第45卷第7期 中国科学技术大学学报 Vol. 45, No. 7
2015年7月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Jul. 2015

Article ID:0253-2778(2015)07-0575-07

基于声道长度对齐的年龄语音转换

李金中[1,3],李贤[1,3],汪增福[1,2,3]

(1.中国科学技术大学自动化系,安徽合肥 230027;
2.中国科学院合肥智能机械研究所,安徽合肥 230031;
3.语音及语言信息处理国家工程实验室,安徽合肥 230027)

摘要:提出一种基于声道长度对齐的年龄语音转换方法.该方法包含频谱转换和基频转换两个方面,前者在频域依据声道因子和弯折函数对已进行基音标注过的每一帧语音的频谱进行弯折转换;后者对基频特征的转换采用线性变换方法.实验结果表明,通过对同一人不同年龄段的语音进行转换合成,由年龄较大语音向年龄较小语音转换时,转换合成得到的语音频谱平均距离得到明显减小,转换效果较好,而从年龄较小语音向年龄较大语音转换时,频谱平均距离减少较小,同时女性年龄语音转换的效果和自然度都好于男性.

关键词:年龄语音转换;声道长度对齐;基音标注;声道因子;弯折函数 ;线性变换

中图分类号:TP391 **文献标识码**:A doi:10.3969/j.issn.0253-2778.2015.07.007

引用格式:LI Jinzhong, LI Xian, WANG Zengfu. Vocal tract length aligning based mandarin age voice conversion [J]. Journal of University of Science and Technology of China, 2015,45(7): 575-581.
李金中,李贤,汪增福. 基于声道长度对齐的年龄语音转换[J]. 中国科学技术大学学报,2015,45(7): 575-581.

Vocal tract length aligning based mandarin age voice conversion

LI Jinzhong[1,3], LI Xian[1,3], WANG Zengfu[1,2,3]

(1. *Dept. of Automation, University of Science & Technology of China, Hefei* 230027, *China*;
2. *Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei* 230031, *China*;
3. *National Engineering Laboratory of Speech and Language Information Processing, Hefei* 230027, *China*)

Abstract: Vocal tract length aligning was proposed for mandarin age voice conversion which transforms age speech into some required target age speech. In the method, the speech spectrum which has been pitch marked was warped in the frequency domain based on the warping factor and warping function while pitch was converted by linear transformation. The experimental results show that the effect of transforming old age speech into a young one is better than otherwise and that the average spectra distance of the former is markedly reduced. Meanwhile, age voice conversion is better for female voice than for male voice in effectiveness and naturalness.

Key words: age voice conversion; vocal tract length aligning; pitch marker; warping factor; warping function; linear transformation.

收稿日期:2014-10-11;**修回日期**:2015-04-03
基金项目:国家自然科学基金(61472393),安徽省自主创新专项基金(13Z02008)资助.
作者简介:李金中,男,1990年生,硕士生. 研究方向:语音信号处理. E-mail: jinzhli@mail.ustc.edu.cn
通讯作者:汪增福,男,博士/教授. E-mail: zfwang@ustc.edu.cn

0 引言

语音合成正在越来越多地被用于人机通信和其他应用，比如为有语言沟通障碍的人提供口语和阅读帮助. 随着语音合成技术的发展，合成出高质量的语音已经不是一个难题，但它们的自然度仍然有待提高，其中一个原因就是对性别、年龄和情感等语音信息的利用还很少. 近些年，已经有了一些关于性别和情感语音特征的研究工作[1-3]，然而很少有工作尝试合成语音的年龄信息. 目前基于大语料库的语音合成方法并不能有效应用到年龄语音合成，相比较而言，语音转换使用一个小型语料库可实现不同年龄段语音的转换.

影响语音年龄特征的主要参数包括频谱特征和基音频率 F_0[4]. 目前大多数的语音转换方法都是基于平行语料库的研究[5-7]，此外，还要求语音具有高度的自然时间对准和相似的基音轮廓[8]，由于基于平行语料的语音转换应用范围受到限制，这几年又有一些基于非平行语料的转换方法[9-10]，然而年龄语音转换所需语料时间跨度大，现有语料库很难满足.

声道长度归一化(VTLN)[11]试图通过在频率轴上对频谱的幅度进行拉伸或压缩来减小因说话者声道长度不同所造成的影响. 在语音识别领域中，VTLN 旨在将一个说话者的语音标准化以减少说话者个性特征对识别率的影响. 年龄语音转换是个很相似的任务，它是将说话人的声音转换成该说话人其他年龄段时的发音，采用声道长度对齐的算法实现年龄语音转换. 我们采用 TANDEM-STRAIGHT[12-13]作为语音信号合成器，它可以将基频从频谱中分离出来，训练阶段包括频谱转换的弯折因子 和基频转换模型，转换阶段包括频谱转换和基频转换，如图 1 所示.

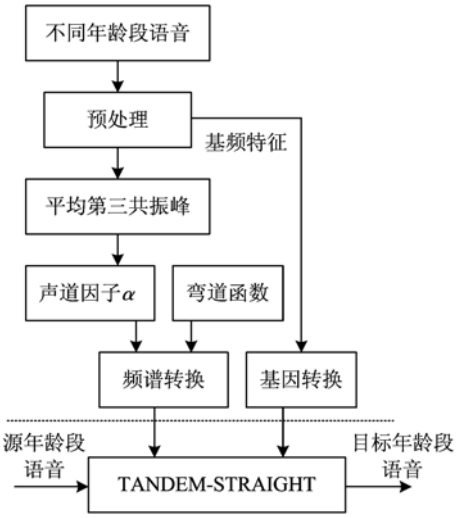

图 1 年龄语音转换流程图

Fig. 1 The flowchart of age voice conversion

1 声道特点

1.1 声道特点

声道包括喉腔、咽头、口腔和鼻腔，它是一谐振腔，有许多自然谐振频率，所以它能够放大某一频率分量，衰减其他频率分量. 声道的形状决定每一瞬间谐振频率的大小，说话时，嘴唇和舌头会不断运动，使声道的尺寸和形状不断发生变化，进而改变谐振频率. 这些谐振频率称为共振峰频率，简称共振峰.

随着年龄和性别等因素的变化，声道长度也会发生变化，对于成年男性，其声道长度约为 18 cm，而成年女性的声道长度只有 13 cm，同时儿童的声道长度比成年人的短，比如 8 岁小孩的声道长度在 10 cm 左右. 如果假设声道是横截面积为 1、长度为 L 的均匀管道，可理论分析得出，声道长度 L 与共振峰频率成反比.

1.2 年龄语音数据库

语音数据库在语音转换研究中非常重要，年龄语音转换之所以难，很大程度上是因为关于同一个人的不同年龄段语音数据难以获取，文献[4]就选用了一个家庭中四代直系亲人的语音进行替代. 为了便于进行年龄语音转换的研究，我们在互联网上搜集数据，建立了一个小型的年龄语音数据集. 数据集中包含了两个人的语音，一个女性，在其 12 岁和 20 岁两个阶段；一个男性，在其 12 岁、18 岁和 23 岁三个阶段，表 1 为用于训练的年龄语音数据集的构成，其中每句话的长度在 5 s 左右.

表 1 年龄语音数据库的构成

Tab. 1 The composition of age voice database

年龄段	女性(语句数)	男性(语句数)
12	50	61
18～20	35	70
23	/	45

2 基于声道长度对齐的频谱转换

2.1 基本原理

不同语音之间的声学差异主要缘于说话人器官

构造的不同[14]，如声道的长度、声带的长度和宽度、肺活量的大小以及嘴唇的位置和形状等等.通常，声道长度对齐的目的就是通过对特征参数空间进行变换，使测试语料和目标模型的声道长度差异最小化.声道长度对齐是通过平移和折叠语音信号的频谱来“改变”声道长度，使不同年龄语音之间的共振峰频率相匹配，主要有以下两种方法：

(Ⅰ)通过对不同年龄语音共振峰频率的估计，直接估计得到频率转换因子；

(Ⅱ)利用最大似然准则来估计每个说话人对应的频率转换因子.

估计得到频率转换因子后，利用谱平移(spectral shift)算法，如频率弯折(frequency warping)，对不同年龄语音的频谱进行归整，从而校正由声道长度不同造成的影响.

声道长度归一化在语音识别领域已经有了相当成熟的研究[11,15]，其核心思想是利用少量的训练语料，估计出说话人的频率转换因子，然后利用谱平移算法对该说话人的语音频谱进行归正处理，归正后的语音频谱很像是一个具有标准声道长度的说话人的语音频谱.

通常可以用级联无损短管模型来描述声道，声道传递函数包括鼻腔谐振、唇辐射和声道响应的影响.忽略唇辐射的影响，不同年龄语音的频谱差异将主要决定于声道面积函数，即频谱差异主要源于不同的声道长度的影响.对于同一个人，其声道长度是随着年龄的变化而变化的，特别是在青春期这种变化更加显著，因而可以通过频率归正算法将一段语音转换成不同年龄段的语音.

2.2　基音标注

随着时间的变化，语音是一系列周期性或非周期性的信号.大多数的语音处理过程将语音信号分割成 10～30 ms 的持续时间帧，每一帧被单独处理，以此来研究语音信号的特征随时间的变化.在语音合成中，单个帧的持续时间是非常重要的，因为它对后续合成语音的质量影响很大.研究发现，当语音被分割成的帧与由语音的基音频率所决定的浊音伪周期相匹配时，合成的语音质量达到最好.将语音分割成基音同步帧的过程叫基音标注.

年龄语音转换的过程基于基音同步帧模式，这就需要对语音数据进行基音标注.对于浊音段有基音周期，而清音段信号则属于白噪声，所以这两种类型需要区别对待.依据 Prasanna 等提出的基音标注评价指标[16]，使用 Praat 程序可以达到很稳定的基音标记结果.Praat 的基音标注算法基于自相关法进行声学周期检测，与其他自相关法相比，Praat 运用 $\sin x/x$ 内插的滞后域，并且 Praat 利用高斯窗替代汉明窗来减少基音判定误差，如图 2 所示.

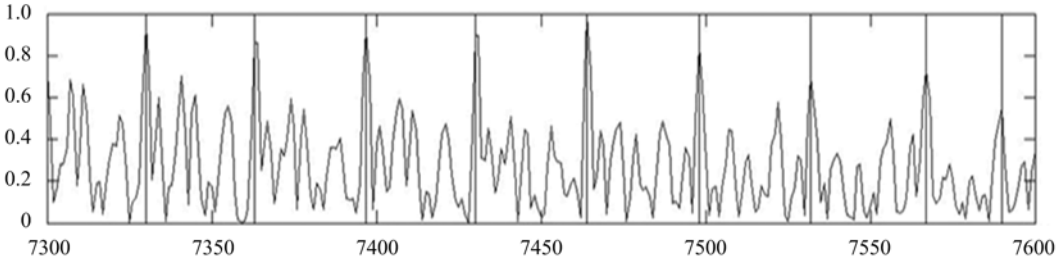

图 2　Praat 基音标注结果

Fig. 2　Result of the pitch marking algorithm by Praat

2.3　声道因子估计

在级联无损短管的声道模型假设下，声道长度 L 与声道谐振频率 F_i 成反比关系[14,17]：

$$F_i \approx \frac{(2i-1)c}{4L}, i = 1,2,3,\cdots \qquad (1)$$

式中，F_i 为第 i 共振峰，c 为声速.实际上，若假设声道形状变化一致，文献[18]指出不同声道长度对应的声道传递函数之间具有如下关系：

$$V_{L=L_2} = V_{L=L_1}\left(\frac{L_2}{L_1}\cdot\omega\right) = V_{L=L_1}(\alpha\cdot\omega) \qquad (2)$$

式中，$V(\cdot)$为声道传递函数，α 为相对声道长度，通常也称其为声道因子，L_1、L_2 为不同的声道长度.由式(1)中声道长度与共振峰频率之间的关系可以得出，不同年龄语音的声道因子可以从其语音频谱的对应共振峰比值中得到，即

$$\alpha = L_2/L_1 = F_i^{L_1}/F_i^{L_2} \qquad (3)$$

利用频谱分布来完成声道因子估计，只需要少量的年龄语音数据.

最大似然法是通过最优化准则来指导声道因子

的估计,其估计方法可以用如下公式来描述:

$$\hat{\alpha} = \mathrm{argmax}_{\alpha} P(X_i^{\alpha} \mid \lambda, W_i) \tag{4}$$

式中,λ 是初始的声学模型,W_i 是对应的基元描述序列,X_i^{α} 是说话人 i 对应的、根据声道因子 α 完成频谱调整之后的特征参数序列,因而最终声道因子估算的结果将对应于最大似然得分. 声道长度对齐过程中,初始模型的选择是这种方法需要解决的问题之一,此外由于它将对特征参数的最优化与参数估计过程联系在一起,从理论上保证了声道因子估计结果在最大似然度量意义下的准确性,其效果要好于频谱估计的方法,但是式(4)的解析解难以得到,只能采用数值方法进行搜索求解,计算复杂度较大.

2.4 弯折函数

为了实现语音的年龄转换,使不同年龄段发音的声道长度对齐,语音每一帧的频谱都要作相应的弯折. 声道长度归一化算法可用于语音帧的频谱弯折,即在范围为 $0 \leqslant \omega \leqslant \pi$ 的频率轴上对频谱进行拉伸或压缩. 频谱弯折是根据弯折函数 $g(\omega)$,弯折函数需要是一个返回值在 0 到 π 之间的单调函数,同时它的形状受上节所求声道因子 α 的影响,弯折函数曲线如图 3 所示. 弯折函数可以有以下四种形式:

对称函数[19]:

$$g(\omega,\alpha) = \begin{cases} \alpha\omega, & \omega \leqslant \omega_0 \\ \alpha\omega_0 + \dfrac{\pi - \alpha\omega_0}{\pi - \omega_0}(\omega - \omega_0), & \omega > \omega_0 \end{cases};$$

$$\omega_0 = \begin{cases} 7\pi/8, & \alpha \leqslant 1 \\ 7\pi/8\alpha, & \alpha > 1 \end{cases}.$$

二次函数[20]:

$$g(\omega,\alpha) = \omega + \alpha\left[\left(\frac{\omega}{\pi}\right) - \left(\frac{\omega}{\pi}\right)^2\right].$$

能量函数[21]:

$$g(\omega,\alpha) = \pi\left(\frac{\omega}{\pi}\right)^2.$$

非对称函数:

$$g(\omega,\alpha) = \begin{cases} \alpha\omega, & \omega \leqslant \omega_0 \\ \alpha\omega_0 + \dfrac{\pi - \alpha\omega_0}{\pi - \omega_0}(\omega - \omega_0), & \omega > \omega_0 \end{cases};$$

$$\omega_0 = 7\pi/8.$$

频谱的值依据频率的弯折位置进行内插,弯折后的频谱的一些区域被压缩,其他区域被拉伸. 图 4 展示了一段语音帧的频谱以及它经能量函数弯折后的频谱.

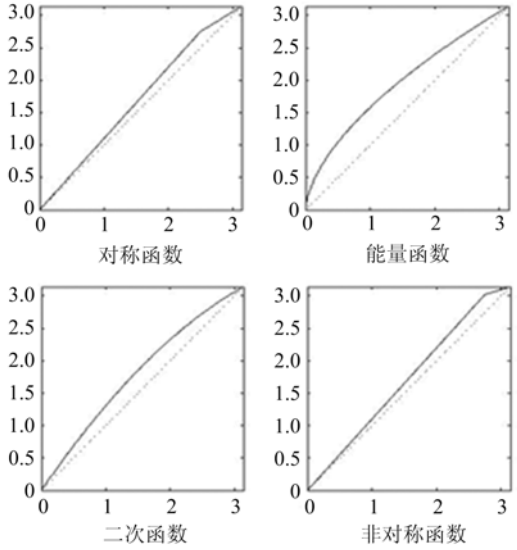

图 3 弯折函数的函数曲线

Fig. 3 Curve of warping functions

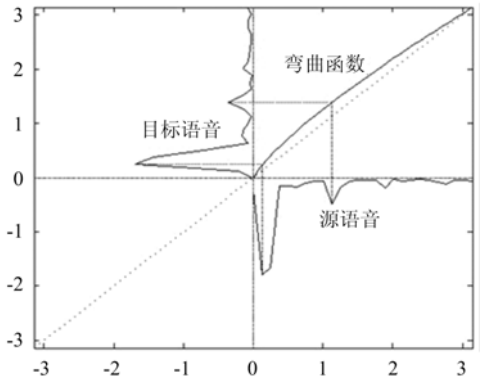

图 4 采用能量弯折函数的频谱变换

Fig. 4 Spectrum transform by a power warping function

3 基频转换

对基音频率 f_0 的转换,采用基于均值-方差的线性变换方法[22],并假定基音频率的均值带有说话者的大量特定信息. 该方法是指转换源说话者的 f_0 值,使得转换后基音均值的轮廓和目标说话者的基音变化范围相匹配,同时保持源说话者的音调模式. 假定说话者的每个 f_0 值都服从有特定均值和方差的高斯分布,线性变换可以被定义如下:

$$s = h(s) = as + b \tag{5}$$

式中,t 是目标基音的瞬时值,s 是源基音的瞬时值,目标就是根据含有均值和方差值的两个高斯分布计算出 a 和 b.

令目标语音的基音值的概率密度函数为 p_t,含

参数(μ_t,σ_t),源语音基音值的概率密度函数为 p_s,含参数(μ_s,σ_s),线性变换的概率密度函数可以写成如下形式:

$$p_t(t) = h(s) = \frac{\partial h^{-1}(t)}{\partial t} p_s(h^{-1}(t)) \tag{6}$$

考虑式(5)线性变换的逆变换,得到

$$h^{-1}(t) = \frac{t-b}{a} \tag{7}$$

$$\frac{\partial h^{-1}(t)}{\partial t} = \frac{1}{a} \tag{8}$$

这样,式(6)就可以被写为:

$$p_t(t)\ \frac{1}{a} p_s(\frac{t-b}{a}) \tag{9}$$

用高斯分布的表达式来代替两边,得到

$$\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp(-\frac{(t-\mu_t^2)}{2\sigma_t^2}) =$$

$$\frac{1}{a}\ \frac{1}{\sqrt{2\pi\sigma_t^2}}\exp(-\frac{(\frac{t-b}{a}-\mu_s^2)}{2\sigma_s^2}) \tag{10}$$

对两边取对数得到

$$\log(\frac{a\sigma_s}{\sigma_t}) - \frac{(t-\mu_t)^2}{2\sigma_t^2} =$$

$$-\frac{(\frac{t-b}{a}-\mu_s^2)}{2\sigma_s^2} \tag{11}$$

二次项式等同于源语音和目标语音基音分布的方差表达为:

$$-\frac{1}{2\sigma_t^2} = -\frac{1}{2a\sigma_s^2} \tag{12}$$

$$a = \frac{\sigma_t}{\sigma_s} \tag{13}$$

将 a 的表达式带入式(11)得到

$$b = \mu_t - \frac{\sigma_t \mu_s}{\sigma_s} \tag{14}$$

一定数目的训练语音可以用来估计说话者的均值和方差的值,在年龄语音转换阶段,公式(5)可以应用于输入语音每一帧的 f_0 值以产生一个目标轮廓.

4　实验结果与分析

4.1　实验条件

实验基于建立的年龄语音数据集,对不同年龄阶段的语音进行相互转换;语音文件的采样率为 16 KHz.首先进行基音标注,对语料进行分帧,计算出语料的基频和共振峰.声道因子的估计可以作为测试语料的平均第三共振峰和训练语料的平均第三共振峰的比值;再加入到弯折函数中得到频谱转换模型,由基频信息的均值和方差得到基频转换的线性模型;最后使用 TANDEM-STRAIGHT 方法提取出输入语料的基频和频谱用作年龄语音转换.

对于每两个年龄段之间的语音转换,选择源语音中的任意 5 句作为测试语料,其他源语音和目标语音作为测试语料.

4.2　年龄语音转换的客观评价

为了对频谱转换进行客观评价,我们定义了频谱平均距离(spectra average distance,SAD)作为客观评价准则:

$$\mathrm{SAD} = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}(\overline{X}_n - \overline{Y}_n)^2}$$

其中,$\overline{X}$ 和 $\overline{Y}$ 分别为目标语音频谱均值和转换合成语音频谱均值,N 为频谱数据的维数.对于数据库中的语音,转换方式包括:

(Ⅰ)W_M2S:将女性 20 岁时的语音转换成该女性 12 岁时的语音

(Ⅱ)W_S2M:将女性 12 岁时的语音转换成该女性 20 岁时的语音

(Ⅲ)M_L2M:将男性 23 岁时的语音转换成该男性 18 岁时的语音

(Ⅳ)M_M2S:将男性 18 岁时的语音转换成该男性 12 岁时的语音

(Ⅴ)M_S2M:将男性 12 岁时的语音转换成该男性 18 岁时的语音

图 5 和图 6 分别是女声和男声年龄转换后的频谱平均距离 SAD.

从各组实验结果来看,从年龄较大语音向年龄较小语音转换时,得到的频谱平均距离得到明显减小,转换效果较好;而从年龄较小语音向年龄较大语音转换时,频谱平均距离减少较小.这可能是由于弯折函数作用于增长声道长度时效果不明显,同时年龄语音数据库的语音质量不是很好,影响了实验结果.

对于弯折函数,在由较大年龄语音向较小年龄语音转换时,除了 M_L2M,采用对称函数时效果最好,SAD 距离也非常小.在 M_L2M 情况下,男性的声道形状在 18～23 岁时基本保持不变,弯折函数对频谱的作用十分有限.在由较小年龄语音向较大年龄语音转换时,采用能量函数的转换效果稍微优于其他几种弯折函数,但整体效果一般.

将本文的方法和基于结构化的高斯混合模型的

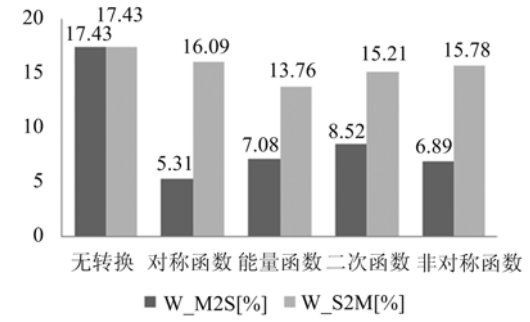

图 5 女声年龄转换频谱平均距离

Fig. 5 Average distance of spectrum of the age voice conversion for female

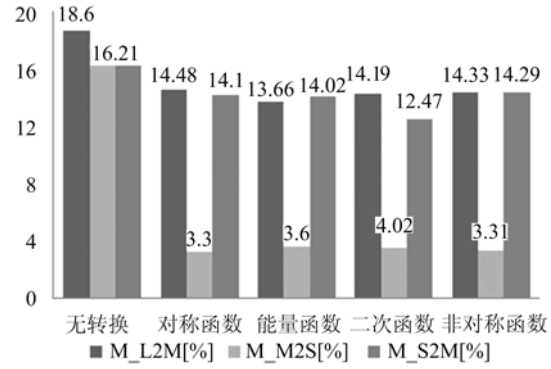

图 6 男声年龄转换频谱平均距离

Fig. 6 Average distance of spectrum of the age voice conversion for male

方法(SGMM)[9]进行比较,计算出 SAD 的结果如表 2 所示. 其中,方法 1 为本文方法中四种弯折函数下的最优情况,方法 2 是 SGMM 方法.

表 2 SAD 结果比较

Tab. 2 SAD results comparison

转换方式	SAD/%	
	方法 1	方法 2
W_M2S	5.31	10.69
W_S2M	13.67	13.08
M_L2M	13.66	12.62
M_M2S	3.3	10.87
M_S2M	12.47	12.31

由表 2 可以看出,在 W_M2S 和 M_M2S 情况下,方法 1 的效果明显好于方法 2 的效果,其余情况下两种方法转换后的频谱距离十分接近. 这主要是因为 SGMM 的方法在训练过程中容易过拟合,有时并不能达到理想的转换效果,同时此方法注重转换后像不像特定目标人的说话,并没有强调年龄信息. 虽然 SGMM 也是针对非平行语料说话人语音转换提出来的,但要求是特定人的,因此,SGMM 应用到年龄语音转换中不合适.

4.3 主观评价实验

为了进行主观评价测试实验,我们准备了一句中年女性和一句中年男性的语音,采用本文算法并选用对称函数作为弯折函数,分别转换成儿童时期的语音,在语音年龄转换效果和自然度两个方面进行打分(非常好为 5 分,非常差为 0 分). 共有 10 名测试者参加的本次试验,结果如表 3 所示.

表 3 选用对称函数的年龄语音转换效果

Tab. 3 Effect of age voice conversion by symmetric function

测试对象	年龄转换效果	自然度
女	3.8	3.9
男	3.1	3.3

从测试者的打分情况可以看出,一个中年女性转换成儿童期的语音,年龄转换效果和自然度两个方面的得分都比较高;而由于男性在成长发育期时其声道喉结的结构发生了变化,年龄语音转换效果明显比女性的年龄语音转换效果差.

5 结论

本文提出了一种基于声道长度对齐的年龄语音转换方法,该方法结包括频谱转换和基频转换两个方面,客观评测和主观评测的实验结果表明,该方法对年龄语音转换是有效的;在主观实验中,将中年女性的语音转换成小女孩的语音效果较好.

针对本文的试验结果,进一步可进行的工作包括:①频谱转换中可引入分段线性频谱弯折函数,可以对不同的共振峰采用不同的弯折方向和强度,从而更加准确地完成频谱对齐;②男性语音的年龄转换效果较差,主要是男性在成长期声道喉结的结构发生了变化,不能简单地用级联无损短管模型来描述,因此用更合适的声道模型来提高男性年龄语音的转换效果值得继续研究.

参考文献(References)

[1] Türk O, Arslan L M. Subband based voice conversion [C]// International Conference on Spoken Language Processing. Denver, USA: IEEE Press, 2002: 289-292.

[2] Tao S B, J H, Kang Y G, Li A J. Prosody conversion from neutral speech to emotional speech [J]. IEEE

Transactions on Audio, Speech, and Language Processing, 2006, 14(4): 1145-1154.

[3] Wu C H, Hsia C C, Lee C H, et al. Hierarchical prosody conversion using regression-based clustering for emotional speech synthesis[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2010, 18(6): 1394-1405.

[4] Schötz S. Perception, analysis and synthesis of speaker age[R]. Lund University, 2006.

[5] Türk O. New methods for voice conversion[D]. Master Degree, Yüksek Lisans Tezi. İstanbul: Boğaziçi Üniversitesi, 2003.

[6] Toda T, Black A W, Tokuda K. Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(8): 2222-2235.

[7] Mashimo M, Toda T, Shikano K, et al. Evaluation of cross-language voice conversion based on GMM and STRAIGHT[C]// 7th European Conference on Speech Communication and Technology. Aalborg, Denmark: ISCA Press, 2001: 361-364.

[8] Kain A, Macon M W. Spectral voice conversion for text-to-speech synthesis [C]// Proceedings of the International Conference on Acoustics, Speech and Signal Processing. Seattle, USA: IEEE Press, 1998, 1: 285-288.

[9] Zeng D J, Yu Y B. Voice conversion using structured Gaussian mixture model[C]// International Conference on Signal Processing. Beijing, China: IEEE Press, 2010: 541-544.

[10] Zhang M, Tao J H. Phoneme cluster based stated mapping for text-independent voice conversion[C]// International Conference on Acoustics, Speech, and Signal Processing. Taipei, China: IEEE Press, 2009: 4281-4284.

[11] Cohen J, Kamm T, Andreou A G. Vocal tract normalization in speech recognition: Compensating for systematic speaker variability[J]. The Journal of the Acoustical Society of America, 1995, 97(5): 3246-3247.

[12] Kawahara H, Morise M, Takahashi T, et al. TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation[C]// International Conference on Acoustics, Speech and Signal Processing. Las Vegas, USA: IEEE Press, 2008: 3933-3936.

[13] Kawahara H, Takahashi T, Morise M, et al. Development of exploratory research tools based on TANDEM-STRAIGHT [C]// Proceedings of International Conference on Asia-Pacific Signal and Information Processing Association. Sapporo, Japan: International Organizing Committee, 2009: 111-120.

[14] Wakita H. Normalization of vowels by vocal-tract length and its application to vowel identification[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1977, 25(2): 183-192.

[15] Lin Q G, Che C W. Normalizing the vocal tract length for speaker independent speech recognition[J]. Signal Processing Letters, 1995, 2(11): 201-203.

[16] Prasanna S R M, Yegnanarayana B. Extraction of pitch in adverse conditions [C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. Pittsburgh, USA: IEEE Press, 2004, 1: 109-112.

[17] Claes T, Dologlou I, ten Bosch L, et al. A novel feature transformation for vocal tract length normalization in automatic speech recognition[J]. IEEE Transactions on Speech and Audio Processing, 1998, 6(6): 549-557.

[18] 卢正鼎，丰洪才. 基于分段线性频谱弯折函数的说话人归一化方法[J]. 小型微型计算机系统，2005，25(12)：2232-2236.
LU Zhengding, FENG Hongcai. Speaker normalization method based on the Piece-Wise linear frequency warping[J]. Mini-Micro Systems, 2005, 25(12): 2232-2236.

[19] Uebel L F, Woodland P C. An investigation into vocal tract length normalisation [C]// 6th European Conference on Speech Communication and Technology. Budapest, Hungary: IEEE Press, 1999: 2527-2530.

[20] Pitz M, Ney H. Vocal tract normalization equals linear transformation in cepstral space[J]. IEEE Transactions on Speech and Audio Processing, 2005, 13(5): 930-944.

[21] Eide E, Gish H. A parametric approach to vocal tract length normalization[C]// International Conference on Acoustics, Speech, and Signal Processing. Atlanta USA: IEEE Press, 1996, 1: 346-348.

[22] Inanoglu Z. Transforming pitch in a voice conversion framework[R]. St. Edmond's College, University of Cambridge, 2003.