

基于混沌特性的情感语音非线性特征研究

张子恒, 孙 颖, 姚 慧

(太原理工大学 信息工程学院, 山西 太原 030024)

摘 要: 根据语音发声和传播过程中表现出的混沌特性, 首先验证了富含不同情感的语音信号是具有混沌特性的。其次采用非线性动力学理论提取了基于情感语音信号混沌特性的 3 种非线性特征: 最小时间延迟、关联维数和最大 Lyapunov 指数。最后, 设计了不同的实验验证了非线性特征的识别性能。实验中选用了 Berlin 语音库中的情感语句。采用了支持向量机进行了情感识别, 其中参数采用十倍交叉验证获得。最后, 对不同的实验结果进行了归纳分析, 对比了不同非线性特征用于识别基本情感时的优劣。

关键词: 情感语音识别; 混沌特性; 非线性特征; 支持向量机

中图分类号: TN912.34

文献标识码: A

文章编号: 1000-7180(2017)04-0065-04

Nonlinear Feature Etraction of Eotional Speech Based on Caotic Characteristics

ZHANG Zi-heng, SUN Ying, YAO Hui

(College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract: Based on the chaotic characteristics of speech utterance and propagation, the experiment firstly verifies that speech signals with different emotions have chaotic characteristics. Secondly, the experiment tries to adopt the nonlinear dynamics theory to extract three kinds of nonlinear characteristics which are based emotional speech signal. The characteristics include the minimum time delay, correlation dimension and Lyapunov exponent. Thirdly, designed different experiments to verify the recognition performance of the non-linear characteristics. Furthermore, the experiments have selected emotional statement from Berlin speech database and have used a support vector machine to identify the emotion. And the parameters are obtained by ten-fold cross-validation. Finally, through the summarizes and analyzes of the experimental results from different experiments. Then it compares the relative merits of different nonlinear characteristics which were used to identify the basic emotions.

Key words: emotional speech recognition; chaotic characteristic; nonlinear features; support vector machine

1 引言

语言是人类沟通最便捷、自然的方式,而语音是语言的载体。现如今人机交互技术飞速发展,为了实现更智能、完善的人机对话,情感语音识别技术应运而生。而情感特征的提取是情感语音识别的主要部分。目前,大部分研究学者提取的有效情感特征参数,大多集中于经过线性处理后得到的声学特征。但是已经有研究学者证明,语音信号的产生和传播过

程是一个复杂的非线性混沌过程^[1],而非线性动力学理论作为混沌理论的分支,在语音信号处理中已经得到了广泛应用^[2-4]。复旦大学闫润成针对语音因素的混沌特性进行了研究,提出了基于信号递归度分析的端点检测方法,取得了较为理想的研究成果^[2]。北京交通大学李响利用了相空间重构的方法对语音信号建模,分别提取了该模型下司机应答语音信号中的混沌特征和传统的语音特征,提高了语音检测疲劳驾驶的准确性^[3]。太原理工大学孙颖根

据语音发声过程中表现出的混沌特性与情感关联性进行研究,提出了几种非线性特征,并将其与声学特征结合,应用于情感语音识别,验证了这些非线性特征融合后对情感区分的有效性^[4].虽然对于语音的混沌特性研究以及对基于混沌特性提取的语音非线性特征的研究有很多,但是并没有单独验证每种非线性特征有效性的相关研究.因此本文首先采用了时间序列分析技术验证了情感语音信号是具有混沌特性的,然后选用 Berlin 语音库中的情感语句对文献^[4]中提到的 5 种非线性特征^[5]中的 3 种进行单独验证,并通过实验,对比提取的 MFCCs^[6]识别结果,逐个验证了最小延迟时间、关联维数和最大 Lyapunov 指数 3 种非线性特征在区分语音情感方面的有效性.

2 语音信号的混沌特性及非线性特征提取

混沌理论是指非线性系统在一定参数条件下展现出周期运动与非周期运动的共存,最终形成某种非周期有序运动的理论.用于描述混沌特性的基本特征有两个,分别是高维相空间重构后的混沌吸引子自身结构具有分形维特征和初始条件对系统影响大^[7].若时间序列具有以上两种特征,则可以判断该时间序列本身是有混沌特性的.基于上述理论,本文分别从功率谱分析^[8]、相空间重构^[9]两个方面来研究情感语音信号的混沌特性^[10].并在高维空间提取情感语音信号的多维动态非线性特征,以作为识别网络的输入.

2.1 功率谱方法

时间序列的周期性或无序性无法从时域波形中直观地观察到,但是通过功率谱却可以发现其规则性.基于功率谱方法判断时间序列是否具有混沌特性,主要从峰值个数以及宽谱特性两个方面判断.如果功率谱图中具有有限个峰值,则说明该时间序列是周期序列;谱图中无明显的峰,呈现出“宽谱”特性,这说明了该时间序列是具有湍流或者混沌特性的.

由于之前已经对实验所用的不同语义的语音进行过相关实验验证,实验证明不同语义的情感语音信号均具有宽谱特性,所以本文只选取 Berlin 语音库中的“happy”、“neutral”、“sad”以及“fear”四种情感中的一句语音进行功率谱分析.从图 1 可以看出,四种情感的语音信号功率谱均呈现出宽谱特性,没有特别的峰值.因此,情感语音信号是具有混沌特性的.

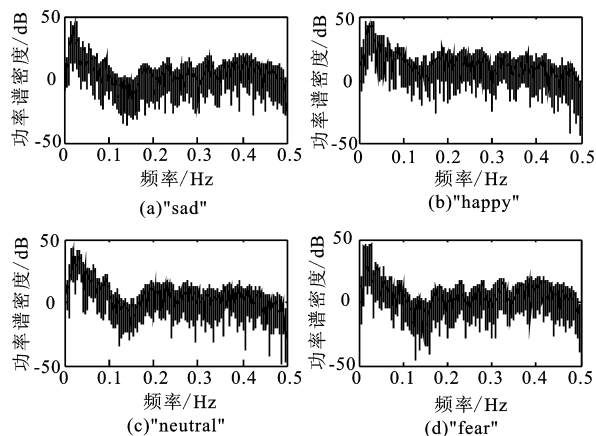


图 1 不同情感语音信号的功率谱仿真图

2.2 相空间重构方法

相空间重构是计算具有混沌特性的时间序列特征量的前提,这些特征量包括最大 Lyapunov 指数以及关联维数等.根据 Taken's 嵌入定理,选择合适的嵌入维数,实现语音在高维空间的等价映射,在高维空间下通过对信号的演化轨迹进行测量和预测就可以得到该信号的定性性质,而在高维空间提取的多维动态信息就是混沌特性.

图 2 分别给出了不同情感语音信号的 3 维相空间重构示意图(选择的语音信号同功率谱方法中相同).从图中可以观察到不同情感的语音信号在高维空间中具有不同的结构;“fear”情感语音信号与“sad”情感语音信号在相空间中的结构相似性较高;“anger”情感语音信号在相空间中的结构与其他情感的结构具有较为明显的差异;所有情感语音信号在相空间的结构相较于一维时域波形图的结构呈现出更显著的差异.

2.3 语音信号的非线性特征

本文以相空间重构方法为基础,实现了情感语音信号与非线性动力学模型的结合.并对映射在高维空间的语音信号提取非线性特征,分别是:通过平均互信息法计算得到的最小延迟时间 τ ,表示两变量间的相关性最小;使用 G-P 算法计算得到的非线性表征量关联维数;使用 Wolf 方法求得的反映相空间中混沌程度最大的最大 Lyapunov 指数.下文将分别提取上述 3 种非线性特征作为识别网络的输入,通过实验,具体论证其应用于情感语音识别的可行性及有效性.

3 实验

3.1 情感语音数据库处理

为了客观全面地评价所研究的非线性特征性

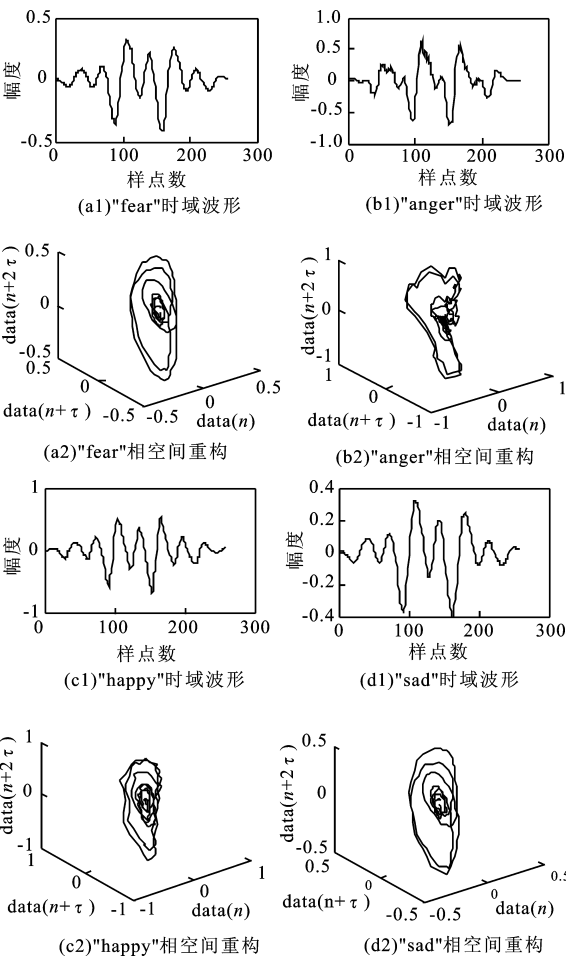


图 2 不同情感语音的时域波形和相空间重构

能,实验选用 Berlin 语音库中的情感语句作为语料来源. 为了避免由于语音时长等因素造成的影响,基于语音时长近似相同的原则,对数据库语句进行了筛选. 最终得到 Berlin 语音库 330 句用于测试和训练的语句,如表 1 所示.

表 1 情感语音实验语句				
实验语句	EMO-DB			
	Happy	Sad	Anger	Neutral
训练	47	42	55	53
测试	24	20	27	26
总共	71	62	82	79

3.2 情感语音信号预处理

为了提取情感语音信号特征,需要先对信号进行预处理,主要包括端点检测、高通滤波预加重和加窗分帧. 其中端点检测采用的是基于过零率和能量的双门限方法,预加重采用的高通滤波器系数设置为 0.97,加窗分帧处理时帧长设置为 256,帧移设置为 128.

3.3 情感语音特性的提取

预处理之后,依次提取每帧语音信号的 3 种非

线性特征及 MFCCs. 提取非线性特征时,首先利用 Taken’s 的相空间重构法通过互信息法和邻接误差法分别得到每一帧语音信号的延迟时间 τ 和嵌入维数 m 两个参数,然后在重构模型下依次提取非线性特征.

3.4 最小延迟时间特征

为了验证最小延迟时间特征是否可以区分语音情感,设计实验将最小延迟时间特征单独作为支持向量机的输入从而得到识别结果. 识别结果如表 2 所示.

从表 2 可以看出,最小延迟时间特征针对 Berlin 语音库中 4 种情感的平均识别率达到 77.61%,识别性能较好. 从单独情感的识别率来看,“Sad”情感的识别率为 90.00%,识别率最高,而“Happy”情感的识别率最低,仅仅为 58.33%. 从这两种情感的误判率可以看出,最小延迟时间特征针对“Anger”和“Happy”两种情感之间的判别较为模糊. 但是针对“Neutral”与“Anger”、“Anger”与“Sad”的两两情感之间的判别效果明显. 因此,最小延迟时间特征可以作为语音信号情感区分的特征.

表 2 最小延迟时间在 Berlin 语音库中的识别率				
情感	Neutral	Anger	Happy	Sad
Neutral	76.92	0.00	15.38	7.70
Anger	0.00	85.19	14.81	0.00
Happy	4.17	37.50	58.33	0.00
Sad	10.00	00.00	0.00	90.00

3.5 关联维数特征

为了验证关联维数特征是否可以区分语音情感,设计实验将关联维数特征单独作为支持向量机的输入从而得到识别结果. 识别结果如表 3 所示.

表 3 关联维数在 Berlin 语音库中的识别率				
情感	Neutral	Anger	Happy	Sad
Neutral	69.23	11.54	19.23	0.00
Anger	11.11	77.78	14.81	0.00
Happy	12.50	20.83	66.67	0.00
Sad	20.00	0.00	0.00	80.00

从表 3 可以得到,关联维数特征针对 Berlin 语音库四种情感的平均识别率为 73.42%. 单独的情感语音识别中最低的“Happy”情感识别率达到了 66.67%,最高的“Sad”情感识别率达到了 80.00%. 从不同情感的误判率来看,关联维数可以明显区分“Sad”与“Anger”、“Happy”. 但是从其他情感的误判率来看,仍没有达到最为理想的效果. 从“Anger”

情感的误判率来看,27 句测试语句中有 4 句被判断为“Happy”情感;从“Happy”情感的误判率来看,24 句测试语句中有 5 句被判断为“Anger”. 因此可以发现,该特征在区分“Anger”与“Happy”情感是存在一定的误差. 综合分析,关联维数特征可以作为语音信号情感区分的有效特征.

3.6 最大 Lyapunov 指数特征

为了验证最大 Lyapunov 指数特征是否可以区分语音情感,设计实验将该特征单独作为支持向量机的输入从而得到识别结果. 识别结果见表 4 所示.

表 4 最大 Lyapunov 指数在 Berlin 语音库中的识别率

情感	Neutral	Anger	Happy	Sad
Neutral	84.62	3.85	3.85	7.69
Anger	7.41	81.48	11.11	0.00
Happy	8.33	8.33	83.33	0.00
Sad	15.00	0.00	0.00	85.00

从表 4 可以得到最大 Lyapunov 指数特征针对 Berlin 语音库 4 种情感的平均识别率为 83.61%,识别性能相较于其他四类非线性特征达到最优. 并且从不同情感的单独识别率来看,4 种情感的识别率达到了 80.00%以上,最低识别率为 81.48%. 值得注意的是,针对“Happy”情感的正确识别率达到了 83.33%. 误判为“Anger”情感的概率明显减小,仅 8.33%. 这说明了最大 Lyapunov 指数特征可以较为有效地区分“Happy”和“Anger”两种情感. 此外,针对“Sad”情感与“Anger”、“Happy”情感的区分,最大 Lyapunov 指数特征依旧体现出了优越性. 因此,最大 Lyapunov 指数特征可以作为语音信号情感区分的特征.

3.7 分析与总结

综上所述,3 类特征针对 4 种情感的识别结果整体上较好. 表 5 给出了 3 类非线性特征特征以及 MFCCs 的识别结果作为对比. 总体来说,基于情感语音混沌特性提取的非线性特征均表现出了较好的识别结果. 从平均识别结果来看,识别率最低达到 73.42%,最高达到 83.61%,最大 Lyapunov 指数特征性能最优. 从每个特征的识别结果来看,非线性特征针对“Neutral”情感与“Happy”情感的区分都体现出了优越性,并且非线性特征及 MFCCs 均可以有效地区分“Sad”情感. 此外,最大 Lyapunov 指数特征作为情感识别特征可以对“Happy”与“Anger”情感的误判情况明显改善,与 MFCCs 相比,虽然个别情感的识别没有 MFCCs 高,但是平均识别率却

优于 MFCCs. 基于以上的实验数据显示,非线性特征是区分不同情感的有效特征.

表 5 非线性特征在 Berlin 语音库中的识别率

	Neutral	Anger	Happy	Sad	平均
最小延迟时间	76.92	85.19	58.33	90.00	77.61
关联维数	69.23	77.78	66.67	80.00	73.42
最大 Lyapunov 指数	84.62	81.48	83.33	85.00	83.61
MFCCs	76.92	88.89	70.83	95.00	82.91

如图 3 所示,将上述由 3 种非线性特征及 MFCCs 得到的平均识别率通过柱状图表达,能更直观地看到每种特征参数的优劣以及它们的有效性.

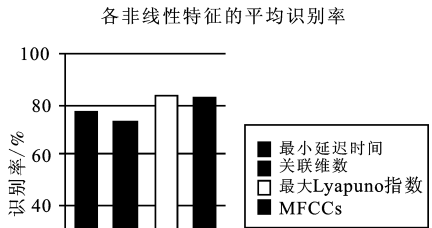


图 3 各非线性特征的平均识别率

4 结束语

本文以混沌理论为基础,验证了语音信号的混沌特性,并结合非线性动力模型,单独提取了 3 种非线性特征,以 Berlin 语音库为情感语料来源,以支持向量机为识别网络,实验分别对比了每种非线性特征用于基本情感区分时的优劣,并与 MFCCs 的识别结果作比较,验证了所提取的 3 种非线性特征分别单独用于情感语音识别的有效性.

参考文献:

[1] Banbrook M, McLaughlin S, Mann I. Speech characterization and Synthesis by nonlinear methods [J]. IEEE Transactions on Speech and Audio Processing, 1999, 7(1): 1-17.

[2] Patricia H, Jesus B A, Miguel A F, et al. Global selection of features for nonlinear dynamics characterization of emotional speech [J]. Cognitive Computation, 2013, 5(4): 517-525.

[3] 李响,谭南林,李国正,等. 一种应用语音多特征检测驾驶疲劳的方法[J]. 仪器仪表学报, 2013, 34(10): 2231-2237.

[3] Heinzellnan,Chandrakasan A P,Bal Krishnan. An application specific protocol architecture for wireless networks [J]. IEEE Transactions on Wireless Communications, 2012, 1(4):660-670.

[4] Ali M,Stewart B G. Congestion adaptive multipath routing for load balancing in mobile WSN [J]. Innovations in Information Technology, 2012, 21 (16): 305-309.

[5] Zhao T. A framework of physical layer technique assisted authentication for vehicular communication networks [J]. Science China (Information Sciences), 2014, 10(12):1996-2004.

[6] Wang D, Zhang Q, Liu J C. Partial network coding: theory and application for continuous sensor data collection[C] // Proc. of the 14th IEEE International Workshop on Quality of Service, IEEE,2012, 35(10): 93 - 101.

[7] 闫新芳,王志龙,闫新申,等. WSN 中基于梯度场节点定位算法的维护更新[J]. 传感器与微系统,2011,30 (8):56-58.

[8] 罗守山,杨文川. 基于矢量定位技术的 wsn 节点定位的模型及研究[J]. 北京邮电大学学报,2013,31(3): 5-9.

[9] Wang Y H. Evidence-based trust a mathematical model geared for multivalent systems [J]. ACM Transactions on Autonomous and Adaptive Systems, 2013, 5 (3):1-25.

[10] Hao Xiaochen, Jia Nan, Liu Bin. Multi-path optimizing routing protocol based on predicting congestion for wireless sensor network [J]. Journal of Electronics & Information Technology, 2011, 33(5): 1261-1265.

[11] Luo Shoushan, Yang Wenchuan, Xiao Qiang. Model and research of WSN node localization based on vector localization technology [J]. Journal of Beijing University of Posts and Telecommunications, 2013,31 (3): 5-9.

[12] Nguyen D, Tran T, Nguyen T. Wireless broadcast using network coding [J]. IEEE Trans. on Vehicular Technology, 2013, 58(2): 914-925.

[13] Jolliffe D, Tran T, Nguyen T. Node location in wsn [J]. Vehicular Technology, 2014, 58(2): 914-925.

作者简介:

黄煜栋 男,(1981-),硕士,副教授. 研究方向为网络通信、人工智能. E-mail:hugydong1981gd@sina. com.

郝 平 男,(1961-),硕士,教授. 研究方向为移动网络、模式识别与数据挖掘.

(上接第 68 页)

[4] 孙颖,姚慧,张雪英,等. 基于混沌特性的情感语音特征提取[J]. 天津大学学报(自然科学与工程技术版), 2015,48(8):681-685.

[5] 吕金虎,陆君安,陈士华. 混沌时间序列分析及其应用[M]. 武汉:武汉大学出版社,2002:20-60.

[6] 丁伟. 基于 MFCC 说话人识别算法研究[D]. 合肥:安徽大学,2009.

[7] 赵力,黄程韦. 实用语音情感识别中的若干关键技术 [J]. 数据采集与处理, 2014, 29(2): 157-170.

[8] 韩贵丞,李锋. 股票指数信号的似混沌特性研究[J]. 计算机工程与设计, 2011, 32(2): 711-715.

[9] 陈鑫. 相空间重构在语音情感识别中的研究[D]. 长沙:长沙理工大学,2014.

[10] Faúndez Z M, McLaughlin S, Esposito A, et al. Non-linear speech processing: Overview and applications [J]. Control and Intelligent Systems, 2002, 30(1): 1-10.

作者简介:

张子恒 男,(1991-),硕士研究生. 研究方向为语音信号处理.

孙 颖(通讯作者) 女,(1981-),讲师,硕士生导师. 研究方向为语音信号处理. E-mail:tyutsy@163. com.

姚 慧 女,(1991-),硕士研究生. 研究方向为语音信号处理.