

关于 DIVA 模型中语速对语音生成影响的研究

刘燕燕, 张少白

(南京邮电大学 计算机学院, 江苏 南京 210003)

摘要:着重研究语速对语音生成的影响。过去有不少人对此进行了研究,指出语速的变化对元音和辅音的运动产生了不同的影响,但是具体是什么样的影响并没有明确指出。基于 DIVA 模型,通过修改 ODV 和 AVV 活动的公式对语速对元音和辅音运动产生的影响进行了研究和探讨。仿真实验表明,语速的增加导致了辅音运动速度的增加,但元音运动速度只增加了少量的值,有时甚至还会减小。最终得出的结论是:尽管语速对元音和辅音的运动速度产生了不同的影响,但最大速度与运动距离比值的增量却是大致相同的。

关键词:DIVA 模型;语音生成;凸形区域;语速

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2011)12-0033-03

Study of Speaking Rate Effects to Production of Speech Sounds in DIVA Model

LIU Yan-yan, ZHANG Shao-bai

(Computer College, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)

Abstract: Mainly study the speaking rate effects to the production of speech sounds. Many researchers in the past twenty years has investigated this, and they pointed that changes in speaking rate have differential effects for the movements corresponding to vowels and consonants of speech sounds, but they didn't point what the exact effects are. Based on the DIVA (Directions Into Velocities of Articulators) model, discussed and investigated these effects by modifying the ODV (Orosensory Direction Vector) and AVV (Articulator Velocity Vector) equations. The simulation results show that increasing rate causes an increase in the velocities of movements corresponding to consonantal gestures, but it causes less of an increase, or even a decrease, in the velocities of movements corresponding to vowel gestures. The final conclusion is: despite changes in speaking rate have differential effects for the movements corresponding to vowels and consonants, the increase in ratios of maximum velocity to movement distance for the vowel and consonant gestures are almost same to vowels and consonants.

Key words: DIVA model; speech production; convex region; speaking rate

0 引言

在过去的二十年里,有不少人就语速的变化对语音生成的影响进行了研究。有些研究者发现不同的说话者采用不同的策略来增加语速:有些人依赖于提高速度来增加语速,称为速度策略;而有些人则依赖于动作幅度的减小,称为振幅策略。

速度策略已经应用于 DIVA 模型(是一种为了生成单词、音节或者音素,被用来控制模拟声道运动的自适应网络模型^[1]),即控制动作命令的 Go 信号。增强

Go 信号就增加了运动速度,也就等于增加了语速。振幅策略可以通过在 DIVA 模型中改变凸形目标区域的大小来实现,如图 1 所示。语速越慢,凸形目标区域越小,产生的音位越清晰,越准确。

增加了运动速度就增加了凸形目标区域的大小^[2]。在语速变快的语音生成期间增加凸形目标区域的大小是一种在语音生成领域中以速度交换精确度的情况,即语速越快,凸形目标区域越大,生成的语音精确度越小。

过去的这些研究得出了一个共同的结论,就是语速的变化对语音的生成产生了不同的影响^[3](Kuehn and Moll, 1976; Ostry and Munhall, 1985),但具体是什么样的影响并没有明确指出。文中通过修改 ODV (Orosensory Direction Vector) 和 AVV (Articulator Ve-

收稿日期:2011-05-10;修回日期:2011-08-15

基金项目:国家自然科学基金(61073115)

作者简介:刘燕燕(1986-),女,山东德州人,硕士,研究方向为人工智能;张少白,教授,研究方向为人工智能。

locity Vector) 活动的公式对语速对元音和辅音运动产生的影响进行了研究和探讨。

1 语速对语音生成的影响

1.1 语速对元音运动和辅音运动的影响

对一个语音而言,在 DIVA 模型中可以很容易地实现凸形目标区域的收缩:简单地向所有的 ODV (Orosensory Direction Vector) 细胞中增加一个很小的正值输入^[4]。因为向所有的 ODV 细胞中加入了相同的输入,所以这个输入被称作是非特定的。为了弄清楚非特定的输入使凸形目标区域缩小的原因,就来看看与 orosensory (体觉) 空间单一维对应的一对 ODV 细胞。因为只有正的 ODV 活动才能产生运动,而且在此维方向上的语音凸形目标区域的大小与 orosensory 反馈值的范围相符。如果一个正的非特定的输入加到了该对细胞上,orosensory 反馈值的范围使得两个 ODV 细胞产生非正的活动,则凸形目标区域就会减小。如果将相同的非特定的输入加到所有的 ODV 细胞上,凸形区域目标就会沿着各个维向中间收缩,如图 1 所示。

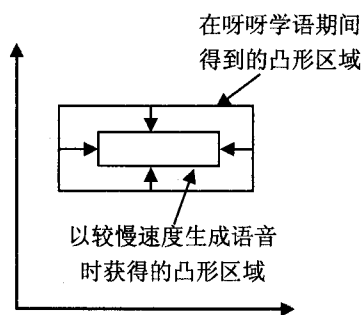


图1 改变语速的振幅策略可以通过收缩凸形目标区域来实现

因为声调越大的活动会产生越小越精确的目标,因此,输入的大小应该与运动速度有关。为此,可以修改 ODV 细胞活动的公式,如下:

$$d_{j+} = \sum_i s_i z_{ij+} - f_{j+} + R(1 - G) \quad (1)$$

$$d_{j-} = \sum_i s_i z_{ij-} - f_{j-} + R(1 - G) \quad (2)$$

其中, G 是 G_0 信号的值(0 到 1 之间), R 是一个参数,与说话者相对于速度策略而言更喜欢振幅策略的程度相符。对 ODV 细胞非特定的输入就是 $R(1 - G)$ 。向 ODV 细胞增加正的输入值会产生不可预期的副作用:能够使此对细胞产生积极的活动,这就像命令 orosensory 变量既增加又减少一样。这个问题比较简单,可以通过修改 AVV (Articulator Velocity Vector) 活动的公式来解决:

$$a_{k+} = \sum_j [[d_{j+}]^+ - [d_{j-}]^+]^+ w_{j+k+} + \sum_j [[d_{j-}]^+ - [d_{j+}]^+]^+ w_{j-k+} \quad (3)$$

$$a_{k-} = \sum_j [[d_{j+}]^+ - [d_{j-}]^+]^+ w_{j+k-} + \sum_j [[d_{j-}]^+ - [d_{j+}]^+]^+ w_{j-k-} \quad (4)$$

这些公式表明了 ODV 阶段敌对细胞间的竞争作用。

在 DIVA 仿真中,同时运用了速度和振幅策略来提高语速^[5]。在模型中通过改变参数 R 的值来表示说话者所采取的策略。当 R 值接近于 0.0 时,表明说话者更倾向于使用速度策略,当 R 值较大时则表示说话者更倾向于振幅策略(见仿真实验 1)。

实验 1 的仿真结果表明:当 R 为 0.0 时(模型优先采用速度策略,这是振幅策略完全没用到的极端的情况),对于元音和辅音而言运动的最大速度都增加了。当 $R = 0.2$ 模型优先地采用振幅策略时,尽管模型采用相同的策略生成元音和辅音,但随着语速的增加,元音运动速度减少了,而辅音的运动速度却增加了。这就证明了元音和辅音具有不同的控制结构。

此实验还得出了一个重要的结论,就是无论采取什么样的策略,元音和辅音运动的最大速度与运动距离的比值都会随着语速的增加而增加。采取振幅策略时,即使最大速度减小了,比值也会增加。

1.2 元音和辅音表现出不同行为的原因

为什么模型中元音和辅音表现出了不同的行为?答案就在于在呀呀学语期间学到的这两种声音类型的凸形目标区域性质的不同。即使沿着比较重要的 orosensory 维,如舌头相对于上颌的位置,元音也能在相当大的范围里产生。而辅音却具有非常严格的控制以确保完全关闭或者摩擦。在呀呀学语期间,模型对凸形目标区域的学习过程反映了这些特点,如图 2 所示。现在来看当这两个凸形区域根据公式 1 和公式 2 在更慢的语速中向中心缩小时发生的情况。声道在慢语速和快语速中移到目标的距离分别标为 D_s 和 D_f 。当这两个凸形目标区域缩小时,元音的凸形目标区域缩小的距离(即 $D_s - D_f$)比辅音缩小的距离要大得多,这种趋势在人类语音中很常见,被称之为元音缩小^[6]。

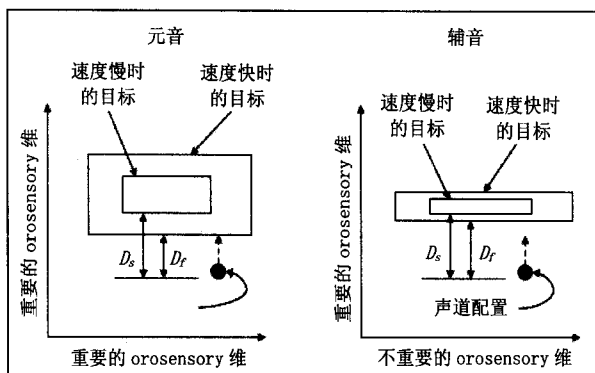


图2 凸形目标区域的缩小对元音和辅音产生的不同影响

此外,早期的研究结果说明了在 DIVA 模型中运动速度与运动距离是直接相关的。因为在语速较快的发音过程中,运动距离对于元音来说,相对于辅音而言减少的要多得多,当采用振幅策略时,在语速较快的发音过程中元音发音的速度增加的很少甚至还减小了。

1.3 语速与最大速度/运动距离的关系

在 DIVA 中,Go 信号和运动距离间的关系(如公式 5 所示)^[7]经重整后得出公式 6:

$$v_k = G \times [a_{k+} - a_{k-}]$$

(5)

$$\frac{v_k}{[a_{k+} - a_{k-}]} = G$$

(6)

其中, v_k 是沿着第 k 维发音自由度的速度, G 为 Go 信号, a_{k+} 和 a_{k-} 是与第 i 维发音自由度相符的敌对的 AVV 活动。由公式 1 到 3 可以看出,活动的大小 a_{k+} 和 a_{k-} 反映了到当前目标的距离。公式 6 表明,速度与距离的比例随着 Go 信号而变化。因为提高语速是通过增强 Go 信号来实现的,显然,速度与距离的比值会随着语速的增加而增加。在元音运动采取振幅策略的情况下,由于运动距离的减小(减小的幅度超过了 Go 信号增加的幅度),于是就发生了语速越快、发音速度越小的情况;因此,尽管最大速度降低了,但是最大速度与距离的比值却增加了。

此外,在 DIVA 仿真中采取振幅策略的情况下,随着语速的增加,元音和辅音的下巴运动幅度都减小了。这种情况通过实验 2 的仿真实验进行了说明。这种现象随着参数 R 的减小消失了,当 R 等于 0.0 时,无论语速快慢动作幅度都是相等的。有些说话者语速的快慢对下巴移动的幅度几乎没有什么影响。此处的仿真结果表明这些说话者可能采用的是速度策略,而 Gay 等人的研究对象采用的是振幅策略^[8]。

文中的研究表明:辅音运动速度随着语速的增加而增加,而元音运动速度增加的很少,甚至还会降低。尽管语速对元音和辅音的运动速度产生了不同的影响,但最大速度与运动距离比值的增量却是大致相同的。此外,说话者采取什么样的策略来提高语速由唯一的参数 R 来决定^[9]。

2 仿真实验

文中应用 MATLAB 进行仿真实验,内容如下:

实验 1 在 pap 发音时,语速对元音和辅音运动影响的仿真实验。

表 1 显示了在两个不同的速度和不同的 R 参数值的情况下生成词语 pap 时模型的仿真结果。用于生成语音的最大动作速度在每个表的前两行给了出来。 R 为 0.0 时,模型优先采用速度策略^[10],这是振幅策略完全没用到的极端的情况,可以看到对于元音和辅音

而言运动的最大速度都增加了。
但是当 $R=0.2$ 时模型优先地采用振幅策略的情况下,尽管模型采用相同的策略生成元音和辅音,但随着语速的增加,元音运动速度减少了,而辅音的运动速度却增加了^[11]。这就是元音和辅音具有不同控制结构的证据。

通过实验结果还可以观察到无论采取什么样的策略,元音和辅音动作的最大速度与运动距离的比值都会随着元音和辅音语速的增加而增加。采取振幅策略时,即使最大速度减小了,比值也会增加。

表 1 在 pap 发音时,语速对元音和辅音运动影响的仿真结果

测量值	速度策略($R=0.00$)		改变的比例
	慢速比($G=0.5$)	快速比($G=1.0$)	
元音最大速度	0.010	0.019	+90%
辅音最大速度	0.088	0.176	+100%
元音最大速度/距离	0.011	0.020	+81%
辅音最大速度/距离	0.020	0.040	+100%
振幅策略($R=0.20$)			
元音最大速度	0.034	0.019	-44%
辅音最大速度	0.144	0.176	+22%
元音最大速度/距离	0.010	0.020	+100%
辅音最大速度/距离	0.025	0.040	+60%

实验 2 采用振幅策略发出 apapapa 时下巴移位的仿真实验。

此实验的仿真结果说明了随着语速的增加,元音和辅音的下巴运动幅度都减小了。但是这种现象随着参数 R 的减小消失了,当 R 等于 0.0 时,在语速慢或快时,动作幅度是相等的。还有一些说话者语速的快慢对下巴移动的幅度几乎没有什么影响,此仿真结果表明这些说话者可能采用的是速度策略,而 Gay 等人的研究对象采用的是振幅策略^[12]。

表 2 采用振幅策略的情况下,发出 apapapa 时下巴移位的仿真结果

动作类型	下巴的平均位移	
	慢速比($G=0.5$)	快速比($G=1.0$)
元音	48.7	37.8
辅音	43.4	34.1

3 结束语

文中研究了语速对语音生成的影响。在语速变慢时,对 ODV 细胞进行的非特定的输入可以缩小凸形目标区域的大小从而获得更清楚的发音,这种现象是与元音缩小有关的数据相一致的。实验结果表明:即使在生成元音和辅音时采用了相同的过程,提高语速时也会对这两种声音类型造成不同的影响:辅音运动速度随着语速的增加而增加,而元音运动速度增加的很

型的交易成功率有了较大的提升,说明本信任模型能够较有效地提高交易成功率,有效地抑制恶意节点的行为,改善网络性能。实验结果证明了该信任模型的可行性和有效性。

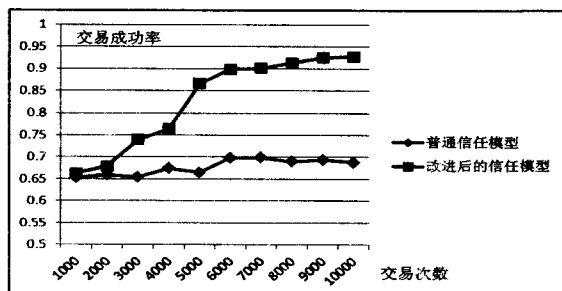


图 2 普通模型和改进后的信任模型交易成功率

4 结束语

文中针对网络安全的特点,研究了网络安全中的信任关系和已有的网格信任模型,并对其进行改进,将信任值分类细化,并分别给出了域内直接信任值、推荐信任值及域间直接信任值、推荐信任值的计算方法,在计算值时引入时间衰减因子和加权因子,使得节点在抉择时更为客观。通过模拟实验证明了该模型的可行性,下一步的工作是给出合理的时间衰减函数以及加权因子的选取方法,同时完善自动更新机制,考虑节点可能属于多个信任域的情况,进一步提高交易的成功率,降低交易失败的风险,使得此网格信任模型更加实现网络的动态性和安全性。

(上接第 35 页)

少,甚至还会降低。尽管语速对元音和辅音的运动速度产生了不同的影响,但最大速度与运动距离比值的增量却是大致相同的。

参考文献:

- [1] Zhang Shaobai, Ruan Xiaogang, Cheng Xiefeng. A new constructing method of cerebellum model applying to DIVA model [C]//Control and Decision Conference, CCDC 2009. Guilin, Chinese; [s. n.], 2009.
- [2] Lacer F, Klintfors E, Gustavsson L. Multisensory information as an improvement for communication systems' efficiency [C]//Proceedings from Fonetik 2005. Gothenburg, Sweden; [s. n.], 2005.
- [3] Guenther F H, Ghosh S S. A model of cortical and cerebellar function in speech [C]//Proceedings of the XVth International Congress of Phonetic Sciences. Barcelona; [s. n.], 2003: 169-173.
- [4] Guenther F H, Perkell J S. A neural model of speech production and its application to studies of the role of auditory feed-

参考文献:

- [1] 朱春鸽. 网络环境下信任的研究与应用 [D]. 北京: 北京邮电大学, 2007.
- [2] Beth T, Borchering M, Klein B. Valuation of trust in open networks [C]//Proceedings of the European Symposium on Research in Security (ESORICS). Brighton; Springer-verlag, 1999: 59-63.
- [3] 黄刚, 王汝传, 田凯. 基于 RBAC 策略的可信网络访问控制模型 [J]. 计算机应用研究, 2010, 27(4): 1473-1476.
- [4] 王莉苹, 杨寿保. 网络环境中的一种信任模型 [J]. 计算机工程与应用, 2004, 40(23): 50-53.
- [5] 王东安. 网络计算中信任模型及其应用研究 [D]. 北京: 中国科学院研究生院, 2006.
- [6] 杨璐. 基于信任域的网格信任模型研究 [D]. 西安: 西北大学, 2008.
- [7] 邓勇, 张琳, 王汝传, 等. 网络计算中基于信任度的动态角色访问控制的研究 [J]. 计算机科学, 2010, 37(1): 51-54.
- [8] 张亮. 网络安全中信任关系以及访问控制的研究 [D]. 合肥: 合肥工业大学, 2007.
- [9] 李文娟, 王晓东, 傅仰耿, 等. 几种网格信任模型的研究 [J]. 福州大学学报 (自然科学版), 2006, 34(1): 189-193.
- [10] 王晓华. 网络环境下基于行为的信任模型研究 [D]. 济南: 山东大学, 2009.
- [11] 姚寒冰. 网络环境中访问控制与信任模型研究 [D]. 武汉: 华中科技大学, 2006.
- [12] 高承实, 王建政, 沈昌祥, 等. 网络环境下层次信任模型研究 [J]. 计算机工程与应用, 2007, 43(13): 132-136.

back in speech [M]//Speech Motor Control in Normal and Disordered Speech. Oxford: Oxford University Press, 2004: 29-49.

- [5] 杜华英, 赵跃龙. 人工神经网络典型模型的比较研究 [J]. 计算机技术与发展, 2006, 16(5): 1-3.
- [6] Guenther F H, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements [J]. Psychol Rev, 1998, 105: 611-633.
- [7] 刘雅琴, 智爱娟. 几种语音识别特征参数的研究 [J]. 计算机技术与发展, 2009, 19(12): 67-70.
- [8] 王晓敏, 刘希玉, 戴芬. BP 神经网络预测算法的改进及应用 [J]. 计算机技术与发展, 2009, 19(11): 64-67.
- [9] 林联明, 王浩, 王一雄. 基于神经网络的 Sarsa 强化学习算法 [J]. 计算机技术与发展, 2006, 16(1): 30-32.
- [10] 葛哲学, 孙志强. 神经网络理论与 MATLAB R2007 实现 [M]. 北京: 电子工业出版社, 2007.
- [11] 钟淑瑛, 李陶深. 基于 MATLAB 的 BP-LVQ 神经网络组合分类模型 [J]. 计算机技术与发展, 2006, 16(2): 114-116.
- [12] Guenther F H. Neural models of adaptive sensory-motor control for flexible reaching and speaking [D]. Boston, MA: Boston University, 1992.