



现代噪声科学系列讲 座选讲之二

——噪声的产生

王世谦

(中国科学院声学研究所)

1988年9月14日收到

现代言语声学理论与实验已大体上说明了正常(normal)语音产生的原理。这指的是,我们对语音产生(除神经外的)的生理过程、声道内(以下在许多词组中声与音二字沿习惯用,具有同意;除声明外,声源均指经喉部声带调制气流转化成声的声源;“声道”包括声带;“声道内”指声门以上,不包括声带)声源和声道滤波的物理过程已基本搞清。但是对噪声在不同声区时的声源情况还了解较少;声源与声门上/下的作用还颇待研究;对于控制噪声系统各部统一有机运动的语言规律了解不够;对神经系统工作还知之极少;对噪声信号的分析与合成较清楚,但识别有限。相对言语声来说,我们对声乐声的认识更是不够。

一、噪声按产生分类

按噪声的产生,可分为单发(sustained)与连续(continued)。单发音又可按发音方法与部位不同分为清音与浊音,辅音与元音,以及不同的元、辅音。其中单发元音在声乐演唱中经常运用(例如它常出现在歌曲演唱高潮的字腹拖唱中),且其发声方法对嗓音健康也十分重要;元音在音节中占有的音长和音强比例较大,它也是语音研究的重点。因此,我们将着重以单发元音为例,来介绍噪声的产生。

由声源激励方式来看,按声带的振动与否可分为清(辅)音与浊音;再按在声道当中的声源激励方式可分为擦音和塞音。由这三种声源声的不同组合可形成一些其它的噪声,如清/

(浊)的塞/(擦)音,或清/(浊)的塞擦音。

从浊声喉源的振动模式(包括声带振动的不同情况和与其匹配的声道调节)来看,声源的产生可分为下列几种:在声乐中以声区(register)来区分,通常有胸(低)声区、中(混)声区、头(高)声区等等,但声区问题迄今还没认识很清楚,而且声区的不同术语很多(有人统计一下,多达100多个)。Wilcox建议用重机制(heavy mechanism)与轻机制(Light mechanism)以避免声区之争。这一对术语类似(但不全同)于我国民族声乐术语中的大(真)、小(假)嗓。在言语声学中,按语音声源的特点又可分为如下几种模式(register/mode):正常声(normal),[在声乐中相应为畅声(flow voice)];紧喉声(pressing voice);呼气声(breathy voice)[在声乐中,相应为漏气声(breathy voice)];嘎声(creak)和气泡声(vocal fry);耳语声(whisper):哨声(whistle)。

从噪声的传送、共鸣和辐射的主要通路来看,又有鼻音和口音之分。还有,按气流强弱来分,有送气与不送气之分。

按在声道中发音方法与部位的特点并结合感知特征来看,噪声可分为元音和辅音;不同音位的元音;不同音位的辅音。元音和辅音在声道内的不同在于:元音在声道中不受阻,辅音受阻;发元音时声道内各部较均衡紧张,而发清辅音时仅成阻部位特别紧张;发辅音时比发元音时气流急。对浊辅音则是兼有元音和清辅音两者的特征。

发声时的声源与声道在上述不同情况下的

匹配组合就能使嗓音带有种种不同的音色，从而使嗓音成为具有不同音位的元音和辅音。

1. 元音 它的音色特征可分为

(A) 元音音色 (vowel color) 它形成不同的元音音位。这主要通过舌位/(口腔的开闭)的高(闭)、低(开),前、后,及唇的圆、展来区分。这是一种由本体感受的(proprioceptive)办法来区分音位。音色、音位是听觉感知特征。舌位的生理位置与感知位置大体一致,但并非完全一致。关于舌高点的标志一直有争论,这也反映了嗓音的生理本体感受和心理感知的矛盾关系。

(B) 音乐音色 (music timbre) 它形成不同声乐感知特征。这表现为亮、暗,前、后,薄、厚。这除了与轻/(重)机制、开放/(掩盖)等发声方法有关,也由于高(头),中(混)、低(胸)音共鸣不同比例的调配。音乐音色主要通过调节嗓音源与声门上半(如舌、咽腔、下腭、软腭、喉腔、喉位、唇等)/(下半)呼吸系统之间的交互作用使其恰当匹配来达到要求。要获得一定音乐音色的生理调节比一般语音音色的生理调节要精细复杂的多,它常常需要多年有效的专业训练才有可能掌握这种调节技术。

2. 辅音 辅音的音色决定于其发音的方法(如声带振动与否、受阻与除阻的方式),气流的受阻部位、程度,和气流通路、强弱等。由发音方法来分,除前曾叙述的清音与浊音、塞音和擦音之外,辅音中还可有鼻音、塞擦音、颤音、闪音、边音、半元音、送气与不送气音等。由发音部位来分有:双唇音、唇齿音、齿间音、舌尖前音、舌尖中音、舌尖后音、舌叶音、舌面前音、舌面中音、舌根音、小舌音、咽头音、喉音等。

上述的不同嗓音一般都有不同的声学特征,通过实验的办法可加以区别。其主要方法有:声谱上有无共振峰(区别清浊)及共振峰模式(用以区别一些元音、浊辅音,即响音),能量集中区(用以区别一些辅音);浊始时长,也称相对清浊度,即VOT-voice onset time(用以区别清、浊,送气与否,和短暂辅音的发音方法等);过渡音征,即T-transition(用以区别一些暂

音)。

不同的元、辅音的部位与方式可总结而列为下面的元音舌位图和辅音表。(略)

二、嗓音产生模型

下面我们只拟介绍嗓音产生的基本过程中的生理解剖和物理声学方面,对于神经系统对发声肌肉运动的操纵、协调和指令过程,则不拟讨论;并在此基础上归结成数学模型。

1. 嗓音产生的生理解剖方面(图略)

人的嗓音系统包括:声门下半系统(肺、气管、喉下腔);声带;声门上系统[喉腔,咽腔,口腔(舌、悬雍垂、软腭、硬腭、齿龈、齿和唇),鼻腔,鼻窦,还有口、鼻出口],以及与它们活动有关的肌肉和神经。嗓音产生的基本过程是:由于呼吸肌收缩(声乐发声中吸肌也同时收缩,形成对抗平衡)使由肺而来的空气经过声门声带振动、声道成阻或共鸣后,部分转化为声能且带有语言信息,再经口、鼻孔发射而成为嗓音。声门下半系统的主要作用是供应呼吸和发声的(平流)气流;声带的主要作用是进行振动,调制气流成为准周期的浊声源;声门上系统的主要作用是通过积极发音器官的运动来调节声道形状和长度,形成一定的浊声共鸣腔或形成阻碍,调制气流成为噪声源。嗓音产生生理的研究通常包括对嗓音系统的以上三个部分在发声中的作用和机理。在一般语音的产生中,呼吸运动及声带振动可近似地被视为单一运动模式,而声道的变化则被视为形成不同语音的变量,所以已往的语音研究大都集中于声道内的一般发音器官,尤其是口腔各部与鼻腔。而在声乐发声研究中,由于呼吸的方式(如胸腹联合式呼吸的支持)及其细微的调节,声带振动机制(如轻、重机制)、模式(如由不同的开闭商、速度商所形成的声带各种振动模式)和声区(如高、中、低、声区)的多样性,喉和咽腔的形状变化,以及气管、鼻腔和鼻旁窦的共鸣等都对音色及发声功能影响很大,所以唱声产生的研究遍及于嗓音系统的各个方面。七十年代中以来,在嗓音产生的

研究中,对声门及声门与其上、下系统的交互作用已有了很多由现代手段研究的新成果(将在“近来关于噪源及其与其上、下的交互作用的研究”详述)。

2. 噪声产生的物理方面

人的噪声系统可视为乐器,从物理观点来分析,可分为四个部分,即

(A) 能源 它由肺内气压形成并经气管输送的气流供给,能量由声门体积流而定。呼吸在歌唱、言语、安静时的特点是互不相同的。为了适应发声和新陈代谢的需要,在声门下形成或保持在一定时间内所需要的一定气压(喉下压在言语发声时约为十几个 cm 水柱,声乐发声时其动态范围较大)这三种情况下的呼吸的频率、时长比、呼吸量以及参予的肌肉活动都是不同的。歌唱和言语中呼吸的频率、时长比都是按发声需要和呼吸生理能力而定的。发声呼吸的频率可在每分钟几次到几十次之间,呼吸时长比可高达十多倍;一般地唱比说的呼吸频率要低,时长比要大。呼吸量在安静时是大约 500ml,言语约为其 2—3 倍;歌唱约为其 3—5 倍。在唱声中为了保持稳定的空气体积流,呼气时吸肌群仍需保持适当的收缩状态,以能和呼肌群收缩力构成对抗平衡,维持稳定渐变的气压。声乐家是靠胸腹联合式呼吸法来达到此目的的。

(B) 声源 呼吸气流经声道调制而转化成声,即形成噪声源。噪声源涉及呼吸气流和气压转换成声能的方式和位置。噪声源基本上可归结成两种调制类型:

(a) 对浊音来说,是由声带振动时声门的开闭(声门面积变化范围约在 0—20mm²)对由肺而来经声门的直流气流进行准周期的交流调制,这种由调制而成的声源,即声门喉源波,也称为声门激励源。声带的振动频率是该噪声的基频。通过声门喉源波的频率、强度和频谱的变化,可提供噪声的韵律和感情的信息。声门关闭瞬间的声门体积速度波形斜率的变化决定噪声高频谐波的产生和其包络。

(b) 对清音来说,是由声道的局部成

阻把气流(部分能量)转换成了噪声。在声道内的噪声源产生中,擦音的情况是:高速运动的平(气)流经过声道狭窄成阻处之后,当雷诺数(Reynolds number)超过临界值,就转换成为湍流,它为准随机噪声激励源。无量纲的雷诺数 $Re = \nu h / \mu$, 其中, ν 为气流速度, h 为成阻狭处截面有效径长, 气体动粘滞系数 μ 是粘滞系数与气体密度之比,在噪声系统擦声产生中, μ 可视为常数并等于 0.15cm/sec。雷诺数临界值依赖于管道的几何形状(以及表面糙度、声道外部振动等其它因素),例如对圆管,其值约为 2100;而对声道狭处成阻(远异于圆管)情况,据 Meyer-Eppler (1953)的模型试验结果,其值约为 1800。另外,气流速度和窄处截面形状大小决定着气流质点速度转换成湍流噪声声压的效率。 啞音 [s, z, f, ʒ] 和非啞擦音使用不同的产生机制:前者,例如 [s] 是使气流通过成阻狭部加速,再经齿缘障碍形成涡流,它在声谱上能量分布在 5000—9000Hz 频带内,且集中在 5500, 6500, 8500Hz 等附近;而后者则属导管型湍流噪声源 (Peterson & Shoup, 1966; Stevens 1971),其能量,例如 [f] 差不多均匀分布在 3500—10000Hz。对塞音的情况是:声道某处完全阻塞气流,在其前后形成压差,然后突然除阻,产生跃阶压变和摩擦噪声(或送气),形成瞬时(噪声)激励源。其特点是,这种跃阶压变发生于很短的时间区间内,并伴随着质点高速运动;它具有脉冲;其连续谱按 6dB/oct 下降。由不同的声源方式与不同的噪声源部位的特征,再和气流强弱、气流发射通道、气流受阻程度等特征进行各种发声组合,就可形成了多种噪声源。在声源进行组合时,有交互作用产生,并非线性叠加。

(C) 共鸣 声道可看做是在口端开、声门端闭的“半闭声学管”。对清辅音来说,对应声谱上不出现共振峰,但可有能量中心。对元音(浊辅音类似)来说,通过积极发音器官的运动来调节声道形状和长度(由截面积函数表示),构成一定的共鸣腔,形成共振,声谱上有相应的共振峰模式。浊音的振动模式主要由声道中三

个生理变量决定: 主要收缩位置, 收缩程度和圆/(展)唇程度。这三个变量的具体组合和它们之间的交互作用可确定元音产生的声学共振参数。为区别各共振峰的感知特点, 不妨划分成两类共鸣。其一是咬字共鸣, 它使嗓音形成不同元音音色。它主要由声道内口腔的各积极发声器官(特别是舌)的运动, 使声道形成一定的形状, 从而对应一定的共振频率分布模式(其中最低的两个共振峰就可确定元音)。其二是音乐共鸣, 它使嗓音带有一定的音乐音色和个性特征。这是由于声道在保持咬字共鸣腔的基本形状(即 F1、F2 维持在其元音音位变体范围之内)的基础上, 对口、喉、咽、鼻等腔进行适应声乐发声要求的调节, 而且气管和鼻旁室的共鸣也赋予嗓音以特别音乐音色特征。在声谱上它主要表现为高共振峰的增强和歌手共振峰的出现以及各言语共振峰的偏移。Sundberg (1974, 1977) 认为喉管是高歌手共振峰的共振腔。Granán & Boves (1987) 等认为: 气管是低歌手共振峰的共振腔, 它的共振频率约为 500Hz, 1400Hz 和 2300Hz, 其中 500Hz 对声区转换有最重要的意义 (Titze, 1988); 许多学者认为副鼻窦是增强高频的共振腔体。迄今, 对这些解释还是存在很多分歧的。由于这些共鸣作用, 使得嗓音成为带有明亮或宽厚等音色特点的声乐声 (Wang, 1986), 而且由高共振峰的表现样式, 又反映了嗓音的个性特点 (Ladefoged, 1982)。

(D) 辐射 嗓音的辐射器是口、鼻腔的出口(皮肤封闭部分可略不计)。语音基频较低(一般成人在 400Hz 以下), 口、鼻出口径长相对波长很小, 且变化不大, 可视为常数, 这样, 它的辐射特性是 $R(f) = \frac{jf\rho}{4\pi R} e^{-j\frac{r}{c}}$, 其中 c 为声速, r 为到唇距离, f 为频率, ρ 为空气密度; 其振幅可以取作 $R(f) = \frac{f\rho}{4\pi R}$, 与频率成正比, 即其斜率为 6db/oct。但在声乐发声中, 基频常很高, 甚至可达 1000Hz 以上。随着音高上升, 波长变小, 而且发音口形渐大, 口径与波长差别

渐小, 则口部辐射就渐不能视作简单点源, 且头部的衍射作用渐大, 情况就复杂了。根据 Flanagan 的实验 (1960), 基频在 300Hz 以上时, 高频辐射振幅增强, 至少比语音每倍频多增加两个 dB。另一点应注意的是, 口音与鼻音的差别不仅是由于口道及鼻道共鸣的使用, 更是因其辐射出口的不同。对鼻音来说, 由于软腭的下降而使鼻道与口道连通耦合, 嗓音的大部或全部经由鼻孔辐射, 它与口音的音色可以形成音位差别。还有, 声乐声中的鼻腔共鸣与语音学中的鼻音并非等同。在声乐声中的鼻腔共鸣主要是指歌手高共振峰的形成(涉及喉管、鼻旁室的共振), 而且也指嗓音在主要经口道辐射的同时兼有适当的比例经由鼻孔辐射。通常在发具有鼻腔共鸣的嗓音时, 面部能感觉到振动, 这即声乐声中的所谓“面罩”(mask)。

3. 嗓音产生的线性数学模型

为了方便, 我们打算采用“终端模型”的办法来处理, 即我们要建立一个线性系统模型, 用一组参数, 对其控制, 使其输出能和真实语音相同。这种模型在终端处和真实语音等效, 但其内部结构未必和嗓音产生的物理过程一致。

我们已知, 嗓音可看成是由两种声源途径产生的, 它们各自产生不同的输出形式。在元音发声时, 在声门嗓音源的激励下, 声道根据其特有的共振模式, 产生不同的元音。其实, 嗓音产生的实际过程是十分复杂的。这是因为: 嗓音源机制的多样性; 声带的多质量肌弹振动; 声道的复杂时变性; 声道壁随发声法和声区而变的表面张力和弹性, 以及由热传导和粘滞性而引起的能量损失; 声道中声的激发; 鼻腔与气管的耦合, 以及声乐声的多腔共鸣和多孔辐射特性; 声门与其上、下系统的交互作用。要全面考虑这些因素, 会使问题极困难复杂。为了便于处理, 可把元音产生在一小段时间(语音在 10—20 ms, 拖唱元音在更长时间)内视为一个线性时不变的物理过程, 而且把声门、声道、口鼻辐射的作用视为各自独立, 即

声源激励—声道共鸣—口鼻辐射—嗓音输出
由声电类比线路可以建立嗓音产生的电模型并

进至数学模型。这样，在口外的辐射声波 $s(t)$ 应是激发源 $g(t)$ 、传递特性 $v(t)$ 和辐射特性 $r(t)$ 三者的卷积。即，

$$\text{在时域有 } s(t) = g(t) * v(t) * r(t)$$

$$\text{在频域对应为 } S(f) = G(f)V(f)R(f)$$

发元音时，源频谱是线谱，谐频的大小 $G(f)$ 与 f 的平方成反比；声道滤波函数 $H(f)$ 可变且在声道共振(峰)处有极点；辐射特性大小 $R(f)$ 与 f 成正比。为方便处理，用 z 变换，则有

$$S(z) = G(z)V(z)R(z)$$

这就是噪声产生的数学模型。

将激励源看做一个源滤波，可令 $G(z) = U(z)E(z)$ ，其中 $U(z)$ 是源滤波器，则

$$S(z) = E(z)U(z)V(z)R(z)$$

把源滤波器、声道滤波器和辐射滤波器对噪声辐射谱的贡献合并成一个滤波器，即再令

$$H(z) = U(z)V(z)R(z)$$

$$\text{则有 } S(z) = E(z)H(z)$$

这成了一个简化的噪声产生终端模型。式中的 $E(z)$ 是声门处的体积速度。对浊声来说，它是周期性脉冲；把它也可推广到清音，那就是白噪声；而浊辅音可简化视为元音和清辅音的组合。

对鼻音和擦音来说，频谱有谷点，即其传输函数应有零点。如果零点不太靠近极点的话，对音色影响不大。我们不妨设 $H(z)$ 是无零点仅有极点的全极系统，其传输函数形如

$$H(z) = \frac{G}{1 - \sum_{k=1}^m a_k z^{-k}}$$

把它代入 $S(z)$ 并取逆 z 变换，可得噪声产生模型的计算表达式为

$$s(n) = \sum_{k=1}^m a_k s(n-k) + Gu(n)$$

其中，增益参数 G 和 $\{a_k\}$ 是依赖于面积函数的参数。这个表达式也可用线性预测的办法得到。那时，式中 $s(n)$ 为第 n 个样点的语音信号， $Gu(n) = e(n)$ 为误差， a_k 则为预测系数，它可用最小平方误差的方法加以确定，用线性预测计算语音参数十分简便。

这个模型对元音，特别是持续元音(如声乐中拖唱母音)是很适用的；对鼻音则受到限制；对擦音情况稍好(必要时也可用补零点的办法来解决)；对塞音这样的暂音，这模型不太好，但仍可用；对浊擦音也不太好，这是因为摩擦与声门流的峰值有关，应考虑它们的交互影响。但在线性预测法中把阶数取得足够高，全极点的模型几乎可以表述所有的语音。

目前一般噪声的，特别是言语声的分析、合成、识别等，大都是以此数学模型和表达式为基础而进行信号处理。对噪声分析来说，是根据上述模型及其表达式，从辐射噪声信号中提取声源激励的参数或声道传输特性参数；对噪声合成来说，则要根据上述模型调制声源特征参数和声道特征参数合成噪声。此外，许多语音识别也是根据上述模型及其表达式先对语音信号分析，提取参数再进行模式匹配的。

显然，进一步改进噪声的产生模型的精度，尤其是要使其符合各种机制的声乐发声，需要考虑更多的因素，这有待进一步研究。在近些年的文献中，已有一些在上述基础上改进了的噪声产生模型报导。

主要参考文献

- [1] Borden, G. J. & Harris K. S., *Speech Science Primer*, The Williams & Wilkins Co. 1980.
- [2] Cranen, B. & Boves L., *J. Acoust. Soc. Am.*, 81(3), (1987).
- [3] Broad, D. J. (editor), *Topic in Speech Science*, Speech Communication Research Lab., Inc., Los Angeles, California, 1977.
- [4] Denes, P. B. & Pinson E. N., *The Speech Chain*. Anchor Press, 1973.
- [5] Fant, G., *Acoustic Theory of Speech Production*. The Hague, Mouton. 1970.
- [6] Fry, D. B. *The Physics of Speech*. Cambridge Univ. Press, London, 1979.
- [7] Ladefoged, P. A., *A Course in Phonetics* (2nd ed.), Harourt Brase Jovanouich, Inc, 1982.
- [8] Rabiner, L. R. and Schafer, R. W., *Digital Processing of Speech Signals*, Prentice-Hall Inc, New Jersey, 1978.
- [9] Massachusetts Institute of Technology, *Speech Communication: Special Summer Course* (Lecture and Lab notes), 1982.
- [10] Minifie, F. D., Hixon, T. J., and Williams, F. (ed.),

(下转第42页)

把握好三个最佳点、三个保证、一个合理，就能取得高的超声加工效率。

三、期望和展望

超声加工业，自然离不开超声加工设备。据我们多年的实践经验，理想的超声加工设备，应具备以下几个条件。1. 要有一套从总体设计上就匹配得很好的振动系统及其工艺资料；2. 要有一套能体现超声加工工作原理的机床结构，其中包括负气压系统，磨料循环系统，静给压力系统，机械调节系统等；3. 稳定、可靠，操作简便、舒适，维修方便。我们希望，从事超声加工的科研单位和生产厂家，能结合生产中的需要，研制与制造出理想的超声加工设备。

超声应用于玉石加工业，不仅开拓了超声的应用领域，也改变了玉石加工业中的一些传统工艺。这不仅提高了工作效率及产品质量，而且也解决了用手工加工玉石所遇到的某些困难乃至不可克服的技术问题。我们深信，随着科学技术的发展及在生产实践中不断总结经验。超声在玉石加工业中的应用领域会不断深化、扩大。

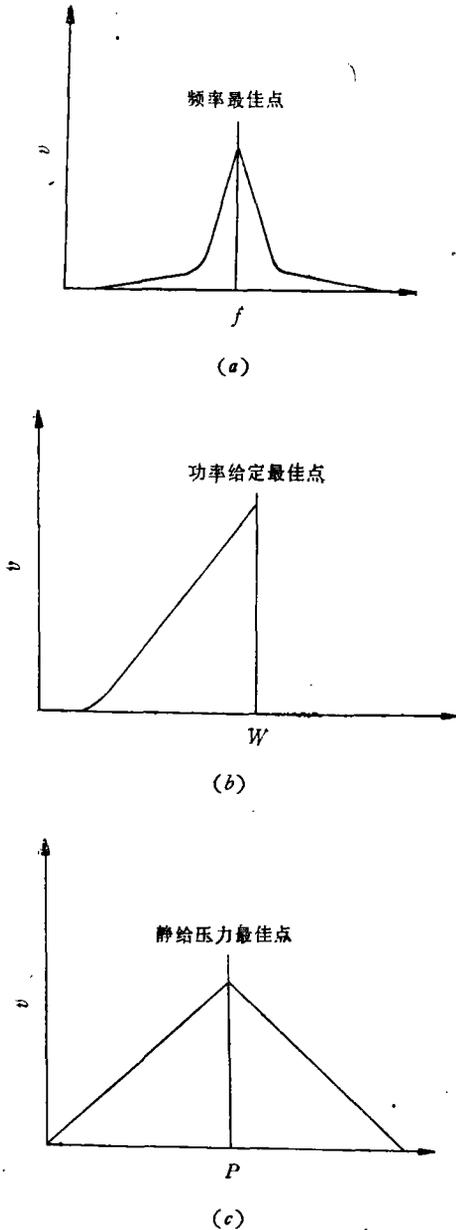


图2 最佳工作状态选取示意图

(a) 频率调谐最佳点，即振动系统的真正谐振点；(b) 功率给定最佳点，即振动系统不失配情况下的最大允许功率；(c) 静给压力最佳点，即在一定功率下的最大允许静给压力。

(上接第47页)

Normal Aspects of speech, Hearing and Language, Prentice-Hall Inc. New Jersey, 1973.

- [11] Sundberg, J., Voice tract resonance in singing, NATS, Mar/Apr., 1988.
- [12] Titze, I., Subglottal resonance, NATS, Mar/Apr., 1980.
- [13] Wang, S., *J. Acoust. Soc. Japan* (E), 7-6 (1986).
- [14] Wang, S., The extra formant, a cluster of F3F4F5 and Others, Proc. 12th ICA (1986 Toronto), K1-3.
- [15] 高名凯、石安石, 语言学概论, 中华书局, 1963.