

## 利用逆滤波和相平面获取高自然声门波的研究 · 论文 ·

胡 琼, 赵春宇

(上海交通大学 电子信息与电气工程学院, 上海 200240)

【摘 要】利用迭代自适应逆滤波的方法对传声器提取的声音信号进行了逆滤波, 滤除了语音信号的共振信息, 并在不借助其他设备的情况下利用声门波在相平面的图形和相应参数来评估和改善声门波的质量。通过实验表明由这种方法得到的声门波合成出的语音有很高的自然度, 可以直接作为语音合成浊音的激励源, 为后期结合统计学语音合成提供了基础。

【关键词】逆滤波; 相平面; 声门波; IAIF; 语音合成

【中图分类号】TN912.33

【文献标识码】A

Study of Improving the Quality of Glottal Flow Using Inverse Filtering and Phase Plane

HU Qiong, ZHAO Chunyu

(Dept. of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

【Abstract】Using the iterative adaptive inverse filtering, the effects of vocal tract and lip radiation from the signal taken from high quality microphone are removed in this paper. In addition, through the figures and the corresponding in phase plane, the quality of the glottal flow without other facility is estimated and improved. The experiment at last also presents that the glottal flow used by this way can be directly served as acoustic excitation for speech synthesis using statistics due to the high naturalness.

【Key words】inverse filtering; phase plane; glottal flow; IAIF; speech synthesis

## 1 引言

语音的发音(speech)一般分为调音(articulation)和发声(phonation)两部分, 调音主要指由发音器官协作用形成不同的声道形状, 从而产生共鸣合成不同声音; 发声主要指声门开启时释放出来的空气流, 其中浊音的声源是一准周期的声门波, 清音是无规则的白噪声, 在言语中, 浊音占了大部分比重<sup>[1]</sup>。因此声门波作为浊音的激励源是后期高质量语音合成的基础, 需要有很高的自然度和精度, 本文的目的就是获取高自然度和精度的浊音激励信号。

利用外部传感器如舌位仪(EGG)判断语音的闭相位段可以直接获取声门波信号(signal of electroglottalgraph), 但这需要专门的外部设备, 给录音人带来诸多不便; 另一种是通过逆滤波的方法来获取, 虽然口腔面罩(如Rosenberg面具)或气流计可以用来提取声门波气流, 但是面具会影响人们的正常发音, 也需要其他辅助设备; 更广泛的方法是直接用高质量传声器录取声音信号, 对语音信号进行逆滤波来消除共振峰的影响来得到声门波信号, 且不需要借助其他设

备<sup>[2]</sup>。Wong等人在1979年提出用最小平方声门逆滤波的方法<sup>[3]</sup>, Alku在1992年提出了迭代自适应逆滤波的方法(Iterative Adaptive Inverse Filtering, IAIF)<sup>[4]</sup>, Plumpe等人在1997年提出了通过共振峰调节的方法来进行声门波倒数的估计<sup>[5]</sup>, 这些方法都可以进行声门波的估计。

通常由逆滤波得到的声门波在不借助其他设备的条件下很难对其质量进行评估, 并且常常由于系统本身的误差, 在声门波中还带有一些共振峰波纹, 若直接进行语音合成必然会进一步扩大误差, 为了平滑误差, 常用的方法是提取逆滤波后的声门波参数并进行建模, 用人工合成的声门波作为激励源<sup>[6]</sup>。常用的声门波建模方法有Rosenberg声门模型, Milenkovic六阶多项式模型和普遍采用的Fant微分声门模型, 由于自然声门波常伴随有基频抖动(jitter)和幅值抖动(shimmer)<sup>[7]</sup>, 简单地进行参数建模合成出的语音带有很强的机器味。因此人们常在激励模型的基础上对噪声建模来模拟自然语音的变化。但是由于声门发声的复杂性, 想要准确地逼近自然的声门波非常困难, 并且要花费大量的精力<sup>[8]</sup>。

笔者利用芬兰教授 Alku 提出的迭代自适应逆滤波 (IAIF) 的方法对自然语音进行逆滤波来得到声门波信号, 并利用其在相平面的图形和参数来提高声门波精度<sup>[9]</sup>。最后利用逆滤波产生的声门波和参数合成的声门波分别对同一声道模型进行激励, 对合成出的语音进行对比, 结果表明本文使用的声门波合成出的语音有更高的自然度。

## 2 发声和逆滤波原理

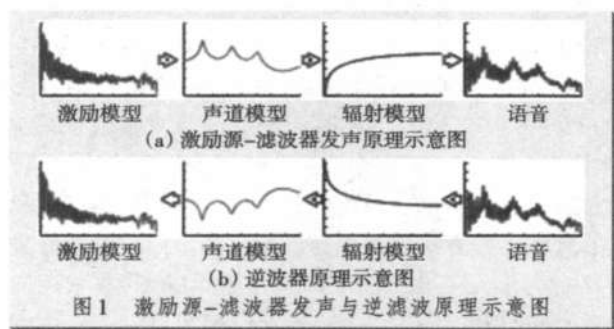
发声模型基于 Fant 的激励源-滤波器原理, 将声音的产生用三部分表示: 激励模型 (Source Excitation Model)、声道模型 (Vocal Tract Model) 和辐射模型 (Lip Radiation Model), 并假设声源和声道是线性无关的, 由声门波的信号来激励一个表示声道的二阶谐振器, 之后通过口唇辐射作用产生出语音。

所以语音信号可以表示为

$$S(z) = E(z) \times G(z) \times V(z) \times L(z) \quad (1)$$

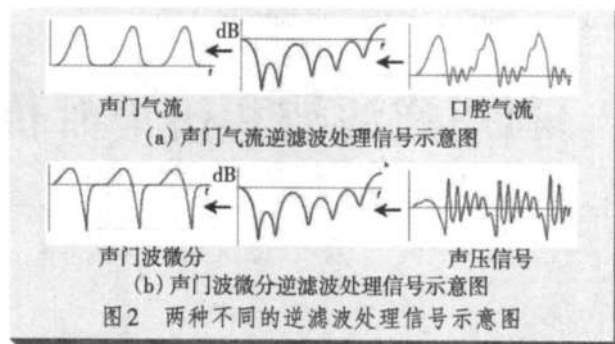
式中,  $S(z)$  表示语音信号;  $E(z)$  表示脉冲激励;  $G(z)$  表示声门形状模型;  $V(z) = \frac{1}{1 + \sum_{m=1}^p a(m)z^{-m}}$  表示声道模型;  $L(z) = 1 - \rho z^{-1}$  表示口唇辐射模型, 通过声道形状模型对一组脉冲激励信号滤波便可以模拟激励模型。

逆滤波可以看作是语音产生的逆过程, 语音信号去除口唇辐射作用后, 用一个传递函数为谐振器倒数的滤波器对其进行滤波便可去除语音中的共鸣信息, 最终得到声门波, 激励源-滤波器在频域的发声与逆滤波过程如图 1 所示。



在实际中, 逆滤波处理的语音信号可以是口腔气流也可以是声压信号, 两者的区别是口腔气流经逆滤波后可直接得到声门气流 (Glottal Airflow), 如图 2(a) 所示; 而声压信号经逆滤波后得到的是声门微分信号 (Differentiated Glottal Airflow), 对其进行积分去掉口唇辐射可得到声门气流<sup>[10]</sup>, 如图 2(b) 所示。

实验中, 采用传声器提取的是声压信号, 需对逆



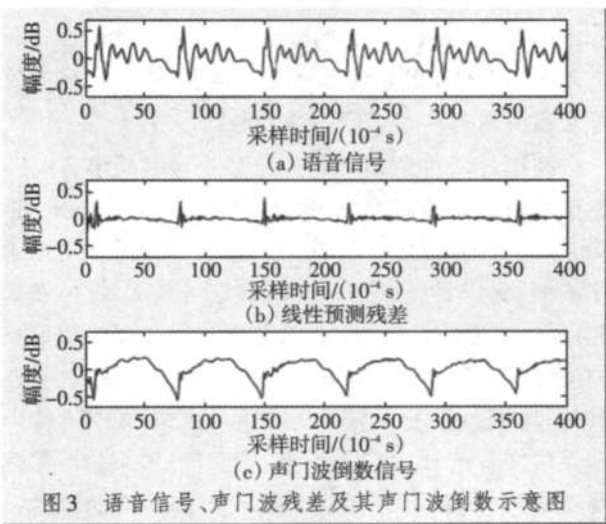
滤波后的信号积分才得到声门波, 但在实际中有时也直接使用声门波倒数来估计声门波。如图 1 中口鼻辐射起到微分算子的作用, 声门波脉冲的一阶导数在闭相位有明显的负脉冲, 在进行说话人识别和声源匹配上更方便, 如 LF 模型和 KLATT88 共振峰参数合成器都是利用声门气流的微分形式来建立模型的, 所以有时也将辐射效应和声源放在一起, 设  $E' = E(z) \times G(z) \times L(z)$ , 则语音信号为

$$S(z) = E'(z) \times V(z) \quad (2)$$

声门波一阶倒数为

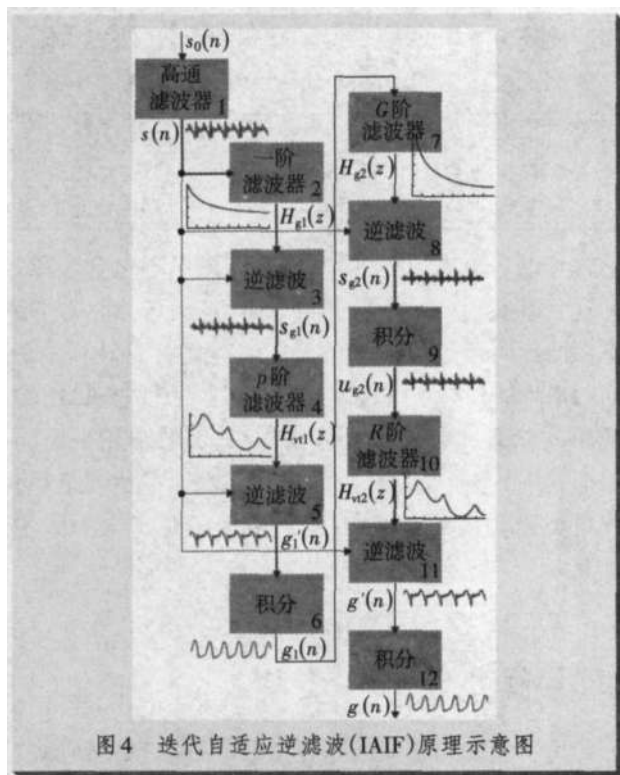
$$E'(z) = \frac{S(z)}{V(z)} \quad (3)$$

因此  $1/V(z)$  即为声道的逆滤波器,  $V(z)$  的准确估计便成为声门波倒数估计的关键, 常使用最小误差估计的线性预测 (Linear Prediction) 方法来进行声道模型  $V(z)$  的计算, 但是这里不能直接使用声道参数  $1/V(z)$  作为逆滤波器去估计声门波倒数。如图 3 所示, 可以看到在声门闭合时刻 (GCI) 由于声门气流的快速变化, 线性预测出现了最大的负脉冲, 逆滤波需要精确定位声门闭合时刻来避开声门脉冲激励项的影响, 否则得到的声门脉冲将很不准确<sup>[9]</sup>。为了避免确定声门关闭时刻, 本文采用 IAIF 方法。



### 3 迭代自适应逆滤波(IAIF)原理

IAIF算法思想和其他逆滤波思想一样,都是对声道模型建模并通过逆滤波消除共振峰的影响。IAIF首先从原始语音的频谱中消除声门激励对它的影响,通过线性预测的方法(Linear Prediction Coding, LPC)或离散全极点的方法(Discrete All-pole Model, DAP)精确估计出声道的模型,最后通过逆滤波得到声门信号<sup>[2]</sup>。其中声门和声道传递函数的估计在IAIF算法中重复了两次<sup>[4]</sup>。原理如图4所示,具体步骤如下。



语音信号 $s_0(n)$ 先通过高通滤波器1来消除传声器带来的低频信号,设截止频率为80 Hz。声门信号频谱以-12 dB/oct斜率衰减,所以可以用一个一阶滤波器 $H_{g1}(z)$ 估计出声门波对语音的影响,然后再通过逆滤波 $S_{g1}(z) = S(z)/H_{g1}(z)$ 来消除声门对语音的影响。之后用一个 $p$ 阶全极点滤波器 $H_{v1}(z)$ 对声道进行第一次估计,声道对语音的影响便可以通过逆滤波的方式来消除,这时就得到了声门波的倒数 $g_1'(n)$ ,对其积分就可以消除唇辐射的影响,得到声门波信号 $g_1(n)$ ,这是对声门波的第一次估计。

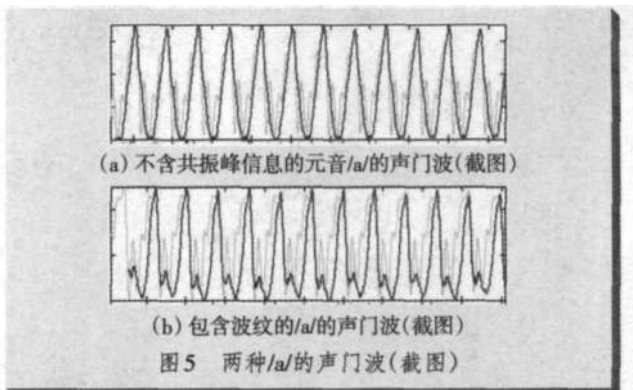
类似地,方框(7)~(12)完成了对声门波的第二次估计。 $g_1(n)$ 通过一个 $G(2\sim4)$ 阶滤波器 $H_{g2}(z)$ 重新对声门波的作用进行评估,然后通过逆滤波的方法来消除声门波的影响得到 $s_{g2}(z)$ ,唇辐射的作用通过

积分来消除得到 $u_{g2}(z)$ ,之后一个信号 $R$ 阶滤波器 $H_{v2}(z)$ 来对声道信号进行评估,再通过逆滤波的方式得到声门波倒数信号 $g'(n)$ ,最后通过积分来消除唇辐射的影响,得到精确的声门波信号 $g(n)$ 。

### 4 利用相平面对声门波质量进行评估以提高逆滤波精度

#### 4.1 评估的必要性

由IAIF算法可以看出,大部分参数的设定(如滤波器的阶数和参数)都是人为主观决定的,所以逆滤波不可避免地会带来误差,但是在不借助EGG等其他设备的条件下又很难估计出由逆滤波得到的声门波质量的好坏,所以客观地评价声门波的质量就变得非常必要了。而主要影响逆滤波得到的声门波质量的因素是共振波波纹<sup>[11]</sup>,如图5所示,这是由于没有完全将声道的共振峰滤除的结果,而这种误差可以借助声门波及其倒数在相平面上的图形来进行简单估计。



#### 4.2 相平面评估原理

通过逆滤波得到的一个理想声门波应不含有任何共振信息,而相平面便可以用来评估逆滤波后的声门波中是否含有共振峰信息,并以此判断声门波的质量。假设声道用串联的二阶谐振器表示<sup>[12]</sup>,声门信号用一个二阶谐振方程来表示时,声门波信号为 $x(t)$ ,声门波倒数为 $\frac{dx}{dt}$ 。则有

$$\frac{d^2x}{dt^2} + x = 0 \quad (4)$$

为了将式(4)在相平面上表示,重新整理为

$$\begin{cases} \dot{x} = x_1 \\ \dot{x}_1 = x_2 \end{cases} \quad (5)$$

因此根据式(4)~(5)可以得到



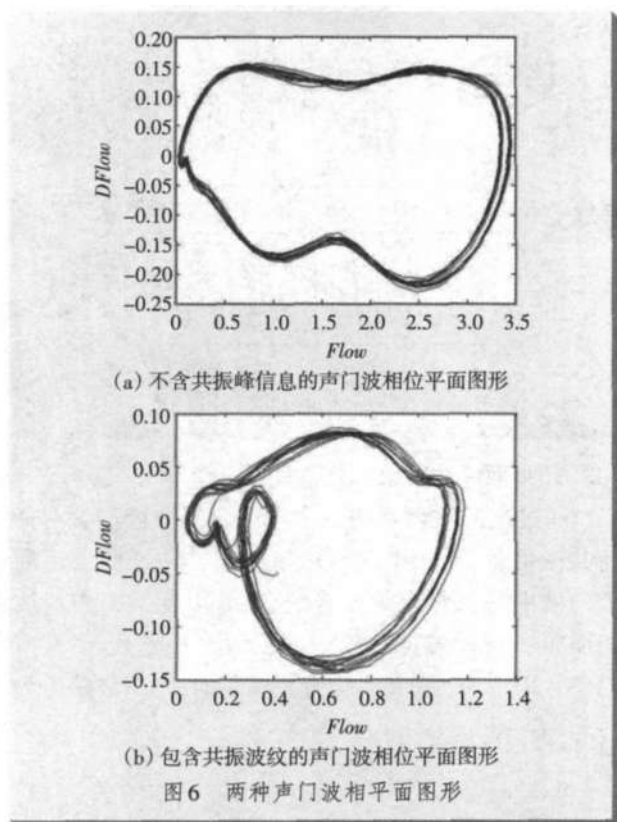
$$\begin{cases} \frac{dx_1}{dt} = x_2 \\ \frac{dx_2}{dt} = x_1 \end{cases} \quad (6)$$

所以有

$$x_1^2 + x_2^2 = C \quad (7)$$

这里  $C$  是常数。

由图5可以看出,元音的声门波信号是准周期性的,即  $f(t, x) \approx f(t+T, x)$ , 因此谐波方程的解  $x_1, x_2$  在相平面上的图形便是一组闭合的同心圈(不一定是圆形),声道对不同频率的声音的谐振作用会产生不同周期性的解,所以当逆滤波把共振作用完全去掉后,相平面的图形应是仅由基频作用产生的单一闭合圈,如图6(a)所示;但如果共振作用还存在的话,就会存在由谐振作用其他频率产生的小圈,如图6(b)所示,而共振峰波纹和小圈又有紧密的联系,小圈的面积越小说明逆滤波越精确,所以根据相平面圈的形状就可以判断逆滤波是否成功。



#### 4.3 利用相平面提高逆滤波精度

仅仅靠相平面是否含有其他圈来评价声门波还是需要人的主观判断,文献[13]提出了客观评价声门波质量的方法,通过计算每一基频所对应的相平面中圈的数量便可估计出声门波的质量并以此选择合适

的逆滤波参数。理想的声门波中,每个基频仅对应着一个闭合圈,因此所对应的比例  $R$  应为1,如果这个比例接近2或更高,则说明还存在其他闭合圈,在实际应用中一般以1.3为一个阈值,当超出这个阈值时,可以估计小闭合圈的面积,从而进一步评估声门波质量。假设小闭合圈的形状为一圆形,则其面积  $a$  和周长  $L$  的平方成正比,因此所有小闭合圈的面积和可近似为  $A=Na=NL^2$ , 其中  $N$  为小闭合圈的数目。比例  $R$  越小,闭合圈面积  $A$  越小,则表明逆滤波质量越好。

表1为对应于图5中不包含共振波纹的声门波(a)和包含共振波纹的声门波(b)在相平面的对应参数。

表1 声门波(a)和(b)相平面对应参数

参数	声门波(a)	声门波(b)
每基频对应圈数 $R$	1.012 60	1.943 60
小圈的平均长度 $L$	—	0.134 82

可以看出声门波(a)的相平面图形不包含小圈,并且每基频对应的圈数  $R \approx 1$ ,而声门波(b)对应的相平面图形明显包含了小圈,  $R=1.943 60$ ,并且计算了小圈的平均周长  $L=0.134 82$ 。所以通过不断调整IAIF的逆滤波参数,同时参考声门波及其倒数在相平面的图形,使每基频所对应的圈数和小圈面积达到最小,就可以得到高精度的声门波,同时避免了人为的主观判断。

## 5 参数合成声源和逆滤波声源语音合成实验对比

为了验证用相平面进行参数改进的IAIF逆滤波的声门波质量,课题后期用参数拟合的声门波和用逆滤波得到的声门波分别对同一声道模型进行激励,通过分析合成出的语音来选择高自然度和精度的声门波。

由于逆滤波对原始语音信号非常敏感,需要语音的提取设备在整个带宽中都保持线性相位,本文采用高质量的电容传声器(Condenser Microphone)作为录制设备,避免了相位的非线性,同时选择低频截止频率小的电容传声器来避免设备本身的低频干扰。环境要求无噪声,录制过程中保持口唇与传声器的间距不变。对喉头无病变的女声陈述句发/a/音进行采样,采样精度为16 kHz,量化精度为16 bit。利用语音处理软件Praat对语音共振峰的频率和带宽进行提取,利用串联型滤波器  $H(z) = \frac{A}{1 - Bz^{-1} - Cz^{-2}}$  对声道进行建

模。利用上面的IAIF方法对语音信号进行逆滤波,利用声门波在相平面的参数和图形再次选择IAIF的滤波参数,使得每个基频对应的圈数和小圈面积达到最小,得到最终逆滤波的声门波信号A;对于参数合成的声门波的建构,采用Rosenberg模型对声门波进行参数拟合得到人工合成的声门波信号B。用A, B对声道模型进行激励得到合成的语音信号,图7分别为对应的声门波图形和声门波频谱图。

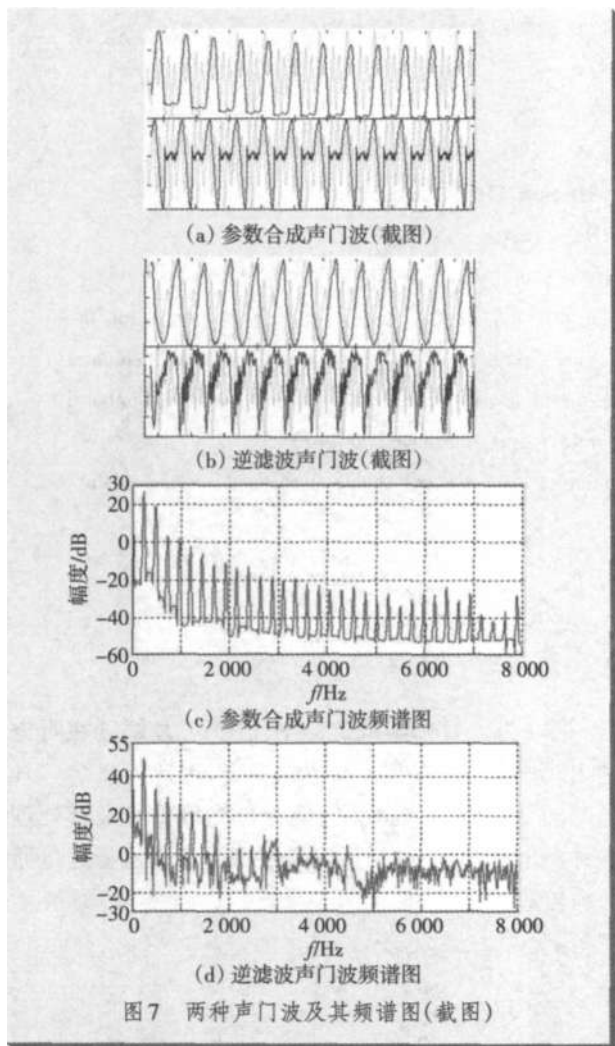


图7 两种声门波及其频谱图(截图)

可以看到人工合成的声门波在高频部分有很强的谐波信号,因此在试听的时候发现有很强的刺耳声(buzzy)和机器味。从图中可以看到逆滤波得到的声门波保持了自然语音的基频抖动和幅度抖动,并且没有共振波纹的存在,经试听后,自然度也有很大改善。

## 6 小结与展望

本文利用相平面来选择合适的IAIF参数对自然语音进行逆滤波可以得到高自然度的声门信号,并且由于去除了共振峰波纹,可以达到很高的精度,因此

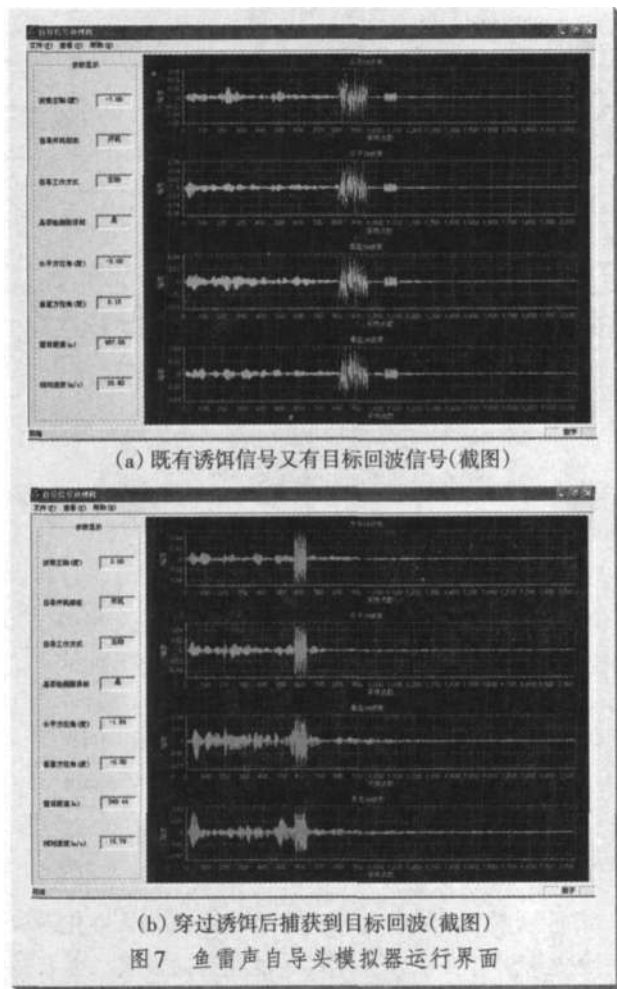
可直接在语音合成中作为激励源,避免了声门波建模和噪声的逼近,为后期高质量语音合成提供了基础。但是这里需要强调的是共振峰实验本身是为了验证逆滤波后声门波的质量,而且随着可训练语音合成的发展,可以结合逆滤波原理和隐马尔科夫原理,利用IAIF得到的参数,用本文得到的自然声门波作为激励源进行HMM合成<sup>[14]</sup>。所以课题今后的工作是对逆滤波得到的声门波进行参数提取,并且结合统计声学建模来灵活地进行参数调整,以满足多情感语音合成的需要。

### 参考文献

- [1] 孔江平. 语言发声研究的基本方法[EB/OL]. (2010-08-01)[2010-11-01]. <http://www.phonetics.ac.cn/research/report%20paper/2008paper/1kong.pdf>.
- [2] AIRAS M. TTK aparat: an environment for voice inverse filtering and parameterization[J]. Logopedics Phoniatrics Vocology, 2008, 33(1): 49-64.
- [3] WONG D, MARKEL J D, GRAY A H. Least squares glottal inverse filtering from the acoustic speech waveform[J]. IEEE Trans. Acoustics, Speech, and Signal Processing, 1979, 27(4): 350-355.
- [4] ALKU P. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering[J]. Speech Communication, 1992, 11(2/3): 109-118.
- [5] PLUMPE M D, QUATIERI T F, REYNOLDS D A. Modeling of the glottal flow derivative waveform with application to speaker identification[J]. IEEE Trans. Speech and Audio Processing, 1999, 7(5): 569-586.
- [6] 孙俊. 基于激励源及其韵律特征的源——目标说话人声音转换研究[D]. 合肥: 中国科技大学, 2006.
- [7] LALWANI A L, CHIDERS D G. Modeling vocal disorders via formant synthesis[C]//Proc. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing. [S.l.]: IEEE Press, 1991(1): 505-508.
- [8] RAITIO T. Hidden Markov model based finnish text-to-speech system utilizing glottal inverse filtering[D]. Espoo: Helsinki University of Technology, 2008.
- [9] AIRAS M, PULAKKA H, BACKSTROM T, et al. TTK aparat ver. 0.3.1. [S/OL]. (2010-08-28)[2010-11-01]. <http://aparat.sourceforge.net>.
- [10] GOBL C. The voice source in speech communication—production and perception experiments involving inverse filtering and synthesis [D]. Stockholm: Department of Speech, Music and Hearing, KTH, 2003.

(下转第73页)





(1)在阵元声信号的层次上实现对鱼雷声自导系统进行建模与仿真,凸显了目标、环境、鱼雷自导的声学特性,真实地反映了鱼雷自导系统的工作性能。(2)采用工控机+并行处理系统的硬件架构,由8片TS201处理器构成的多DSP并行处理系统实现高性能的实时运算,满足系统实时性要求。(3)采用模块化+多线程的软件设计,具有很好的可扩充性和可重用性,添加相应的功能模块或仿真节点即可将当前系统升级为

水声对抗仿真系统。(4)采用阵元信号产生与自导信号处理分离的分布式系统设计,各自承担相应的仿真功能,具有良好的可移植性,将目标环境模拟器或鱼雷声自导头模拟器配备实物设备(声对接阵或声学基阵)即可单独或共同参与到其他半实物仿真系统中,方便快捷地实现数字仿真到半实物仿真的转换。

#### 参考文献

- [1] 陈军,黄建国,潘小京. 自航式声诱饵对抗声自导鱼雷有关问题研究[J]. 电声技术,2008,32(8):41-44.
- [2] 杨云川,崔怀林,李志舜. 基于水声物理场的鱼雷制导半实物仿真方法及应用[J]. 系统仿真学报,2005,17(3):755-759.
- [3] 王刚,黄飞,乔纯捷,等. 基于以太网的鱼雷声自导电信号半实物仿真系统[J]. 兵工学报,2009,30(5):567-571.
- [4] 刘永丰,关国枢,张西建,等. 鱼雷声寻的制导半实物仿真系统研究与实现[J]. 舰船科学技术,2003,25(3):26-29.
- [5] 王新晓. 自导系统仿真建模及仿真系统研究[D]. 西安:西北工业大学,2002.
- [6] 沈玉利,沈云涛,赵俊渭,等. 水声对抗分布交互仿真系统的研究[J]. 系统仿真学报,2002,14(6):726-731.
- [7] 聂卫东,康凤举,苏颖,等. 鱼雷声自导检测目标过程的建模与实时仿真[J]. 计算机仿真,2005,22(6):57-60.
- [8] 周明,初磊. 高频噪声干扰器干扰主动声自导鱼雷仿真研究[J]. 兵工学报,2010,31(3):327-330.
- [9] 由延军,康凤举,唐凯. 综合海洋战场环境分布交互仿真系统研究[J]. 系统仿真学报,2009,21(19):6048-6052.

#### 作者简介

杨占龙, 博士后, 主要研究方向为水下信号与信息处理、系统仿真。

[责任编辑]丁雪

[收稿日期] 2011-01-13

(上接第63页)

- [11] 金学成. 基于语音信号的情感识别研究[D]. 合肥:中国科技大学,2007.
- [12] EDWARDS J A, ANGUS J A S. Using phase-plane plots to assess glottal inverse filtering[J]. Electronics Letters,1996,32(3):192-193.
- [13] BACKSTROM M T, AIRAS M, LEHTO L, et al. Objective quality measures for glottal inverse filtering of speech pressure signals [C]//Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 05). Philadelphia, PA, USA: IEEE

Press,2005:897-900.

- [14] RAITIO T, SUNI A, YAMAGISHI J, et al. HMM-based speech synthesis utilizing glottal inverse filtering [J]. IEEE Trans. Audio, Speech, and Language Processing,2011,19(1):153-165.

#### 作者简介

胡琼, 硕士研究生, 研究方向为语音处理与合成;

赵春宇, 副教授, 研究生导师, 研究方向为语音处理、语音合成和嵌入式技术等。

[责任编辑]丁雪

[收稿日期] 2010-11-30