
Destroyengine	幅度判定条件约束	0.7541	0.8156	0.8659	0.8967
	本文改进判定条件算法	0.7693	0.8460	0.9066	0.9479

虽然本文提出了融入几何相位信息的噪声抑制算法，但本文研究仅局限于理论和算法层面，与达到实际工程上的运用还有一定的距离。未来本文还应对噪声的估计算法进行深入研究，这势必会对实验结果的精度造成一定的影响，下一步应该深入研究噪声估计算法，以进一步提高算法数据的精确度。

参考文献

- [1] Loizou, P. (2007) *Speech Enhancement: Theory and Practice*. CRC Press LLC, Boca Raton, Florida.
- [2] Kim, G. and Loizou, P.C. (2011) Gain-Induced Speech Distortions and the Absence of Intelligibility Benefit with Existing Noise-Reduction Algorithms. *The Journal of the Acoustical Society of America*, **130**, 1581-1596. <https://doi.org/10.1121/1.3619790>
- [3] Scalart, P. and Filho, J. (1996) Speech Enhancement Based on a Priori Signal to Noise Estimation. *Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing*, Atlanta, GA, 9-9 May 1996, 629-632. <https://doi.org/10.1109/ICASSP.1996.543199>
- [4] Chung, K. (2004) Challenges and Recent Developments in Hearing Aids: Part, I. Speech Understanding in Noise, Microphone Technologies and Noise Reduction Algorithms. *Trends in Hearing*, **8**, 83-124.

<https://doi.org/10.1177/108471380400800302>

- [5] Ma, J., Hu, Y. and Loizou, P.C. (2009) Objective Measures for Predicting Speech Intelligibility in Noisy Conditions Based on New Band-Importance Functions. *The Journal of the Acoustical Society of America*, **125**, 3387-3405. <https://doi.org/10.1121/1.3097493>
- [6] Hu, Y. and Loizou, P.C. (2007) A Comparative Intelligibility Study of Single-Microphone Noise Reduction Algorithms. *The Journal of the Acoustical Society of America*, **122**, 1777-1786. <https://doi.org/10.1121/1.2766778>
- [7] Hu, Y. and Loizou, P.C. (2007) Subjective Comparison and Evaluation of Speech Enhancement Algorithms. *Speech Communication*, **49**, 588-601. <https://doi.org/10.1016/j.specom.2006.12.006>
- [8] Loizou, P.C. and Kim, G. (2011) Reasons Why Speech Enhancement Algorithms Do Not Improve Speech Intelligibility and Suggested Solutions. *IEEE Transactions on Audio, Speech, and Language Processing*, **19**, 47-56. <https://doi.org/10.1109/TASL.2010.2045180>
- [9] Rangachari, S. and Loizou, P.C. (2006) A Noise-Estimation Algorithm for Highly Non-Stationary Environments. *Speech Communication*, **48**, 220-231. <https://doi.org/10.1016/j.specom.2005.08.005>
- [10] Lu, Y. and Loizou, P.C. (2011) Senior Member, Estimator of the Magnitude-Squared Spectrum and Methods for Incorporating SNR Uncertainty. *IEEE Transactions on Audio, Speech, and Language Processing*, **19**, 1-14.
- [11] Lu, Y. and Loizou, P.C. (2008) A Geometric Approach to Spectral Subtraction. *Speech Communication*, **50**, 453-466.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-6753, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: jisp@hanspub.org