

电子切音器及其在语音实验中的应用

吴阿华

(北京大学无线电系)

1983年9月17日收到

在实验语音学中,为获得语音的时长参数,需要采用切音的办法。我们研制了电子语音切音器,既可以方便准确地切取不同部位,不同长度的音段,还可以做语音时长的测量,语流的重新组合等语音学实验,所得结果用数字显示。

用它做了一些音节互换,在听感上造成声调变化的初步实验。结果表明,对语音学研究有一定的意义。

一、引言

研究语音的时长,可以发现人类语言感知中的许多重要现象,切音则是实现这种研究的重要手段之一。

不论是单元音还是复合元音,切取其一定的时长可以获得识别阈值^[1-3]。辅音的时长因素也是语音的基本参数之一^[4],实验证明,如果将 *ma* 中辅音 *m* 逐渐从头部向尾部切短,使 *a* 前面的辅音部份只剩下 5ms,听起来不再是 *ma* 而像 *ba*。北京话四声的区别虽然主要表现在音高上,但四声的音长也是不同的。可见音长在声调的感知中也占有一定的地位。

切音的一种办法是把一个稳定的语音信号送到语图仪上,从所得的语图观察时长,再按一定的要求来抹音。如果不用这种办法,也可以通过剪接磁带或抹音的办法用录音机来实现切音。这种机械的办法,不仅花费时间,而且切得的音长很难准确。如果抹音不干净,更难得到良好的效果。

我们设计制做了一种数字显示电子切音器,用它做了一些语音实验。结果表明:它所切得的时长比较准确,而且能把数据直接显示出来。下面将这种切音器的基本原理,设计框

图和使用方法作一简单介绍。

二、基本原理

假设语音信号为时间 t 的函数 $p(t)$,按傅里叶变换,其频谱为:

$$P(\omega) = \int_{-\infty}^{\infty} p(t) e^{-j\omega t} dt$$

而

$$p(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) e^{j\omega t} d\omega$$

用 $g(t)$ 表示一个宽度为 τ ,高度为 A 的矩形信号,即

$$g(t) = \begin{cases} A & |t| \leq \tau/2 \\ 0 & |t| > \tau/2 \end{cases}$$

$g(t)$ 的频谱为

$$G(\omega) = \int_{-\infty}^{\infty} g(t) e^{-j\omega t} dt = A\tau \frac{\sin\left(\frac{n\omega}{2}\tau\right)}{\omega\tau/2}$$

$$n = 1, 2, 3 \dots$$

使上述矩形信号与语音信号相乘,得到

$$p^*(t) = p(t) \cdot g(t)$$

取 $A = 1$, 则

$$p^*(t) = \begin{cases} p(t) & |t| \leq \tau/2 \\ 0 & |t| > \tau/2 \end{cases}$$

变动矩形信号的宽度 τ ,则语音信号 $p^*(t)$ 的宽度也连续可变。 $p^*(t)$ 的频谱为

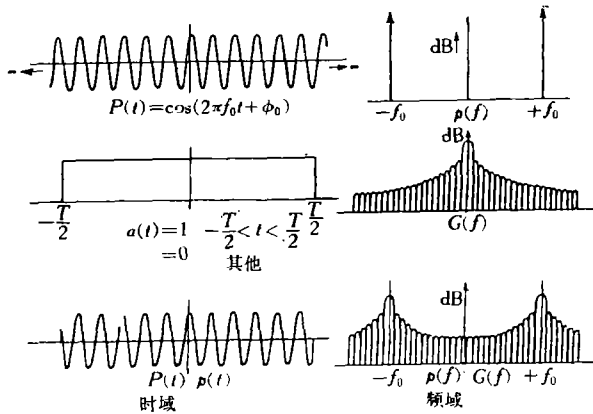


图1 正弦信号被矩形窗截取后的波形及其频谱

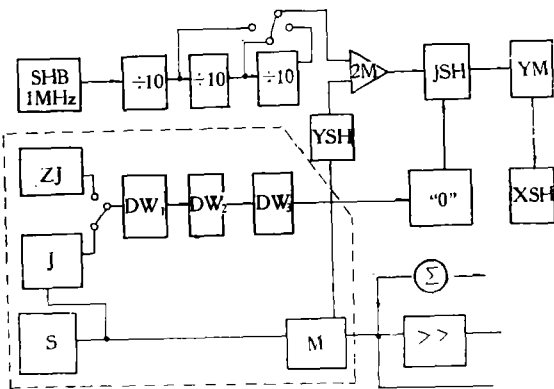


图2 数字显示电子语音切音器方框图

$$P(\omega) * G(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\nu)G(\omega - \nu)d\nu$$

两信号频谱的卷积。

为直观地说明语言波形被切取后，在时域上及频域上的变化，我们取语音信号 $p(t)$ 为单色正弦波的例，将 $p(t)$ 与 $g(t)$ 相乘，所得波形及其频谱示于图1中。

三、切音器设计框图

数字显示电子语音切音器的电路方框图示于图2。图中用点线包围的部分是切音部分，其余部分的作用是保证切时准确，显示切音时间。下面着重讨论切音部分。图2中 DW——单稳，JSH——计数，M——门，SHB——时标，XSH——显示，YM——译码，YSH——延时，JZ——自校，2M——闸门。

为了说明切音的原理，先从信号源 S 开始。信号是由循环磁带输出的。假设将已录音的循环磁带展平，示于图3。图中 T_1 是接头处到有语音信号的间隔时间， T_2 是语音信号的持续时间， T 是循环一周所需要的时间。循环磁带输出的信号波形如图4所示。图

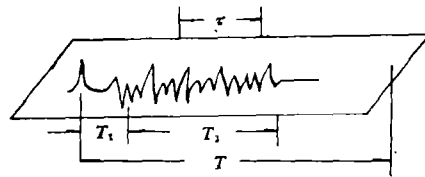


图3 展平了的循环磁带

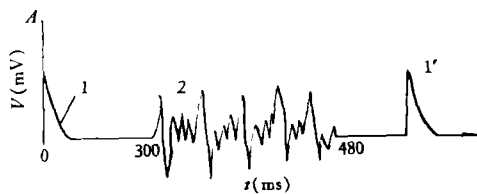


图4 待分析的语音信号

中横轴为时间 t ，以 ms 计，纵轴为振幅，以 mv 计。标有 1 的部分是标记脉冲信号。它可以是在录音前用一个特殊信号打上去，也可以是直接利用磁带接头处的缝隙形成的脉冲，但其幅度必须大于实际语音信号的幅度。标有 2 的部分，为实际语音信号。

信号源 S 的输出一路送给 M，另一路接 J。J 为标记脉冲 1 的识别器，它识别标记脉冲 1 后，即触发单稳态电路 DW_1 (DW_1 的门限电压为 2—3V 可调，视磁带输出情况具体调整之)，封锁第二个标志信号 1' 以前的一切语音信号，以保证正确识别第二个标记信号 1'。为与循环时间 T 相适应， DW_1 的持续时间要长，在设计上要采用长时单稳电路。

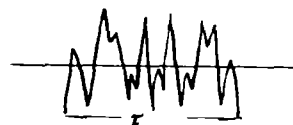


图5 门电路输出

单稳电路 DW_2 的作用是在接收到 DW_1 的

下跳变信号后延时 T_1 , T_1 的时长应从标记信号开始到要取出的语音信号处为止。

DW_2 的输出触发单稳电路 DW_3 , DW_3 输出的脉冲宽度 τ 根据所要取出的语音音长来确定。 τ 就是所取出语音信号的时长。

M 是电子模拟开关^[9], 因某些语音学实验的要求, 电路采用模拟式的。 DW_3 的脉冲前沿送到 M 的控制端, 将门打开, 语音信号便直接通过, 用 DW_3 输出脉冲后沿关闭 M, 使语音信号不能通过。因此, M 的开门时间正好是 DW_3 输出的脉冲时间 τ 。这样从 S 来的语音信号受到 DW_3 和 M 的控制, 获得所需部位(从标志脉冲到 DW_2 的后沿)及所需时长 (DW_3 的脉冲宽度 τ) 的输出。其波形如图 5 所示。

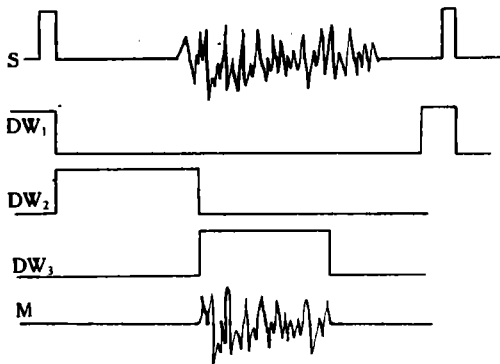


图 6 切音器各级波形

最后, 将门电路输出信号放大, 以便进行分析或主观听测。

以上所述实现切音部分的各级波形示于图

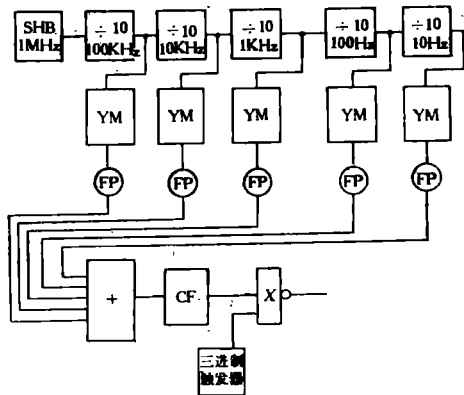


图 7 数字化电路代替三级单稳电路

6 中。

图 2 方框图中的 Σ 号是线性相加器, 可以同时输入几路信号, 实行线性相加。

为了实现全部数字化, 提高测时的精度, 图 2 中三级单稳态电路可改由数字电路组成, 我们给出数字化的框图示于图 7。

图中 SHB 为时标, YM 为译码, FP 为分配, CF 为触发。

为了减少语音拼接时的附加频率成份, 除了矩形窗外, 还设有海宁窗, 这样拼接时自然度会好一些。

四、切音器的有关技术指标

1. 单稳态电路

DW_1 脉冲宽度 T : 1s—60s, 分档可调; 误差: $\pm 0.1ms$ 。

DW_2 脉冲宽度 T_1 : 1ms—5000ms, 分档可调, 误差: $\pm 0.01ms$ 。

DW_3 脉冲宽度 τ : 0.1ms—1000ms, 分档可调, 误差: $\pm 0.01ms$ 。

以上三级单稳态电路的脉冲宽度, 均以 1MHz 的石英晶振时钟来计量。

电源电压变化对脉冲宽度的影响 电源电压从 220V 增大到 230V, 显示数不变; 从 220V 降至 200V, 显示数增加 0.01%。

2. 开关门指标

动态范围 信号由 3mV—3000mV, 失真度 < 1%。

频率响应 20 Hz—20 kHz, 输出曲线起伏 < 0.01%。

3. 放大器

动态范围 60dB. 增益 35dB.

频率响应 60 Hz—15 kHz, 输出曲线起伏 < 0.01%。

输入阻抗 配录音机输出阻抗。

输出阻抗 配有小功率输出, 阻抗为 $8\Omega, 600\Omega$, 可直接接音箱或记录设备。

该仪器已改为全数字式切音器, 时间最小为 $10\mu s$, 最大 60s, 误差 $\pm 1\mu s$ 。

五、使用方法和实验结果

我们用切音器来测量时长,切取语音,重新组合语流,以及通过模拟法合成元音,使用时仪器连接的总框图见图 8,其中 1,4 分别为循环录音机甲和乙,2 为数字显示电子语音切音器,3 为电子延时开关,5 为音箱,6 为语图仪,7 为记录仪。

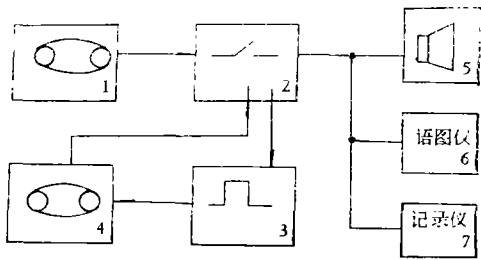


图 8 实验方框图

1. 使用方法介绍

(1) 语音信号时长的测量

测量语音信号的时间长度,可将图 8 中的部件 3,4 去掉。将待测信号录制到循环录音机 1 中,并将其输出信号送给切音器 2 的输入端,调整 DW_1 的时长 T_1 ,使与录音机 1 的循环时间相适应,再调节 DW_2 的延迟时间 T_2 ,使语音信号刚能出现,这时已找到了语音信号的起始点。然后再用 DW_3 寻找语音信号的结尾点。在调节 DW_3 时,数字显示器上不断显示出所得时间。因此,当找到语音终点时所显示的时间即为语音信号的时间长度。

(2) 语音信号的切取

单纯切取语音信号,其框图与测量语音信号时长的框图相同,方法和步骤也大体相同。但为了定出不同部位,不同长度的语音信号,需要测两个时长,即 DW_2 及 DW_3 的时长。只要拨动相应的旋钮,数码管上即显示相应的时长。切音输出可以直接进行主观听测,也可用语图分析和仪器记录。这三项可以同时进行,也可以分别进行。

(3) 语流的重新组合

语流中某些音段的重新组合,需要加接部

件 3,4。例如,“大 $d\dot{a}$, 密 mi ”两个音节包括 d, a, m, i 四个音素,要将这四个音素重新组合,其实验框图如图 9 所示。图中 A—同步录音机, B—同型号录音机, 1—切音器 1, 2—切音器 2。

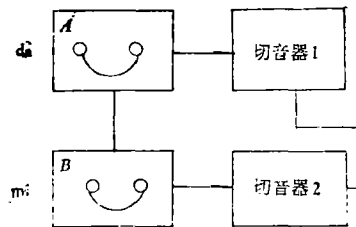


图 9 音素组合实验框图

用同步录音机 A 录得 $d\dot{a}$, B 录得 mi , 分别送给切音器 1, 2 切取信号, 使切音器 1 去掉 d , 从切音器 2 中取得 m , 送到切音器 1 的合成器(线性相加器)输入端, 就可以从切音器 1 中输出一个 $m\dot{a}$ 。反之, 如果使切音器 2 去掉 m , 从切音器 1 取得 d , 加到切音器 2 的合成器输入端, 则可以从切音器 2 中输出一个 $d\dot{i}$ 。

如果两个音节相隔时间较长, 则按图 8 连结启用延迟器 3, 使它延迟两音节的时间间隔再动作。

表 1 低平与高平结合产生中升的实验

	声 调					合成	听感
	低音调词	高音调词	从低音中取出	从高音中取出			
词	1	施行	施行	施	行	实行	✓
	2	七十	七十	七	十	其实	✓✓
	3	支援	支援	支	援	职员	✓
	4	发球	发球	发	球	罚球	✓
	5	花车	花车	花	车	滑车	✓
组	6	居士	居士	居	士	局势	✓
	7	鲜鸡	鲜鸡	鲜	鸡	咸鸡	✓✓
	8	鲜鱼	鲜鱼	鲜	鱼	咸鱼	✓✓
	9	鲜肉	鲜肉	鲜	肉	咸肉	✓✓
	10	青天	青天	青	天	晴天	✓✓
	11	燕京	燕京	燕	京	眼睛	✓✓

注 1. 表 1 听感结果, 例如“七十”在听感上变成“其实”, 可以认为是“其”的轻读“实”的重读和低调中的“七”高调中的“十”结合, 造成“七”变“其”的轻读感。

(4) 模拟法合成元音

切音器中装有 5 路线性相加器, 所以根据某个元音的声学参量, 如 F_0, F_1, F_2, F_3 等将它

们同时输入,即可合成单元音。

2. 实验结果

根据上面的方法,我们做了一些汉语中不同音高的音节互换,在听感上造成声调变化的实验^[4-7],列见表1^[8]。

表中听感栏内的“√√”表示拼接后听起来很好,“√”表示拼后听感尚可。

在上述实验中,我们还看到造成听感上声调变化程度与所取音节的时长有关。如我们取支(zhi)为322.23ms,掇(yuān⁵⁵)为414.99ms,听感上还可以,如果变长或变短,听感上就差一些。我们还注意到声调的变化程度与两个音节的相对强度有关。如fā⁵⁵取5×90mv, qiú⁵⁵取10×90mv(相对强度)听感上形成 fá qiú 较好。以上实验表明,这台切音器切音准确,拼接自然(例如低平与高平组合在一起,变成汉语中自然的升调^[4]),它的操作性能是符合预期目标的。

上述实验也使用电子计算机做过,计算机固然有速度快、调换灵便等优点,但是,本模拟

式电子切音器也有自己的特点:它切出的是模拟信号,拼接以后能够保持发音人的特色。

实验是在北京大学中文系林焘教授主持的语音实验室进行的,并且得到林教授的指导,对此表示感谢。

参 考 文 献

- [1] 吴宗济,声学报,1-1(1964),33—40.
- [2] 里查孙,声学技术概要,上册,章启康等译,科学出版社,1961,285—293.
- [3] G. Petron, *J. Speech and Hearing Res.*, 4(1961) 10—29.
- [4] 齐士铃,张家骏,声学报,7(1982), No.1, 8—13.
- [5] H. 施米德,电子模拟-数字转换,国防工业出版社,1976,63—103.
- [6] 梁之安,生理学报,26(1963), No.2, 85—92.
- [7] C. Gunnak, M. Fant, On the predictability of formant levels and spectrum envelopes from formant frequencies, *Reading in acoustic phonetics*, Edited by I. Lehiste, 109—120.
- [8] 徐世荣,普通话语音知识,文字改革出版社,1980, 100—127.

(上接 31 页)

录音棚与监听室(调音室)间的声级差为63dB(125—4000Hz平均值),因此,棚内不受监听室和环境噪声、空调设备噪声的干扰。

录音棚竣工后经声学测定和试用的结果表明,该棚具有很大的适应性,各种录音工艺均能获得满意的效果。各种可调结构使用很方便。

四、结 语

通过音研所音乐棚的工程实践可归纳如下几点体会:

1. 建立可随录音工艺而改变声学条件的音乐棚是可能的,这种新设想已经付诸现实;

2. 不规则形体的音乐棚,不仅可加强声扩散(与矩形棚相对比),同时,还能节约使用面积;

3. 折叠式声屏障不仅有声学上的功能,同时具有节约面积和便于使用等优点;

应用声学

4. 欲使混响可调幅度在63—8000Hz覆盖频率范围内接近相同,必须做到:(1)吸声面暴露时低频要有足够大的声吸收;(2)而当反射面暴露时,则要尽量减小低频的声吸收;

5. 为使可调混响幅度增大,除了增大可调面积外,还必须选择高效能吸声结构,并防止缝隙对低频的声吸收。

致谢:朝阳区第三建筑公司在施工阶段配合声学测量,音研所康昌琪先生多方协助,本院王峰,陈金京、徐游、葛砚刚等同志参加了声学测量,笔者谨致谢意。

参 考 文 献

- [1] 项端祈 马欣,电影技术,6(1982), 1(1983).
- [2] 项端祈,第二届建筑物理论文集,1(1981).
- [3] L. L. Beranek, *J. Acoust. soc. Am* 50(1971).
- [4] 北京市建筑设计院,空调设备隔振基础参考图,78:—4 (1978).
- [5] 马大猷,同济大学学报,4(1960), 6.