

本章讨论发音的声学方面，并探讨发音在言语交际中的应用。前几小节集中在声源信号上，声源信号通常被定义为通过声门的气流或体积速度。它以周期性的方式变化，反映了振动着的声带的快速打开和关闭。这里不讨论不振动声音（清音）的声源。

在语流中，声源信号是不断调制的，这些调制向听者传送不同种类的信息。其中一些变化是实质性的，听众可以清楚地听到音质的变化，如正常嗓音和气嗓音。有些调节不是以这种方式被感知的，而是以其他方式对信息的意义或者对我们识别说话者的社会或个人特征的能力产生影响。在本章的后续部分中，我们讨论了各种声源变异的决定因素，以及他们在语言学、副语言研究、社会语言学、口语交谈中的超语言/语言以外的非语言因素中的作用。

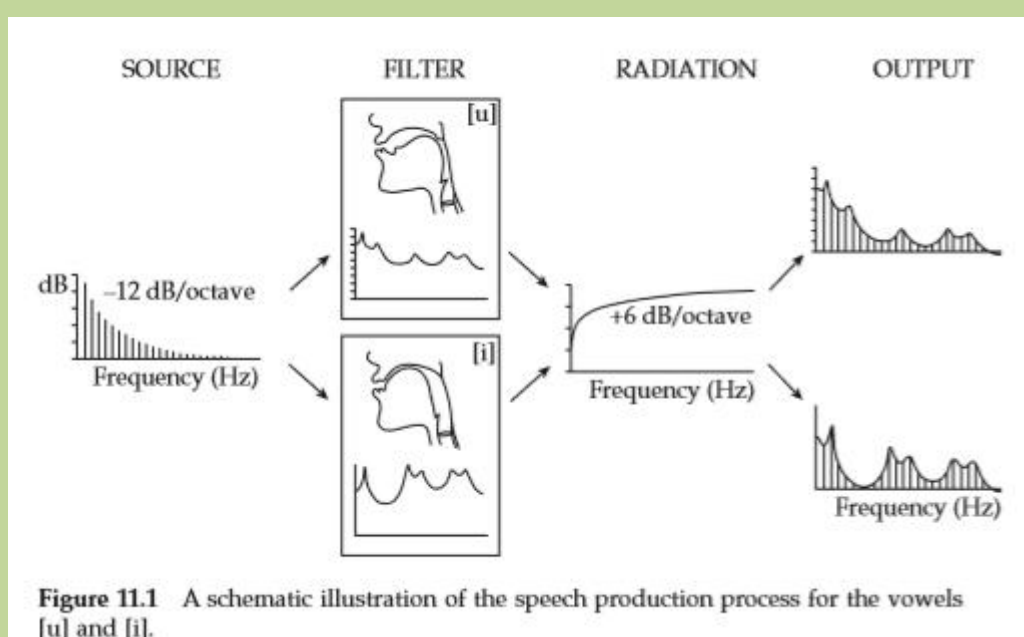


图 11.1 元音[u]和[i]的语音生成过程示意图

声门的气流构成了声道的输入信号，而声道在这个过程中充当了声学滤波器的作用。声门上的发声器官的结构决定了这种滤波器的特殊共振特性。对于一个给定的声源信号，可以根据不同声门上滤波器所产生的特定的共振和反共振模式来区分各个部分。

图 11.1 是两个元音[u]和[i]的发声过程示意图。在这个例子中，声源的(振幅)语图在两种情况下都是相同的：它包含所有谐波分量，斜率常数为-12 分贝/倍频程。这意味着谐波的幅值随着频率的增加呈单调下降的趋势，每增加一倍的频率，幅值就会下降 12 dB。我们应该注意这只是一个理想的情况。真正的声门频谱并没有一个恒定的斜率，并且可能根据声门脉冲的精确形状呈现局部扰动。

由于舌头和嘴唇所处的位置不同，声道的过滤作用（即传递函数），对于两个元音来说是相当不同的。在传递函数峰值处或附近的声源谐波将被滤波器放大。不接近峰值的谐波将不会被放大，并可能被减弱。因此，滤波器的输出，即，口语气流的频谱显示出波峰和波谷，而不是相对均匀的下降的声源频谱，这些决定了我们听到的声音的不同分段性质(在这个例子中表现为[u]和[i]之间的差异)。最后，与口语气流的频谱斜率相比，辐射声压的频谱倾斜幅度约为每个倍频程+6 dB

图 11.1 中的示意图是一个理想化的声源，假定这两个元音是常数/常量。在真实的言语中，声门声源的动态变化反映了声门的结构、可能存在的任何喉部紧张的程度和类型、所使用的呼吸运作，甚至声门上狭窄的空气动力结果。Gobl(1988)举例说明了在使用中性、正常发音模式的单个话语过程中，音源可能发生的变化。如果说话者选择在不同的发音模式之间切换(例如，气嗓音、挤喉音等)，这种差异就会更大。不同的说话者在他们使用的习惯性发音类型上也可能有很大的不同

近几年，人们对滤波器的声学特性进行了大量的研究，它与发音位置和方式的许多部分的不同有关。关于声源，人们对 f_0 的变异有很多了解，以及 f_0 是如何随着语调、音调和重音的变化而变化的。相对而言，对于声源的其他层面以及它在语流中的变化则知之甚少。当然有许多研究语音信号的强度变化的。虽然语音输出的振幅在某种程度上反映了声源的振幅，但是我们应该牢记言语输出的总振幅是一个声源和滤波器相互作用的结果。

在下一节中，将讨论分析和测量声源的方法。在第 3 节中，我们将简要说明不同音质的声源是如何变化的。在第四节中，我们概述了决定语音和语言中声源变化的因素。

2 声源分析

2.1 获取声门气流:逆滤波

对声源的实验研究大多基于逆滤波。这种技术有效地逆转了语音的生成过程。语音信号通过滤波器传递，滤波器的传递函数是声门上传递函数的倒数（两者相反）。原则上，当声道的过滤效果被取消时，将会以过滤前的形式保留声源。图 11.2 说明了这一过程在频域(以信号的频率分量表示)和时域(以声门气流或其衍生物表示)。这里没有显示唇辐射的消除，原因如下（原因在下面会解释）。

逆滤波器应包含一个有确定频率和带宽规定的反共振器(复共轭零点)，以便在任何给定的时间及时的取消共振峰(复共轭极点)的声道传递函数。重要的是要为逆滤波器使用正确的反谐振器的数量，采样频率决定了信号的带宽。共振峰(极点)之间的平均间距由声道的长度决定:对于声道长度为 17.5 cm 典型的男性，声道为平均每 1000 赫兹有一个共振峰。例如，设定采样频率为 10kHz(即，信号带宽为 5kHz)，声道长度为 17.5 cm，反谐振器数目就应该是 5 个。精确的频率和带宽对低共振峰，尤其是 F1 峰来说是非常关键的。这里的任何错误都会导致声门脉冲的某些失真。高共振峰的微小误差对主脉冲形状及其频谱影响不大(Gobl, 1988)。

全极函数充分描述了元音等多种声音的传递函数。对于某些特定的声音，如鼻音和边音，声道传递函数包含零点和极点，原则上这些零点应该通过在逆滤波器中包含相应的极点来抵消。由于通常很难估计传递函数的零点，大多数研究人员在实践中倾向于对所有声音使用全极点模型。尽管这简化了逆滤波器规格，但它确实意味着，其频谱包含零点的声音是不太准确的过滤。Mahshie 和 Gobl(2003)探讨了这种简化如何影响鼻音化语音的声源估测。

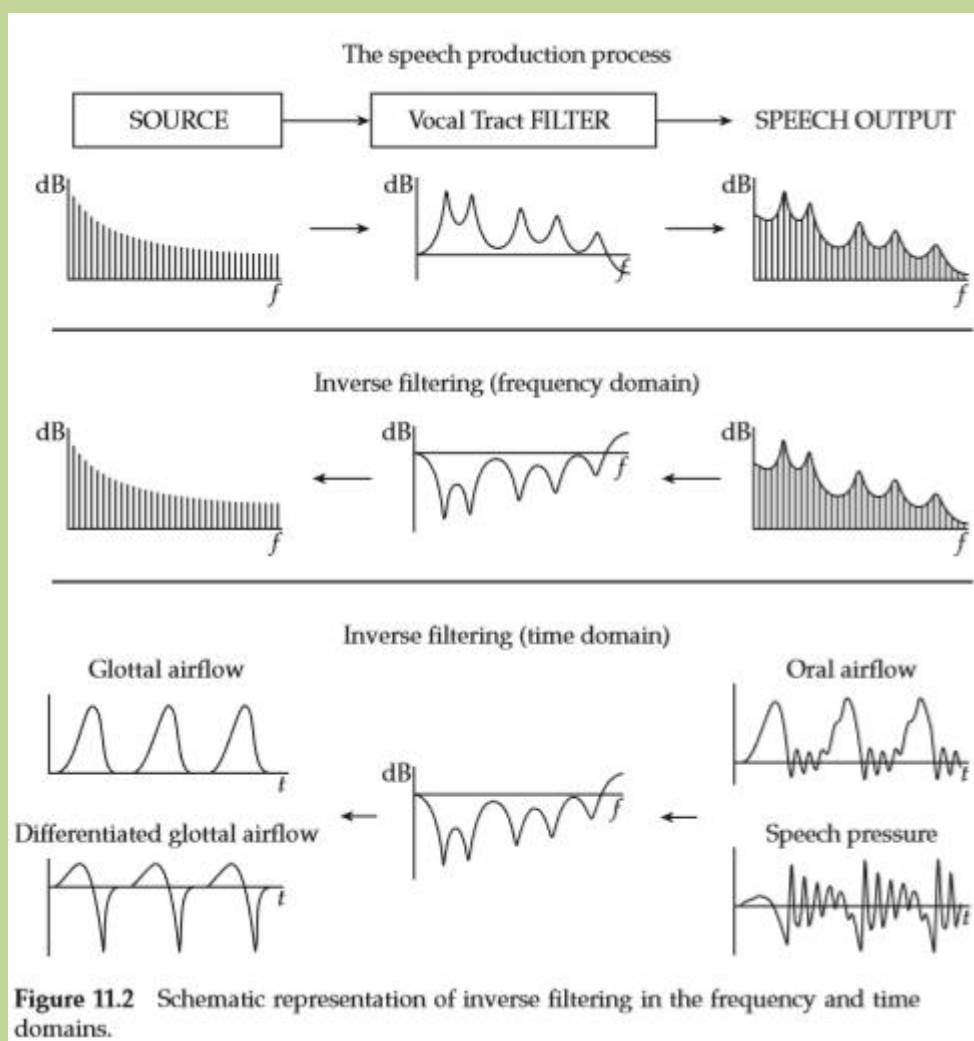


图 11.2 逆滤波在频域和时域的示意图

为了从语音压力波中获得真实的声门气流，还需要消除唇部声辐射的滤波效果。辐射特性可以相对精确地近似为一阶差分(然而，参见 Fant, 1960, pp. 44-5, 以获得更详细的描述)。微分的频谱结果是每倍频程相对增加 6 分贝。这种效应很容易被简单的信号积分(零频率的实极)所抵消，因为这是微分的倒数。如果不消除唇辐射的影响，逆滤波器的输出将对应于分化的声门流，也称为声门流的导数/衍生物。许多研究人员选择使用这种信号而不是真正的声门气流。每倍频程提高 6 分贝的频率可以更精确地模拟声源信号的频谱斜率。声门闭合时最大气流不连续点/处所决定的声门激发/兴奋强度也更容易得到(见进一步的 2.4 节)。将唇部辐射与声源重新组合也会比较方便：这一声源在第一开始不需要时应当去除，然后需要时可以再重新引入。

基于语音压力波的逆滤波可以获得详细的时间和频谱信息。然而，记录设备和空间是至关重要的，任何一种情况下的缺陷都可能导致测量中的工件(参见 Ladefoged 等人的讨论和评论，1987)。理想情况下，应该使用消声室，尽管一个阻尼良好的“半消声”工作室通常也是足够的。即使在非常低的频率条件下，记录设备也必须保持信号的相位特性。

为此，麦克风和放大器的选择至关重要。建议使用高质量的数字或 FM 录音设备，避免使用任何可能引起相位失真的内置过滤器、数据压缩技术等。模拟猿类记录器会扭曲相位:然而，有人提出了如何补偿这个相位的建议(Holmes, 1975; Hunt, 1978; Liungavist and Fujisaki, 1985; Hedelin, 1986)。录音通常需要一些高通滤波，以消除一些在录音室不可避免的听不清的低频压力波。这是至关重要的，以确保这种滤波是用线性相位来响应滤波器。

对口腔/口语气流的记录也可以进行逆滤波。在这种情况下，可以使用一种内置差压传感器的特殊气流面罩/面具:许多研究使用了 Martin Rothenberg (Rothenberg, 1973)设计的环形通气速度记录仪面罩/面具。当口腔气流被反向过滤时，输出是对真实声门气流的估计。如果需要不同的声门流量，则在逆滤波器中添加一个一阶微分器(在零频率下为实数零)。使用口腔气流记录的主要优点是可以测量气流速度的绝对值，而这在语音压力波的记录中是不可能做到的。这对于测量发声过程中的“直流泄漏”特别有用，因为在所谓的闭合阶段，声门循环没有完全闭合。这种方法的主要缺点是面罩/面具的频率响应有限。即使使用 Rothenberg 设计的特殊面罩/面具，频率响应也被限制为略大于 1kHz(见 Badin et al. 1990; Hertegard & Gauffin, 1992)。因此，它没有提供详细的声源的频谱分析。

一个成功的声源分析，对估计声道传递函数的估测当然应该是准确的。许多用于估计逆滤波器的系统都涉及全自动程序，通常基于某种形式的 LPC。不幸的是，它们往往不能为详细的声源分析产生令人满意的结果，特别是在声道过滤器正在经历快速变化或发非正常状态的声音时。目前，最准确的声源信号是通过使用者交互微调逆滤波器的共振峰频率和带宽来获得的。

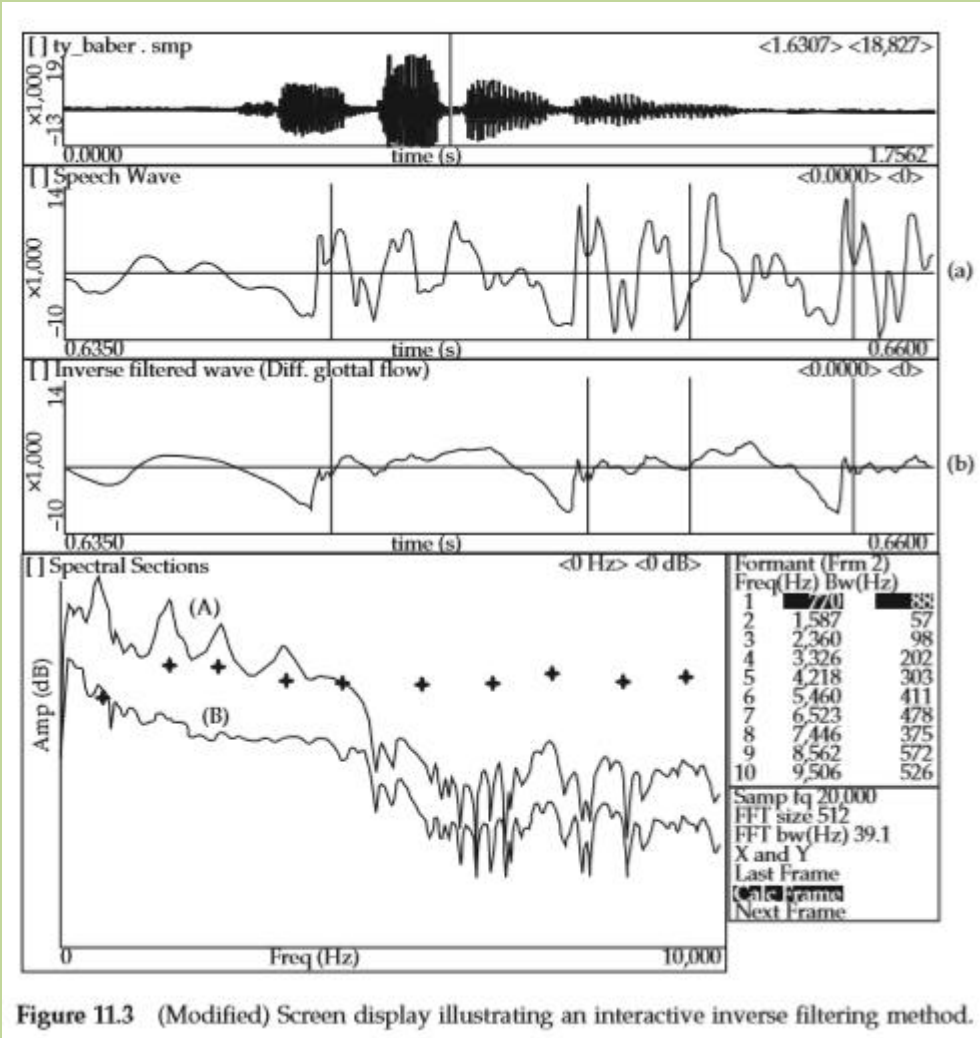


Figure 11.3 (Modified) Screen display illustrating an interactive inverse filtering method.

图 11.3(修改后的)屏幕显示说明了一种交互式逆滤波方法。图 11.3 是一个稍作修改的屏幕显示，显示了时域和频域信息，引导使用者/用户消除共振峰峰值(频域)和共振峰振荡(时域)。上面的窗口显示的是语音波形图，用光标来标记分析的脉冲。第二和第三个窗口显示(a)该脉冲的语音波形和(b) 输出的逆滤波 (差分声门流)。最低的窗口显示相应的频谱(A)语音波形和(B)分化的声门气流。在最低窗口中标记为交叉的点

表示共振峰，决定了逆滤波器的复杂的零点。这些点都可以在水平或垂直方向上移动以分别控制调节频率和带宽。每次操作后，屏幕都会立即更新，以显示新的逆滤波波形(b)及其频谱(b)。有关此特定装置的进一步细节，请参见 Gobl 和 ni Chasaide (1999a)。

目前还没有一种自动程序能达到训练有素的研究人员所能达到的精确度水平。使用时间和频率相结合的信息，可以比其他方法更准确地测量声源的许多方面。然而，这种方法也有缺点。在微调滤波器时，有时需要在时间和频率信息之间进行折衷，这里采用一致的方法是至关重要的。当然，这需要相当多的技能和经验，并带来风险，不同的实验者将采用不同的策略，导致不一致的结果。即使是训练有素的用户，这个过程也存在循环的风险。因为实验者对声门气流有一定的预期。它可以避免看起来不太像但却有效的脉冲形状。但最大的问题可能是手工交互方法不适合分析大量数据。由于分析通常是逐脉冲进行的，因此非常耗时。这一点，以及使用时所需的高度警惕，导致这些类型的研究仅限于在少量详细分析数据时才使用。

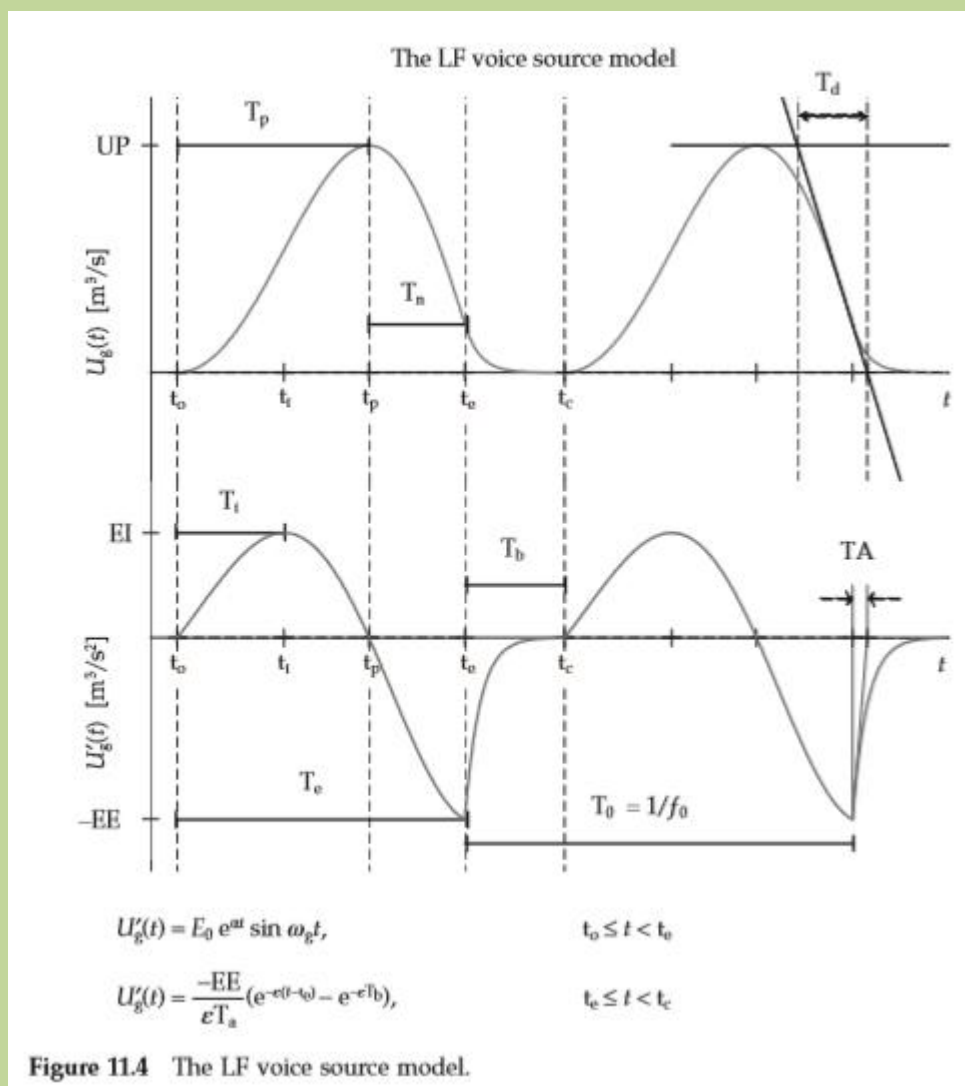
2.2 声源模型

如上所述，逆滤波器的输出信号是对声门气流或其衍生物的估计。视觉检查可能对某些特征有初步的印象，如声音是否有效、是否有气嗓音等，但要进行精细的比较，则需要精确的测量。在若干研究中，这些测量是直接从未测的声源信号入手的(例如， Huffman 1987; Hertegard & Gauffin, 1991; Holmberg 等, 1988; Lofqvist & McGowan, 1991; Laukkanen 等, 1997; Alku 等, 2002)。另一种方法则是将声源模型参数与从逆滤波中获取的脉冲进行匹配，以及从声源模型波形图中推导测量方法。

这一方法如果想获得成功，重要的一点是，该模型是一个很好的真实声源的表示，而且应当足够灵活以便捕捉可能发生的重要变化。传统上，参数化语音合成中的声源采用低通滤波脉冲序列(e. g., liljencrant, 1969; Klatt, 1980)。这种简单声源的控制参数只有 f_0 和脉冲幅度/振幅。它的主要缺点是频谱斜率无法控制:它总是完全规则的，以每倍频程 12 分贝的速度单调下降。另一个缺点是滤波脉冲的相位特性与典型的声门脉冲有很大的不同。与典型的声门波形相比，低通滤波器的脉冲响应是时间反转的。这意味着波形的主要不连续(对应于主兴奋/活跃处)发生在声门脉冲的上升支/部分上，而不是下降支/部分上。

这些缺点导致这种类型的语音合成器合成出来的声音音质不灵活并且往往难以令人满意，这促使我们开发出更精细的声源模型。这些模型都具有更多的控制参数，能显示出更准确的声门波形。因此，它们更能捕捉自然声门波形的频率特性(e. g., 频谱斜率和相位特征)。

2.2.1 Liljencrants-Fant (LF)模型 Fant et al.(1985)中提出的 LF 声源模型，该模型已经得到了广泛的应用，下文将其用于许多示意图中。除了 f_0 外，该模型还有四个参数来控制声门脉冲的形状。



模型由两部分组成，如图 11.4 所示的数学表达式所示。注意，这些表达式根据分化的/微分的声门气流 $U'(t)$ 来定义脉冲形状。然而，综合 LF 模型基本表达式，我们还可以生成与声门气流直接相关的 LF 模型脉冲， $u_8(t)$ 。图 11.4 的下波形为两种分化的声门气流的 LF 模型脉冲的例子，上波形为对应的声门气流脉冲的波形图。

第一部分的模型是一个正弦函数,它的振幅从声门打开的时间点到主活跃区的时间点呈指数增长。三个参数确定这部分的形状(1) $\omega g = 2\pi Fg$, Fg 是正弦函数的频率, (2) α , 决定振幅增加的速率, (3) $E0$, 这是比例项间的比率。

第二段是一个指数函数，用来模拟从主活跃区时间点到声门闭合时间点的气流， t 。声门循环的这部分称为“回流阶段”，决定了当声带闭合时，主活跃区后的剩余气流，即“动态泄漏”。在 LF 模型中，回流阶段的控制参数为 TA。TA 是对回流阶段“有效持续时间”的度量，由 t 时刻切线在时间轴上的投影决定(见图 11.4)。

LF 模型的描述假设 $t_c = t_o$ 。即声门闭合的时间点与声门为即将到来的脉冲周期打开的时间点相同。这意味着模型缺少一个封闭的阶段/闭合段。在实践中,对于比较小的 TA 值,指数曲线将非常接近于零线,在任何程度上都提供了一个封闭的相位。这样可以节省一个参数,而不会显著降低灵活性。此外,为了明确地确定脉冲形状,四个 LF 参数 e_0 , α , ω_g and t_a 按照“面积平衡”的要求组成,即 LF 脉冲的正面积(从 0 到 t)应等于负面积(从 t 到 t_c)。根据真正的声门气流 i_g 。这意味着连续脉冲的基线保持不变。(关于该模型的更详细描述,请参

见 Fant et al. 1985; Fant, 1995;和 Gobl 2003。)

2.2.2 其他声源模型 文献中已经提出了许多其他参数声源模型(如 Rosenberg, 1971; Rothenberg 等人, 1974; 1974 年 9 月, 1970 年 9 月, 1982 年; Ananthapadmanabha, 洛杉矶奥运; Hedelin, 洛杉矶奥运; 1985; 永奎斯特和藤崎, 1985; 价格, 1989; Klatt & Klatt, 1990; Oi 和 Bi, 1994 年 Veldhuis, 1998 年)。根据所建立的声门气流脉冲模型模拟的是真实声门气流脉冲还是分化的声门脉冲可以将这些模型分成两类。它们在参数的数量和用于生成声门脉冲的函数上也有所不同。它们之间的另一个重要区别是, 它们是否包含一个段来模拟声门脉冲的回流阶段。

在 Ananthapadmanabha(1984)和 Ljungqvist 和 Fujisaki(1985)中可以找到这些声源模型的相当详细的比较。图 11.5 总结了七种不同声源模型的一些重要特性。

	Model	Single flow derivative discontinuity	Provision for multiple flow derivative discontinuities	Provision for continuous flow derivative	Waveform realization
Amplitude, width, and skewing of the glottal flow	(a)	yes	(yes)*	no	sinusoidal
	(b)	yes	no	no	sinusoidal
Independent control of flow derivative discontinuity	(c)	yes	no	yes	sinusoidal
Modeling of activity in the glottal closed phase	(d)	yes	no	yes	sin+polyn.
	(e)	yes	no	yes	exp. sin.
	(f)	yes	no	yes	polynomial
	(g)	yes	yes	yes	

* Rosenberg proposed several models, some of which allow multiple discontinuities.

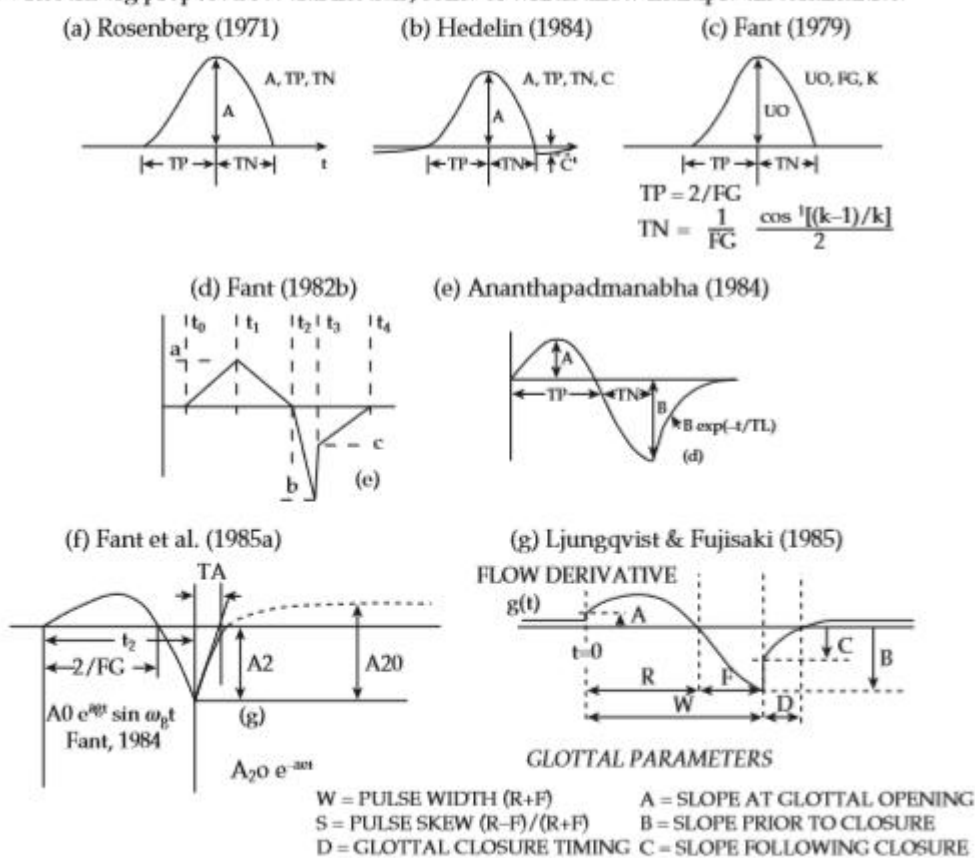


Figure 11.5 Waveforms and equations for seven proposed voice source models. (After Ananthapadmanabha, 1984; and Ljungqvist & Fujisaki, 1985)

图 11.5 总结了 7 种不同声源模型的波形图和方程。(Ananthapadmanabha 之后, 1984; Ljungqvist & Fujisaki, 1985)

2.3 声门信号测量:声源模型匹配

如前所述, 提取声源测量值/参数的方法包括将语音声源模型与逆滤波器输出进行匹配, 并从模拟的波形图中得到测量值。与直接从逆滤波器输出的时间点和振幅点测量参数相比, 该方法具有一定的优点。首先, 模型匹配可以同时考虑时域和频域信息, 因为可以计算出模型的频谱。这对于捕获具有重要频谱意义但难以直接从波形图(例如回流阶段)精确测量的特

征特别有用。进一步的优点是，所建模的声源信号可以快速地实现合成，原则上这应该有助于对所测的各种参数进行感知测试。

与逆滤波一样，模型的匹配可以自动完成(如 Ananthapadmanabha, 1984; Chan and Brookes, 1989; 斯特里克和波夫斯, 1994)，但目前的自动算法并不总是产生可靠的结果。如果采用手工交互式方法如图 11.6 所示，就能得到更精确的测量数据，屏幕显示将在匹配过程中引导用户。

此图的中间面板根据顶部面板的指定脉冲显示了其逆滤波波形(分化的声门的气流)。叠加在这个脉冲上，我们还可以看到匹配的 LF 脉冲(粗线)，其轮廓由四个时间点(垂直线)和一个振幅点(水平线)确定，由实验员手动设置。四个时间点分别为:(1)声门打开时间，到(2)声门气流峰值时间， t ;(3)活跃区时间， t ;(4)估算出来的回流阶段的时间点， t ，($=t$ + 助教)。振幅点(5)为活跃区的振幅， EE 。底部面板为逆滤波脉冲对应的频谱，其上叠加的是 LF 模型脉冲的频谱(粗线)。通过对时间点和振幅点进行微调，对模型脉冲进行优化，从而在时域和频域找到最佳的整体一致性。

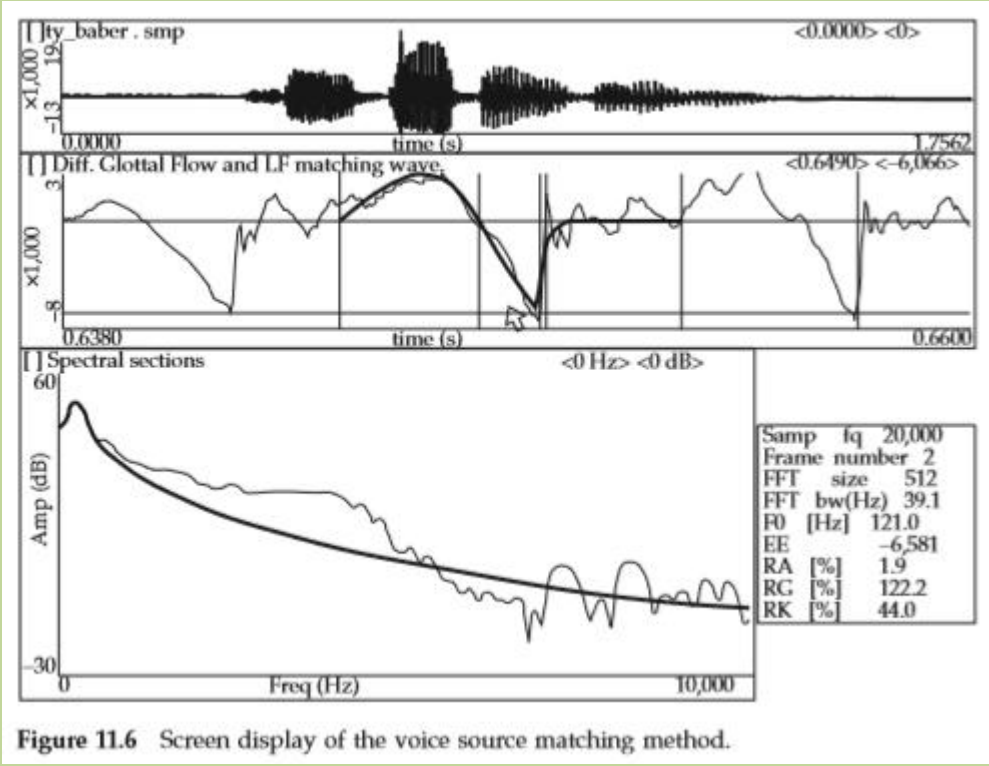


Figure 11.6 Screen display of the voice source matching method.

图 11.6 语音声源匹配方法的屏幕显示。

2.4 一些重要的声源参数

第 2.2 节中概述的 LF 参数决定了声门脉冲的总体形状。在我们的分析中，我们需要测量这个波形的非常具体的方面，即那些被认为在听觉和感知上都很重要的方面，并且这些方面更容易与潜在的生理活动相关联。一旦匹配过程圆满完成，就可以计算出这些声声源参数。我们在这里概述了一些最重要的参数，如图 11.7 所示说明了声门波形的变化如何影响声谱。但是，我们必须记住，我们很难非常精确地说明单个声源参数的语谱结果，因为它们经常以复杂的方式相互作用。我们还应该记住，精确的定义在一定程度上取决于使用的模型。这里我们使用 LF 模型来定义参数，本章后续部分都会使用它来画示意图。

基频， f_0 ——基频= $1/T$ ， T_0 在这里是基本周期，声门周期的时长由两个相邻的声门脉冲之间的主活跃区的时间决定。

活跃强度 EE —— EE 值与声门活跃的综合强度密切相关。它是主活跃区的负振幅，出现在声门气流信号的最大的非连续段的时间点处。该时间点(t)通常与声门脉冲下降信号的最大斜率相重合，即， EE 通常由分化的声门气流中最大的负值给出。在产出水平上，它是由声带的闭合速度和通过它们的气流决定的。在声级上，它与声源信号的总强度密切相关。 EE 的变化基本上会导致所有声源分量的振幅发生相应的变化，除了最低的分量，尤其是第一谐波。因此，该参数与简单脉冲声源的振幅参数最为相似。然而，最低分量的振幅很大程度上是由脉冲形状决定的，因此受活跃强度变化的影响较小。

动态泄漏， RA ——动态泄漏是回流阶段的剩余气流，发生在活跃时间到完全闭合时间之间(如果有直流泄漏，则为最大闭合时间)。就真实的声门气流而言，回流阶段/返回阶段显示为脉冲关闭分支的一个“圆角”。 RA 值是对回流阶段有效时长的衡量， TA ，是对基本周期的归一化， T 。因此， $RA = TA / T_0$ ，就 LF 模型而言， TA 大约是模拟回流阶段的指数函数的时间常数（见图 11.4）。在产出层面上， RA 是指声门闭合段的清晰度，也即，研究声带在整个长度和深度上是瞬间接触还是逐渐/渐进接触。动态泄漏的差异在声学上是很重要的，因为它们影响到声源语谱的斜率。回流阶段指数函数的频率特性近似于一阶低通滤波器的频率特性。截止频率 FA 与 TA 成反比： $FA = 1 / (2\pi TA) = f_0 / (2\pi RA)$ ，i. e.，滤波器截止频率与动态泄漏量成反比。

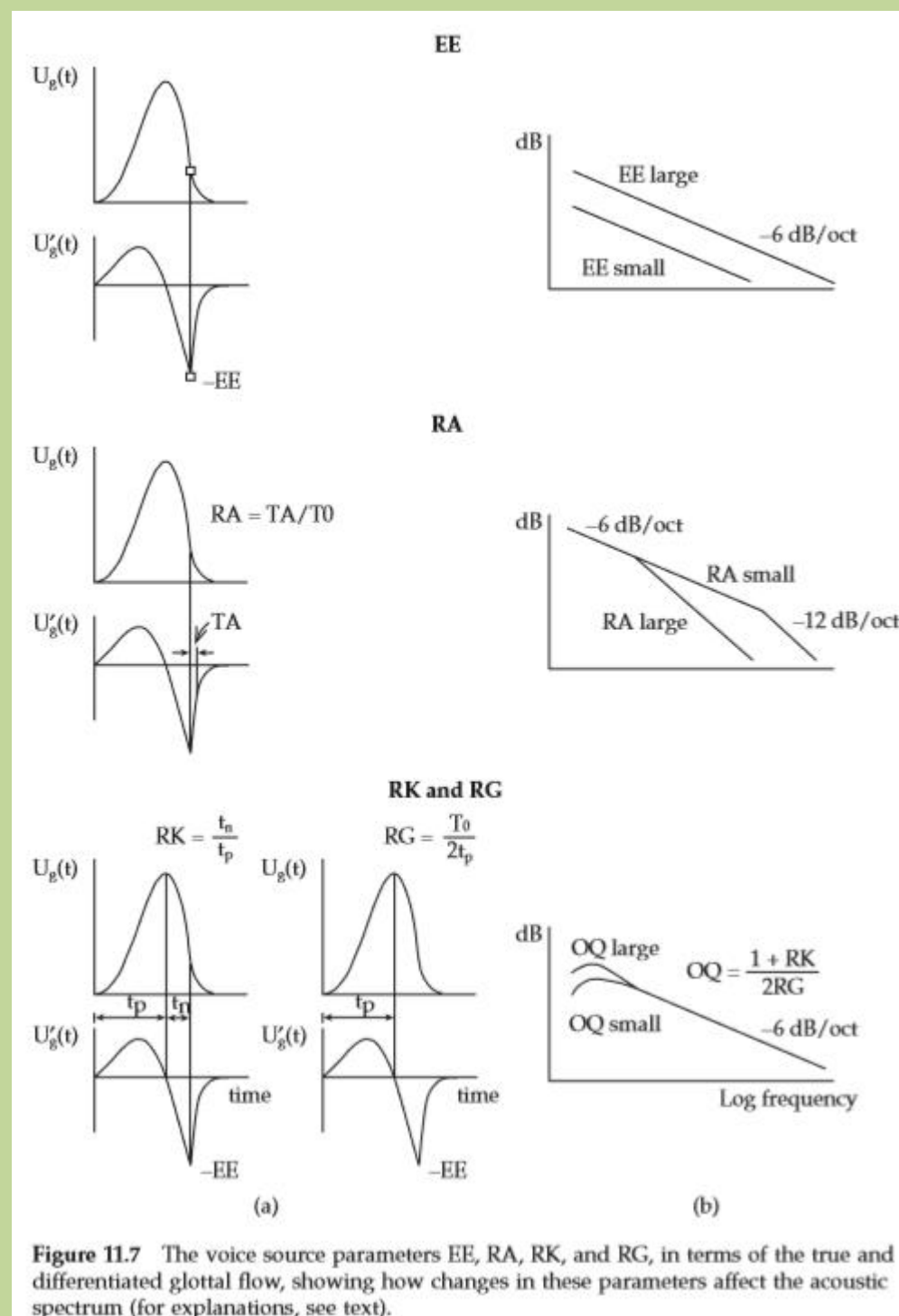


图 11.7 声源参数 EE、RA、RK 和 RG，分别表示真实声门气流和分化声门气流，显示这些参数的变化是如何影响声谱的(详见文本)

开商，OQ——开商是声门循环中声门打开的比例。在声源语谱方面，主要影响较低分量的振幅。尤其是，OQ 值和第一谐波振幅之间有密切的对应关系:但是请注意，对应程度的变化取决于参数 RG 和 RK 的值(见下文)。倾向于与开商共变的一个相关参数就是 UP，也即声门脉冲的峰值体积速度（声门脉冲的峰值体积速度随开商的增大而增大；/一个相关的参数，往往与开商是共变），（分别在图 11.5 中前三个声门模型被标记为 A，A，UO）。

声门的频率、FG——声门的频率是声门的脉冲形状在开放段的特征频率（方特，1979

年), 在 LF 模型中, $FG = \frac{1}{2t}$ (见图 11.4)。FG 可由声门脉冲开口分支的时间周期确定为 $1/(2t)$ 。这个参数的另一个表达式是 RG, 它是归一化为 f_0 的 FG, 所以 $RG = FG/f_0$ 。RG 趋向于与 OQ 和 UP 成反比变化, 因此, 在声源频谱的最低端发现一个衰减水平的高 RG。对于非常高的 RG 值, 声门频率可能接近二次谐波(H2)的频率, 并能够提高其水平。高 RG 和相对较强的 H2 往往是紧音或压音的特征。较低的 RG 值出现在较高的位置, 这有助于提高 H1。因此, RG (OQ 和 UP)对语音输出中 H1 和 H2 的相对振幅有影响, 这是语言学文献中经常使用的一种测量方法(更多信息请参见 2.5 节)。

声门对称性/sketw, RK——与声门面积函数相比, 声门气流脉冲通常向右倾斜, 即, 开放段往往比闭合段更长。声门脉冲的这种倾斜可能是由于声道的惯性负荷造成的(Fant, 1982; Rothenberg, 1983)。脉冲偏斜的声学结果有些复杂。它主要影响声源语谱的较低部分, 使更对称的脉冲形状具有提高较低谐波的作用。然而, 偏斜的程度也决定了声源语谱中缺口(减弱或缺失的谐波)的深度:脉冲越对称, 频谱/语谱下降得越深。缺口的位置由开商和脉冲形状决定(参见弗拉纳根, 1972, pp. 236-42)。对声门脉冲偏斜的效果已给予相当的重视。然而, 有一种情况可能是, 偏倚的感知重要性被相对高估了。偏斜通常与活跃强度高度相关, 其感知贡献容易与活跃强度混淆。如果使用的声源模型缺乏对活跃强度的直接控制, 那么这种混淆的风险尤其高。换句话说, 我们更建议活跃强度是一个重要的参数, 而不是脉冲的倾斜。

送气噪音, AH——发声过程中除了产生准周期性声源外, 还会在声门附近产生不同程度的噪声, 通常称为送气噪音。这种噪音声源的模型往往没有被明确包括在参数化声源模型中, 混合活跃(与预期噪声混合的周期性活跃)的重要性已经被多次提到(例如, Pandit, 1957; Dolansky & Tjernlund, 1968; Fujimura, 1968; Rothenberg, 1974; Ladefoged & Antonanzas Barroso, 1985; Klatt, 1986; Hunt, 1987; Gobl & Ni Chasaide, 1988;)但噪声分量很难定量估计。然而, 即使不是模型的直接组成部分, 也可以使用噪声发生器来提供浊音活跃的噪声分量。例如, Rothenberg(1974)等人讨论了预期噪声的实际频谱内容/情况、控制送气噪音水平的策略以及如何在声门周期内调节噪声等问题。1974 年, 《马岗左公鸡》(1978), 《黑暗与上诉》(1990 年)和 Gobl(2006)。

声源参数从一个脉冲到另一个脉冲的稳定性也是决定音质的一个重要因素。传统上, 抖动和闪烁等测量方法被用来量化脉冲到脉冲(一个个脉冲)的变化。抖动是 f_0 中的随机变化, 闪烁等于脉冲到脉冲振幅的波动。闪烁经常从语音波形图的振幅得到衡量, 由于声源滤波器相互作用的影响, 因此这在一定程度上会产生误差。理想情况下, 闪烁应该直接从声门波形的振幅测量中估计, 例如 EE。高水平的抖动和闪烁经常被发现与嘶哑的声音有关。但是, 请注意, 其他声源参数也可能在某些情况下表现出不稳定性, 这一事实可能也具有感知上的重要性。例如, 请参见图 11.16 中的病态声音和图 11.9 中的正常挤喉声。

Gobl(1988)已经表明, 上述许多声源参数趋于协变/共变。EE 与语音波形的负振幅高度相关, 其他声源参数往往与 EE 相关。例如, 回流阶段/返回阶段通常与 EE 成反比, 因此如果活跃较弱, RA 就较高。RA 和 RK 之间通常也存在协变, 因此长回流阶段/返回阶段(低 EE)对应于更对称的脉冲形状。这些趋势中有几个已被后来的工作证实(Pierrehumbert, 1989; Fant, 1995, 1997), 但并不总是如 Gobl 和 Ni Chasaide(1992)和 Ni Chasaide 和 Gobl (2004a)所述。

随着我们的知识水平的提高, 从几个基本参数预测许多声源参数可能成为可能。例如, Fant (1995, 1997)提出了一种全局脉冲形状参数 Rd, 该参数考虑了一些参数之间的自然协变/共变。从 Rd 值中, 可以使用线性回归分析得到的公式预测 RA 和 RK 等参数的默认值(进一步的细节见 Fant, 1995; Gobl, 2003)。Alku 等人(2002)也提出了类似的参数“归一化振幅商”NAQ。

Rd 和 NAQ 都是基于“直线/排列时间”Td (Fant, 1979a)归一化为基本周期。有人提出(如 Alku 等, 2002), 这些全局参数是与音质差异的松紧维度密切相关的度量。注意, Td 是由声门气流信号及其衍生物($Td=UP/EE$, 见图 11.4)测量的振幅得出的。这种基于振幅的参数可能比基于时间的参数更适于自动处理。声门波形中由特定定时事件定义的参数-例如 OQ、RK 和 RG 对自动声源分析系统中经常出现的逆滤波器估计误差很敏感。在 Gobl 和 Ni Chasaide (2003a)中描述了一组扩展的基于振幅的参数, 其中包括 OQ、RK、FG 和由振幅测量和 f_0 得出的 RG 的近似值。

2.5 与声源相关的频谱测量

在上一节中, 我们主要关注声门声源的时域测量, 并将其与预期的频谱结果联系起来。频域测量也可以在逆滤波器的输出端进行。从上面对声源参数的描述中可能可以清楚地看出, 在试图表征和比较声源语谱时, 可能需要区分非常低的频率和较高的区域。光谱缺口的出现, 甚至是额外的声门下极/零点对, 都可能使图像变得更加复杂。特定的声门脉冲形状(非常对称)也会产生缺口, 例如在气嗓音中就可能发现。此外, 声门被绑架的越多, 与声门下系统的耦合就越大, 声门下共振出现在声源谱中的可能性也就越大。

频谱倾斜可能是在声源频谱中要测量的最基本的参数。Jackson 等人(1985 年, 1986 年)在早期研究中探索了将单一回归线拟合到声源语谱图的可能性, 证明了获得它并不总是一件简单的事情。图 11.12 说明了一种比较声源语谱的可行的方法, 该方法考虑到不同频率区域的电平变化, 并在第 3 节中进行了说明。关于谱倾斜的一个复杂因素涉及准周期和非周期声源的混合活跃。在估计声源频谱的倾斜度时, 我们理想的做法是测量准周期脉冲的形状所引起的倾斜度, 而不受任何其他声源, 例如送气噪音声源的影响。例如, 音质, 如气嗓音和耳语的声音, 通常有一个相对陡峭的频谱斜率(见进一步的第 3 节), 但这种倾斜的增加可以被预期噪声所掩盖。如果送气噪音声源足够强, 可能会引起频谱中高频分量的相对增大, 从而掩盖准周期声源频谱倾斜的增加。

基于语音输出信号的频谱测量也很有用。对于相同的语音项目, 不同的只是音质, 平均语谱(如图 11.11 所示)甚至长期平均光谱可以帮助显示声源差异。常用的一种测量方法是第一谐波(h_1)的振幅水平/level 与某些高频分量的水平进行比较。 h_1 和 F_1 水平的比较已在许多研究中使用(如 Kirk 等, 1984, 1993; ni Chasaide & Gobl, 1993; 及下图 11.10)。另一个常用的测量涉及到前两个谐波水平的比较(见, 例如, Fischer-Jørgensen, 1967; bickley, 1982; Maddieson 和 ladefoged, 1985; 布兰肯希普, 2002)。广泛发现, 占主导地位的 h_1 与气嗓音高度相关, 而相对较强的 h_2 则与紧音或挤喉音相关。

基于语音输出波形的测量对该领域的语言学家特别有吸引力, 因为它们不需要执行逆滤波和模型匹配所需的技术设施。然而, 重要的是要记住, 虽然这些类型的测量反映了声源语谱的差异, 但它们也对其他因素敏感, 因此不能用来推断源谱的实际斜率。重要的是, 实验人员要意识到影响输出频谱不同频率水平的其他因素, 因为必须仔细选择语音材料来考虑这些因素。首先, 共振峰的频率影响它们的振幅水平, 所以比较 h_1 和 F_1 的水平显然不适合不同的元音性质。共振峰层级也部分由共振峰带宽决定, 共振峰带宽反映了阻尼的存在程度。在声门脉冲中很少或没有闭合段的地方, 如在气嗓音中, 可发现高度阻尼。声门上因素也影响阻尼的程度, 从而影响共振峰的带宽。在任何情况下, 共振峰带宽以一种不直接反映声源频谱水平的方式影响输出频谱的水平。

所有这些语谱测量也对 f_0 的差异敏感, 或者更准确地说, 对 f_0 与 F_1 频率比值的任何偏移都敏感。例如, 当 F_1 高和 f_0 低时, h_1 和 h_2 层级水平的比较可能是一个有效的度量。然而, 当 F_1 较低或 f_0 较高时(或者在最坏的情况下, 这两个因素都适用), h_1 或 h_2 的层级水平可能会随着它们接近 F_1 峰值而升高。在这种情况下, h_1 和 h_2 的相对水平会受滤波器

和源因素的影响，因此不再是语音模式的可靠指标。

通过补偿声道的主要影响，可以得到对元音质量差异不太敏感的估计值。例如，Stevens 和 hanson(1995)从语音输出频谱中提出了 $H1^* - H2^*$ 和 $H1^* - A3^*$ 的测量方法。 $H1^*$ 和 $H2^*$ 是 $H1$ 和 $H2$ 振幅的调整量，考虑到了 $F1$ 传递函数对这些谐波振幅的主要影响。 $A3^*$ 是第三共振峰振幅水平 $A3$ 的调整量，其中补偿了 $F1$ 和 $F2$ 变化的影响。换句话说，这些措施包括一个简化的反滤波过程，其中只取消了声道滤波器的主要作用。hanson(1995, 1997)用这些方法分析了女性说话者声门声源特征。最近，Iseli 和 alwan(2004)提出了一种更全面的谐波振幅水平校正公式，以进一步降低声道滤波器对这些措施的影响。

3 一些常见的音质

作为第 4 节的背景，我们在这里简要介绍一些常见的音质。其目的不仅是为了说明这些音质在声学上可能有何不同，而且还为了说明各种有用的测量方法。如 Laver(1980)所述，我们需要处理的音质有正常嗓音、气嗓音、儿语时的声音、挤喉音、紧音和松音。请注意，虽然我们在这里只关注语音质量的喉部方面，但最后提到的两个方面可能涉及整个语音设备/发音器官/发音装置或多或少的紧张程度，而不仅仅是语音系统。

这里的生理描述是关于肌张力的三个假设参数：内收张力、内压和纵向张力(见 1980 年 laver 图 11.8)。这些因素决定了声带的结构和张力设置，并与声门下压力和声门气流相关的气动因素相互作用，产生各种音质。要获得更完整的描述，请参考该文本。也可参见 Catford(1964)和 ladefoged(1971)中对语音质量的描述。内收张力是指杓状肌被拉到一起，从而使软骨声门内收的力。它由杓间肌控制。通过对杓状体发声过程的近似分析，内侧压缩被定义为舌韧带关闭的力。它主要由外侧环状杓状肌控制，但也可累及外侧甲状腺杓状肌。纵向张力是指声带的张力，主要通过声带和环甲肌的收缩来调节，环甲肌的主要功能是控制音高。

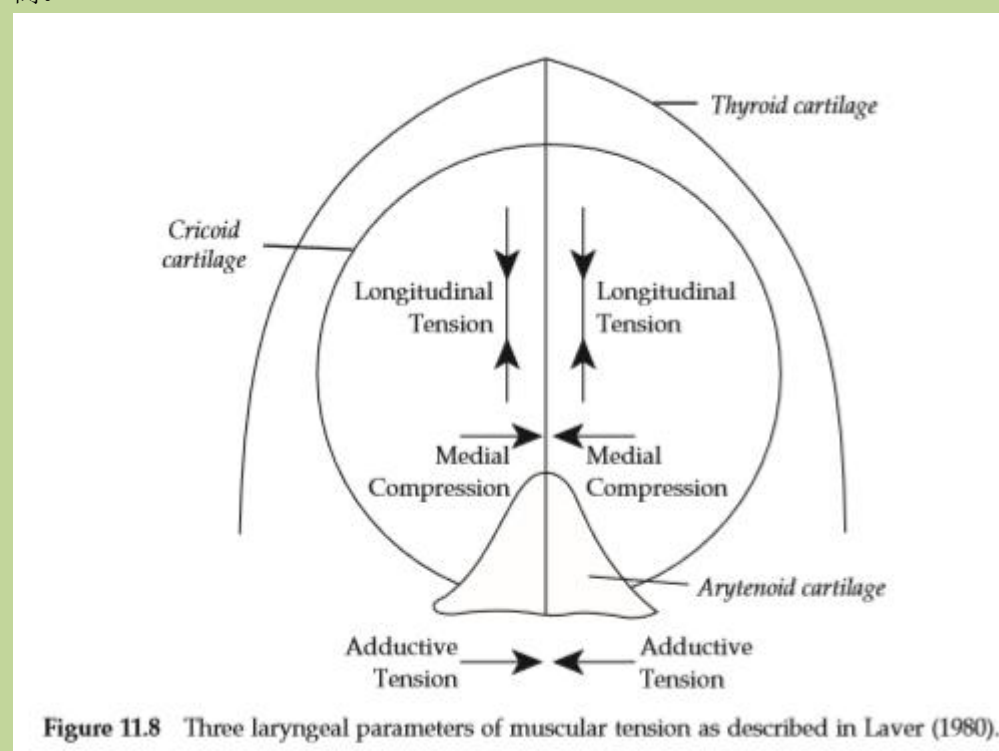


图 11.8 Laver(1980)中描述的喉部肌张力的三个参数。

图 11.9-12 显示了这些音质的一些声学特性，并使用第 2 节中概述的分析技术推导出来。演讲材料由一位熟悉 Laver 系统的男性语音学家制作。图 11.9 显示了声源参数值；图 11.10

和 11.11 显示了相对不重读单词 *strikes* 的语音间隔的语音输出信号的频谱测量。对于其中四个性质，声源语谱的示意图如图 11.12 所示，测量在无意义单词 *babber* 中重音/a/的中点。这里的目的是通过显示每种音质的斜率偏离“理想”声源的程度来促进频谱斜率的比较（即，-6db/每个倍频程用于分化的声门气流）。为了达到这一目的，语谱图被“平面化”，每个倍频程增加 6 分贝相对于振幅层级水平的第一谐波(L0)。将声源语谱分为 0-1、1-2、2-3、3-4 khz 四个频段。每个带中有一个共振峰。4khz 以上的谐波并未测得。然后计算一个频带内所有谐波的归一化(线性)振幅的平均值，并相对于 L0 作图。这个值表示与“理想”声源斜率的偏差，由图 11.12 中 0 db 处的水平线表示。更多细节见 Gobl(1989)和 Gobl and ni Chasaide(1992)。

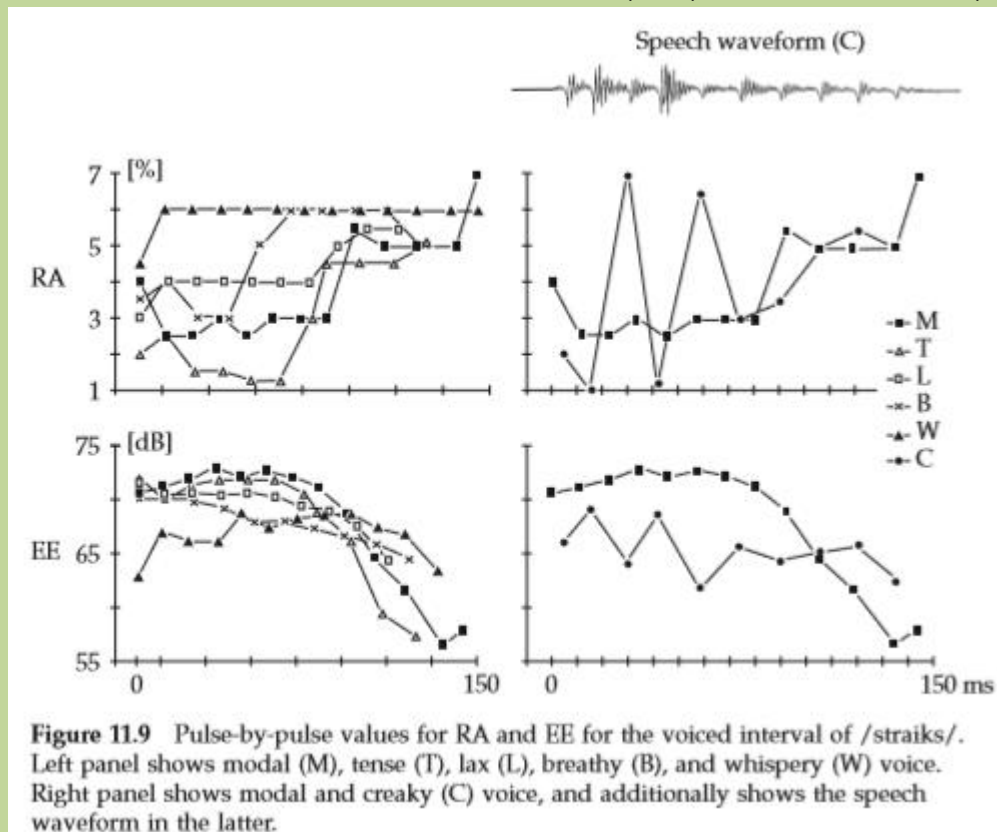


Figure 11.9 Pulse-by-pulse values for RA and EE for the voiced interval of /striaks/. Left panel shows modal (M), tense (T), lax (L), breathy (B), and whispery (W) voice. Right panel shows modal and creaky (C) voice, and additionally shows the speech waveform in the latter.

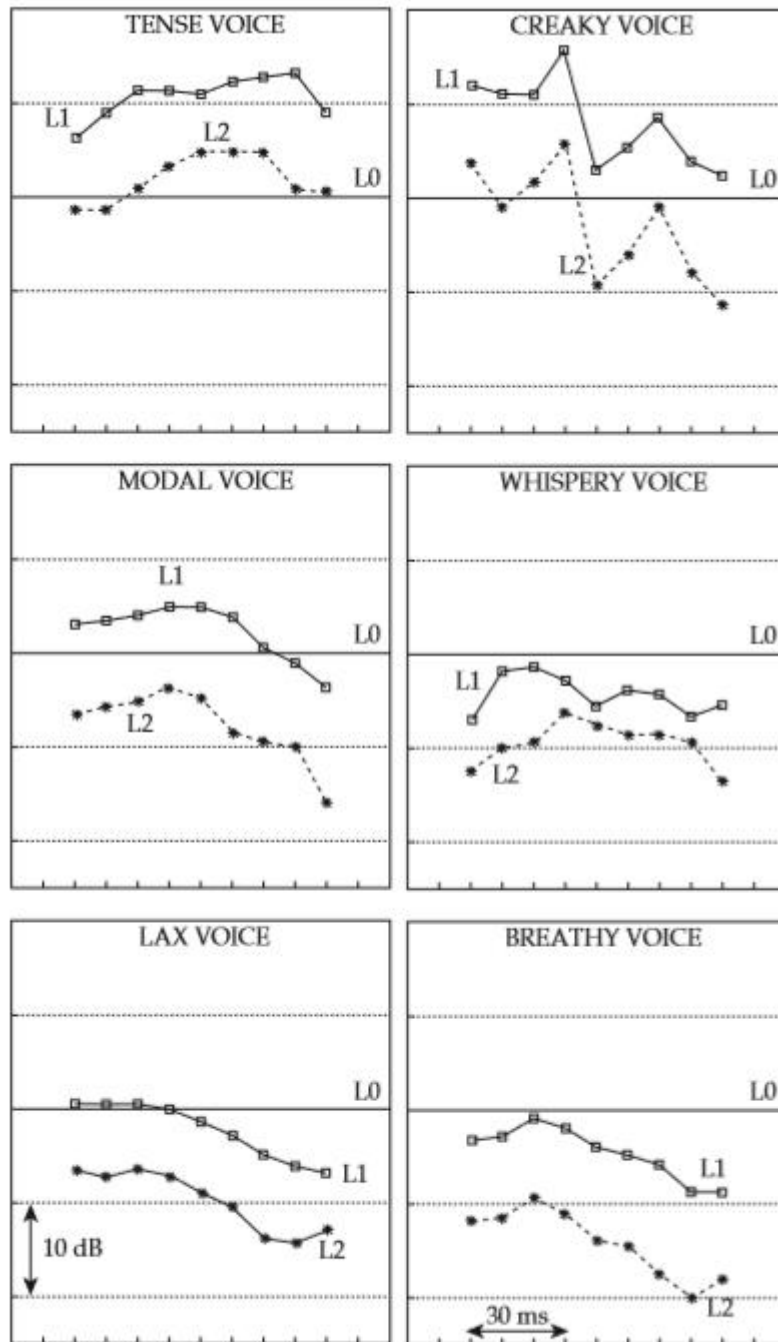


Figure 11.10 F1 and F2 levels relative to the level of H1 (L_1-L_0 and L_2-L_0) are shown for tense, modal, lax, breathy, whispery, and creaky voice qualities for a 90 ms interval in /straiks/.

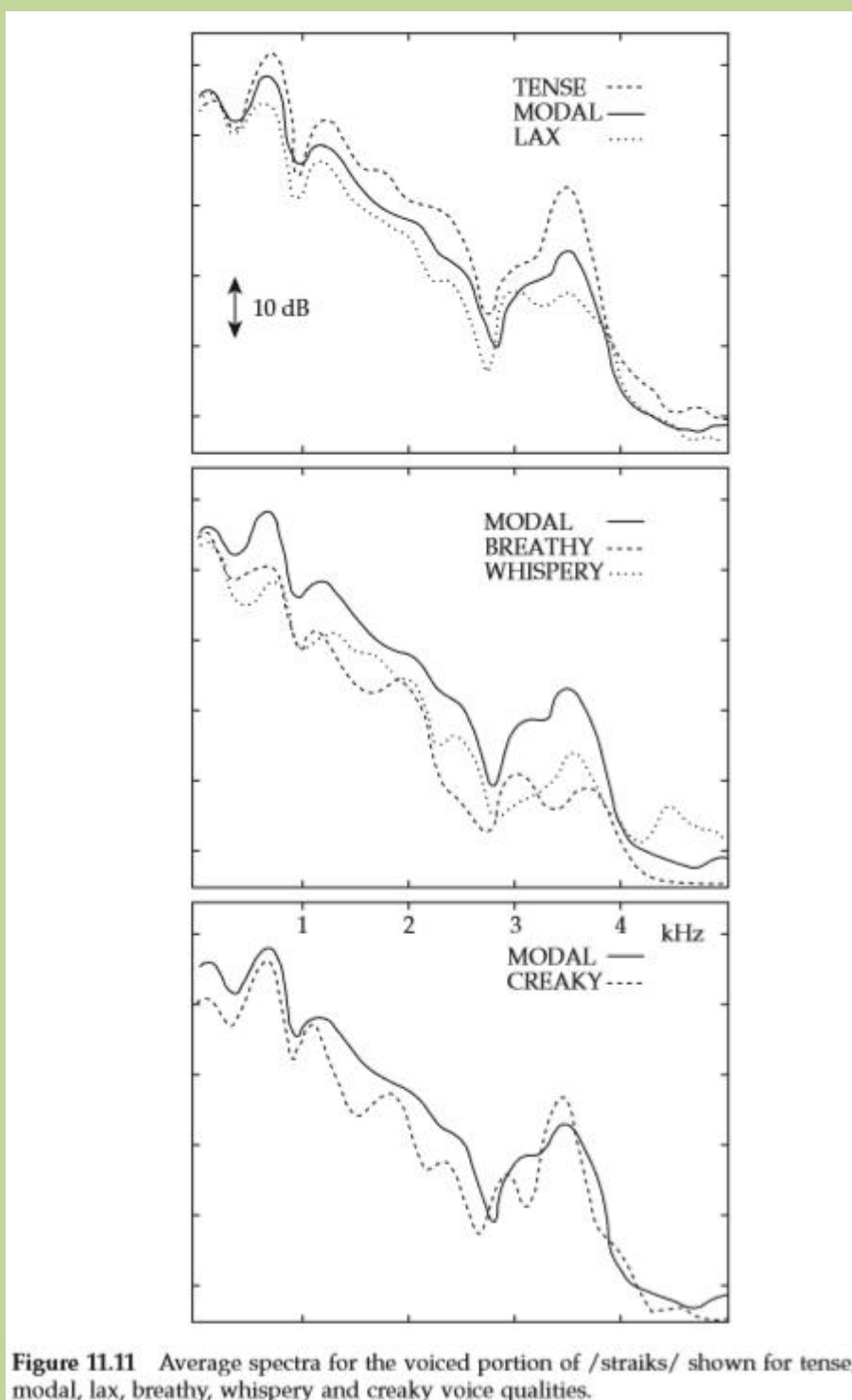
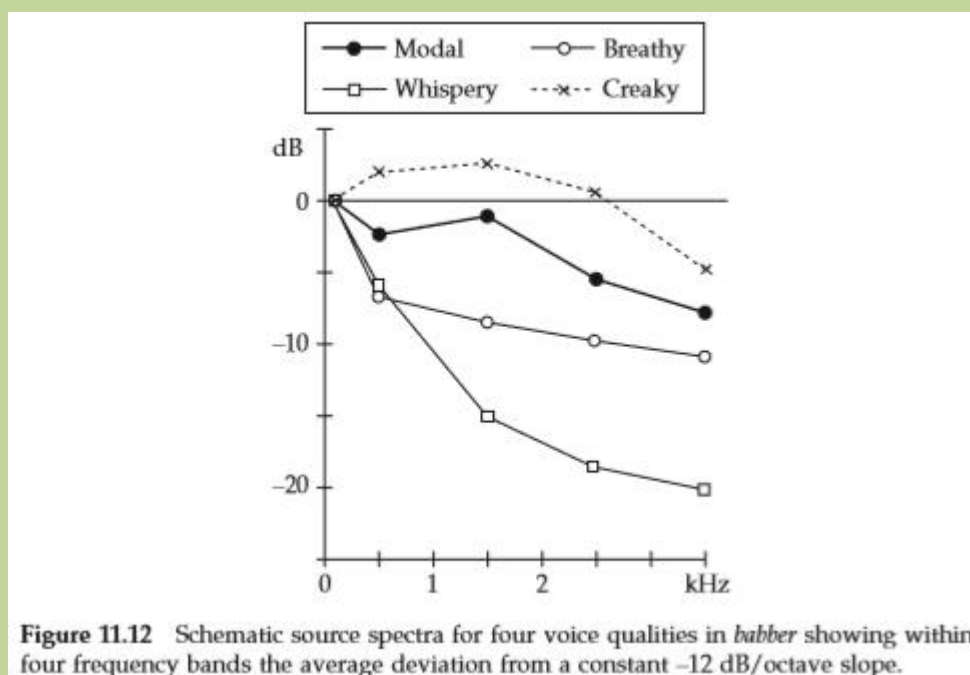


Figure 11.11 Average spectra for the voiced portion of /straiks/ shown for tense, modal, lax, breathy, whispery and creaky voice qualities.

与其他音质相比较，正常嗓音是一种中性的发声方式，“语音理论认为，正常嗓音是在没有明确改变或添加特定特征的情况下，在普通语音中发生” (laver, 1980, p. 95)。对于这种性质，内收张力、内侧压缩和纵向张力被认为是适度的。声门的韧带和软骨是作为一个整体在振动的。laver 进一步将声带振动描述为有规律的、周期性的、有效的，声门完全闭合，因此没有可听到的预期噪声。然而，最近的一些研究表明，即使在被认为是正常嗓音的情况下(例如，见 Sodersten, 1994)，声门闭合不完全也可能非常普遍，尤其是在女性语音中。

图 11.12 中正常嗓音的声源语谱斜率略大于“理想”情况。然而，相对而言，它是一种相当有效的发音方式。重要的是要记住，在使用正常嗓音或任何发音方式说话的话语中，可能

有相当大的声源的动态变化(例如, Gobl, 1988)。在某些环境中, 正常嗓音和气嗓音/耳语声可能会有相当大的收敛, 如图 11.9 中不振动的/清辅音/k/相对靠前的几个周期就可以看到这一点。这是一种上下文效应, 它似乎会影响所观察到的所有音质, 并在 Ni Chasaide 和 Gobl(1993)中进行了详细讨论。



气嗓音被认为包含最小的内收张力, 较弱的内压和较低的纵向张力。声带振动的效率很低, 它们从来没有完全贴合在一起。因此, 有相当大的持续声门泄漏与一些可听到的吸气/预期噪声。这种语音质量的高动态泄漏表现为高 RA 值。因此, FA 比正常嗓音要低得多, 尤其是在 *babber* 这个词的重音元音中, 我们发现气嗓音的值为 500 Hz, 而正常嗓音的值为 1500 Hz。声门脉冲也更对称(高 RK)的气嗓音, 并有一个高开商, OQ。总的来说, 这表明气流通通过声带的速度很高, 这是我们对音质的预期, 这确实是我们计算出来的(声门峰值气流流量)值所显示的。这将产生相对增强的最低谐波, 这种音质中的这一特征已经被广泛报道。声源语谱下端随之产生的陡坡在图 11.12 中清晰可见。在语音输出信号中, 我们从图 11.10 中可以看出, 气嗓音中 h1 的优势明显/占据主导地位, 而正常嗓音中 F1 的优势明显/占据主导地位。这种特殊的语谱测量不仅捕捉到 L0 的相对增强, 而且还捕获了 F1 的高阻尼, 这是预期的气嗓音的音质, 发此音时声带被绑架/翘起。这些影响也可以在图 11.11 的平均语谱图中观察到。

儿语发声的特点是低内收张力, 中等到高的内压, 中等的纵向张力。因此, 有一个三角形的软骨声门开口, 其大小随着内侧压迫的程度而变化。在弱耳语中, 内侧受压适中, 声门开口可包括部分舌韧带和软骨。随着声门强度的增高, 声门内侧受压程度增高, 声门开口变小, 直至只打开软骨声门。喉部振动被认为局限于内收的舌结声门的那一部分, 以及杓状体之间三角形开口的须状部分。这是非常低效的, 有相当程度的可听到的吸气/预期噪声。

laver (1980, pp. 133-4)指出, 儿语发声和气嗓音形成了一个听觉连续体, 两者之间没有明显的分界线。就听觉而言, 它们的区别在于周期和噪音成分的相对优势: 在气嗓音中, 周期成分占主导地位, 而在耳语声中, 噪音成分相对较大(参见第 4 节对这两种声音质量的进一步评论)。

耳语声的声源测量值与气嗓音的声源测量值非常相似, 在某些情况下, 它们与模态值的偏差更大。耳语声与气嗓音的主要区别在于儿语声的 RK 和 OQ 较低, 声门血流脉冲更为倾

斜，闭合分支相对较短。计算出的 UP 值也明显低于气嗓音。由于 UP 与 h1 (L0)水平高度相关，我们发现基本音的增强比气嗓音的增强要小。注意，图 11.12 的声源语谱中没有显示这种差异，因为这些已经归一化为 L0。然而，就整个声源频谱而言，基本成分仍然占主导地位，因为声源频谱的斜率甚至比气嗓音更大。可能由于这个原因，在输出信号中 L1 和 L2 相对于 L0 的测量值(图 11.10)与气嗓音的测量值看起来并没有太大的不同。当然，带宽差异也会影响输出频谱的层级/水平/level。

挤喉音 被认为涉及高内收张力和内侧压缩，但很少纵向张力。音高被观察到是非常低的，而且似乎是由空气动力因素控制的，而不是像其他音质那样通过改变纵向张力。连续声门脉冲的 f0 和振幅是非常不规则的。由于高内收张力，只有声带的韧带部分在振动。这些皱褶相对较厚且被压缩，心室皱褶也可能有些内收，因此它们的下表面与真实声带的上表面接触。这将创造一个更厚的振动结构/这将导致振动结构更厚。平均气流速度被观测到是非常低的。

虽然每个声音的音质在说话的过程中都是动态变化的，但是挤喉音是特别多变的。在连续的声门脉冲不规则的层面上，挤喉音间歇性地出现。它没有出现在重读单词 babber 中，而是出现在相对不重读的单词 strikes 中(参见图 11.9 中的右侧面板，其中还显示了挤喉音的语音波形)。这里，有两种不同类型的声门脉冲交替出现。一个有一个相当高的 EE 值-这仍然低于正常嗓音或松音，和一个非常低的 RA 值，表示一个相当瞬时的声门关闭和强高频成分。另一种脉冲则表现出相反的趋势:EE 值非常低而 RA 值很高，因此这些脉冲的频谱会非常不同。两种类型的脉冲都具有低 OQ、低 RK 和相对较高的 RG 的特征，第一类脉冲的值更极端。在重音词 babber 中，声源值与上面描述的强脉冲没有什么不同，但主要不同之处在于 EE 值要高得多。这种音质的短开相通常与已知的低气流速率密切相关，而低气流速率是在挤喉音中观测到的，其结果是降低了相对于频谱其余部分的那个较低谐波的水平。这种效应可以在图 11.12 的声源语谱示意图和图 11.11 的语音输出信号的平均谱中清楚地看到。类似地，在图 11.10 中，我们发现相对于 h1 级别/层级水平，F1 非常占优势。请注意，在该图的频谱计算中使用的 30ms 汉明窗在某种程度上使脉冲到脉冲的变化平滑。声门脉冲相对较长的闭合段也应有助于窄共振峰的带宽，这可能也可以从图 11.11 平均光谱中的尖峰推断出来。相对较高的 RG 观察到这种声音质量也将倾向于提高 h2 相对于 h1 的区域，这一特征作为挤喉音的一个特点已在文献中指出。

与中性/正常嗓音的环境相比，紧音涉及到整个声道的高度紧张。在喉部，有两个参数显示张力特别增加，即内收张力和内压。这与许多作者使用的术语“压音“pressed phonation”相对应。与紧音相关的肌肉紧张也可能影响呼吸系统，引起声门下压力增加，以及导致更有力的发音的上咽部。从声学上看，这种音质——这里只分析了 strikes——显示出一个非常低的 RA，表示一个尖锐的，完全关闭的声带。相关的频率测量 FA 值高于在此背景下观察到的所有其他音质，因此可以预期高频成分会相对增强。声门脉冲也更倾向于紧张的声音(最低的 RK 值)，有一个小的开商(OQ)和一个高的 RG。一个相对倾斜的脉冲，具有一个几乎是瞬间闭合和长时间的闭合段，这与高内收张力和内压的物理逻辑描述非常吻合。这些声源值表明，较低的谐波应相对于较高的频率衰减。这种效应在图 11.11 中清晰地显示出来，我们在图 11.10 中观察到 F1 在光谱中占主导地位。与挤喉音一样，高 RG 会相对提高 h1 和 h2 的比例，从而影响 h1 和 h2 的比例。对于这个说话者来说，如果忽略了后者有时会出现的从一个脉冲到另一个脉冲的变化，那么在紧音和挤喉音的参数之间会有相当大的相似性。

松音在整个声道中包含较少程度的紧张，通常与紧音具有相反的特征。在喉部，内收张力和内压降低。因此，发声可能与气嗓音相似，听起来比正常嗓音更柔和、音调更低；然而，这些张力参数的变化量往往小于气嗓音。声源测量显示，其参数值类似于气嗓音。除了极低的 RG 值，松音和气嗓音之间的差异表明，在这两种声音中，松音更接近正常嗓音。

对于所有的音质,我们应该记住它们不是固定的。非正常嗓音音质在一定程度上可能发生,也可能不发生,也可能离正常嗓音更远或更近。音质也可以是一种复合类型,例如儿语型挤喉音(有关这方面的更多信息,请参见 *laver, 1980, 1994*)。

就像刚才提到的正常嗓音一样,即使在被认为是单一音质的话语中,也可能存在相当大的动态变化。跨音质更是如此:非正常嗓音音质可能在特定语境中更接近正常嗓音。例如,我们的数据表明重读音节比相对不重读音节的差异更大。另一个例子是挤喉音,挤喉似乎与特定的环境有关。只通过一组转换就实现合成中的音质改变似乎不太可能,我们还需要借助上下文相关规则。

4 声源变异的决定因素

不同的人所使用的发声类型可能存在个体差异。甚至在个人的讲话中,语音也有相当大的动态变化。声源变异的某些方面是在说话者的控制范围内的,可能与话语的语言内容有关,也可能与说话者发出关于情绪和态度的副语言信息的意图有关。由于社会、地区或语言群体可能倾向于频繁使用特定的发音类型,因此一些声源差异在社会语言学方面发挥着作用。但是,除了语言、副语言和社会语言的影响之外,个人的声音还受到许多因素的影响,其中一些因素不在他们的控制范围内,比如他们发声器官的物理特性。本节概述了声源变化的一些决定因素。

4.1 语言学决定了声源的变异

声源的变化可能与语码的音段性或超音段性有关。其中一些变化可以作为听觉发声类型的变化而被听到,这些变化在语言系统中具有对立/区分功能。其他的变化则不一定能作为发声类型的明显变化而被听到,但是它们以多种方式对系统产生影响。

在对比使用时,最常提到的发声类型是正常嗓音、挤喉音和气嗓音。对于后者,*laver(1980)*认为,在他的分类中,可能涉及的是耳语发音而不是气嗓音。然而,由于气嗓音片段/音段的实现有相当大的变异性,它们很可能位于从气嗓音到耳语发音的连续音域上的不同位置(见下文和第三节),在这里它们将被简单地称为气嗓音。

“松紧”也被用来描述基于发声类型的对比,但是从 *Maddieson* 和 *ladefoged(1985)*中可以清楚地看出,这些术语可能具有误导性,而且很可能用于音系而不是语音准确意义上。因此,作者推测,紧音可能意味着在一种正常嗓音的发音中——如佤语,中国西南的 *Mon-Khmer* 语言,它与一个松音形成对立,但后者可能在语音学上属于气嗓音——但挤喉音在另一种语言,比如彝语,也是在中国西南,与松音形成对立,但后者实际上是正常嗓音。

其他更极端的音质也出现了,*Rose (1989)* 考察吴语的镇海话描述有一种尖锐的“growl/咆哮”。这种音质似乎包括脑室皱褶和会厌化,对说英语的人来说,听起来像是一种病态的声音。从文献中描述,这个“咆哮”似乎像非常极端的音质称为“strident/尖锐”,这是发现在 *Xoo* 语的元音上的,是在非洲南部的布须曼人说的克瓦桑语语言,涉及杓会厌皱褶的变窄,以及咽部的收缩和会厌的背侧 (*拉迪福吉德,1982;公司,1985;ladefoged & Maddieson, 1996*)。

4.1.1 音段层面/层级

元音或辅音音质的对比使用在东南亚、南非的印度-雅利安语和美洲土著语言中相当普遍,这些都是许多研究的重点,其中许多研究在加州大学洛杉矶分校进行。虽然元音和辅音在任何一种给定的语言中都可能使用音质对比,但 *ladefoged(1982)*指出,在一个音节的多个位置发现对比是非常罕见的。“register”通常用于描述音质的对立,但它是一个音系层面的术语,包含许多其他语音特征,例如元音音质、元音时长,以及在特定语言中经常与这种对比相关的微小或夸大的 *f0* 差异。*Gordon* 和 *ladefoged(2001)*阐述了这些其他特性,并讨论了它们如何有助于增强音质对比。

作为现有设施的实际结果，特别是在实地工作中，对语言对比声源效果/效应的大多数研究往往集中于基于语音波形的语谱图测量，如第 2.5 节所述。然而，由于上述原因，在共振峰频率或 f_0 同时存在差异的情况下，使用这些测量来表征基本音质差异可能会出现。元音通常是一种双向的对比，例如，Gujerati 语元音中气嗓音的与正常嗓音的对立。一个更不寻常的例子是对!Xoo 语中六向对立的描述 (ladefoged, 1982; 公司, 1985 年)。这种语言区分正常嗓音的元音和气嗓音元音。每一种特质都可以伴随着额外的挤喉音出现，挤喉的正常嗓音和挤喉的气嗓音。(我们应该注意的是，在 laver 的系统中，后者被称为“whispercreaky voice”，气嗓音和挤喉音的组合被认为是不可能的组合。)最后，正常嗓音和气嗓音的浊音都可以出现，并具有额外的极端刺耳的性质，如上所述。以辅音为例，已经提到了塞音、鼻音、边流音和近音/通音的音质对比。据报道，鼻音的正常嗓音和气嗓音的对比出现在聪加语和其他班图语的莫坎比克语和南非语中 (traill & Jackson, 1987)。许多印度雅利安语 (包括尼泊尔语、古杰拉蒂语和印地语) 都有带气嗓音的停顿，它们可能与正常嗓音、不送气清塞音和送气清塞音形成对立。有关印地语对比的描述，请参见 dixit (1987, 1989)。

注意，当辅音被描述为具有相反的音质时，声学表现往往主要出现在元音的起始点或终止点。虽然这些情况下的声学效应归因于与辅音相关的喉部差异，但它会影响到元音的起始或结尾部分。dixit (1987) 极力指出，虽然发送气浊塞音时声门外展发生在闭合的半程，但塞音间隔本身不应被认为与正常闭合间隔不同。因此，气嗓音，虽然取决于辅音，但其主要体现在紧随其后的元音上。与此类似，traill 和 Jackson (1987) 对松加的气嗓音型的浊鼻音的研究表明，声学效应主要与元音的出现有关，而气浊鼻的声带外展始于鼻音。在交替发声的情况下，有一种很强的跨语言倾向，即在辅音的前段和前元音的后段都有挤喉音。Gordon 和 ladefoged (2001) 综述了这一研究结果和其他关于挤喉音的跨语言分布的发现，他们推测，他们的动机可能是需要保存 CV 衔接的在感知上的重要位置和方式线索。

值得注意的是，即使元音使用气嗓音和挤喉音的对立，这种气嗓音或挤喉音也常常局限于元音的一部分，而不是贯穿始终。至于辅音，有人认为，这种倾向也可能反映出保持特定音位对比的其他听觉上的重要层面 (见 Steriade, 1999; Gordon & ladefoged, 2001; 西尔弗曼 2003 年)。

鉴于这些实现趋势，并不奇怪的是，特定音质对立的音域/语音范围并不总是清晰的，而且可能有理由更倾向于将特定的对比视为关于元音、辅音或音节的对比。在墨西哥 Mazatec 的 Jalapa de diaz 方言中，我们发现了正常嗓音、气嗓音和挤喉音的对比。柯克等人 (1984) 提出了两种可能的分析。一种是把语言看成音节层次上的三方面对比。或者，它可以被看作是正常嗓音和气嗓音型的元音之间的对立，或者是正常嗓音和挤喉音型的浊辅音之间的对立。关于吴语音质对立领域的讨论，见《建芬与马迪森》(1989)。

正如在第 3 节中指出的，音质之间的差异是一个标量/范围，而不是一个离散的方式。对于任何一种特定的音质，例如气嗓音，它在不同的语言甚至是同一种语言/方言的不同使用者中都会有不同程度的出现。Maddieson 和 ladefoged (1985) 指出，印度语中气嗓音的发声比例高于彝语，彝语高于聪加语。Wayland 和 Jongman (2003) 认为，在高棉语昌塔布里方言中，元音对立位于松紧连续体上的位置存在着男女差异。关于跨音位变异，ladefoged 假设，虽然说同一种方言的人在使用一个带气嗓音的音段时，他们使用的气嗓音程度可能不同，但是所有的人在连续体中使用不同程度的气嗓音来产生对比。因此，对于本质上带有气嗓音的说话者，通过在相关的地方增加气嗓音，可以实现正常嗓音与气嗓音的对比。因此，试图从说明语言对比的数据中找出可以对语音质量进行分类的绝对的方法，往往会遇到困难，这并不令人惊讶。

到目前为止，只讨论了音质被认为具有对立功能的情况，即，被认为是音系对比的主要语音特征。差异也可能发生在不会被认为是语音上的不同的情况中。不同发音类型的辅音类

似于具有本质上不同的声源特征。关于浊塞音、擦音、鼻音和边音之间的区别,请参见 Gobl 和 ni Chasaide (1999b)。这种类型的变异很可能反映了声门上的结构,并由此产生了与不同类别的音段相关的空气动力条件,因此,人们期望它是通用的。出于同样的原因,不同种类的元音在其声源参数上也表现出相当微妙的差异(ni Chasaide et al., 1994)。

清辅音比浊辅音对声源的干扰更大,虽然这是声道除音动作的必然结果。虽然这些微扰很大,但通常持续时间很短,本质上类似于描述得很好的 f_0 微扰,它与清辅音有关。在某些情况下,在时间上也会发生更广泛的影响,因此元音的很大一部分的音质可能会发生很大的变化,以反映相邻塞音的清浊性质。语言——甚至一种语言的方言——在这方面可以有很大不同,正如 ni Chasaide 和 Gobl(1993)中瑞典语、英语、法语和德语的数据所表明的那样。该研究的数据表明,喉部控制的潜在差异涉及声门外展的精确时间,以及(在某些情况下)喉部紧张度/张力设置(见 Gobl & ni Chasaide, 1999b 中的讨论)。

虽然这些非对立的声源差异对研究声音系统对立单位的语言学家来说似乎相对不重要,但我们确实需要对它们有更多的了解,因为它们为描述性分析提供了基础材料,并对言语的产生提供了重要的预见。此外,这种次对立特征经常导致声音变化。例如,清浊辅音在元音系统中是一个重要的语域对比的历史来源。例如,参见威兰和琼曼(2003)对高棉语发展演变的讨论。

4.1.2 超音段层面

超音段现象,如语调、音调和重音,已经得到了广泛的研究,尽管主要是根据 f_0 (和较小程度的音强)的变化。这些相对容易理解的方面被明确地排除在本文之外。然而,值得注意的是,当 f_0 和音强变化时,声门脉冲的许多其他方面也可能同时变化。但是声源的变化在超语言系统中所起的作用远远超出了那些严格依赖于 f_0 和强度的方面。

声调——在有声调的语言中,发现特定的音质与特定的声调相关联是很正常的。Green Mong(一种苗族方言,在东南亚使用)的 7 个音调中,有 3 个具有相当相似的 f_0 轮廓,并且似乎主要是根据发声类型来区分的(andrucki & ratliff, 2000)。在普通话声调中,挤喉音往往与第三个低音调和第四个降调联系在一起(戴维森, 1991;belot - Grenie & Grenie, 2004)。建芬和马迪森(1989)所描述的汉语吴语的阴调和阳调也具有特定的音质特征。阳调与阴调的不同之处在于它们使用了气嗓音,并以较低的 f_0 开始。另见罗丝(1989)对吴语镇海话变调中阴调和阳调的不同实现。

由于在任何情况下, f_0 差异往往与不同的语音质量相关联,因此发现声调对立的语域相关关系并不令人惊讶,反之亦然。尽管存在这样的相关性,但许多作者都煞费苦心地指出, f_0 和音质/发声类型/发声态是独立可控的,其中一个的变化无法预测另一个的变化。这一点在 Michaud 和 Mazaudon(2006)发表的关于塔芒语、纳西语和越南语的跨语言数据中得到了说明。

音质和 f_0 之间的联系可能具有历史渊源,Maddieson 和 hess(1987)、Jianfen 和 Maddieson(1989)、rose(1989)、abramson 等人(2004)和 Michaud 和 Mazaudon(2006)讨论了声调对立可能是从早期的音质/发声类型对立演变而来的。

正如可以预料的那样,在某些情况下,某些特定语言的对比可能会受到竞争分析的影响,主要涉及到语域或声调。例如,见关于所谓 Mon(李, 1983;diffloth 1985;thongkum, 1987)。Maddieson 和 hess(1987)认为傈僳语的六个声调应解释为四向声调对立,其中两个声调有一个语域对比。rose(1989)主张将吴语镇海方言声调的阴调和阳调的差异作为一种语域对立来解释,它与三种声调对立相互作用。显然, f_0 和音质特征是音源的两个方面,它们可以单独工作,也可以以互补的方式形成声调对立。

语调——尽管人们通常认为语调是由音高动态决定的,但我们在 ni Chasaide 和 Gobl(2004a, 2004b)中已经讨论过,除了 f_0 之外,声源参数的变化也有重要影响。图 11.13 显示

了 f_0 的动态变化,EE 和 RK 这两个参数来看下面的瑞典话 “Inte i detta århundrade” [Inti xdet:a o:rhøndradə]”。通常被认为是在说话过程中 f_0 的逐渐下降的“偏斜”,在声门脉冲激发(EE)强度下降和声门脉冲形状(RK)越来越对称中也有明显的体现。语言学家倾向于用短语和边界声调(以及时间特征)来描述短语边界,然而声源特征也很重要。在图 11.13 中,RK 在短语末尾的急剧上升,表示有气嗓音偏移,是短语边界的典型标记。在一些语言中,如瑞典语和某些英语变体中,陈述句的短语边界也常常与话语末尾的挤喉音联系在一起(参见 Fant & Kruckenberg, 1989;爱普斯坦,2003;ni Chasaide & Gobl, 2004a)。

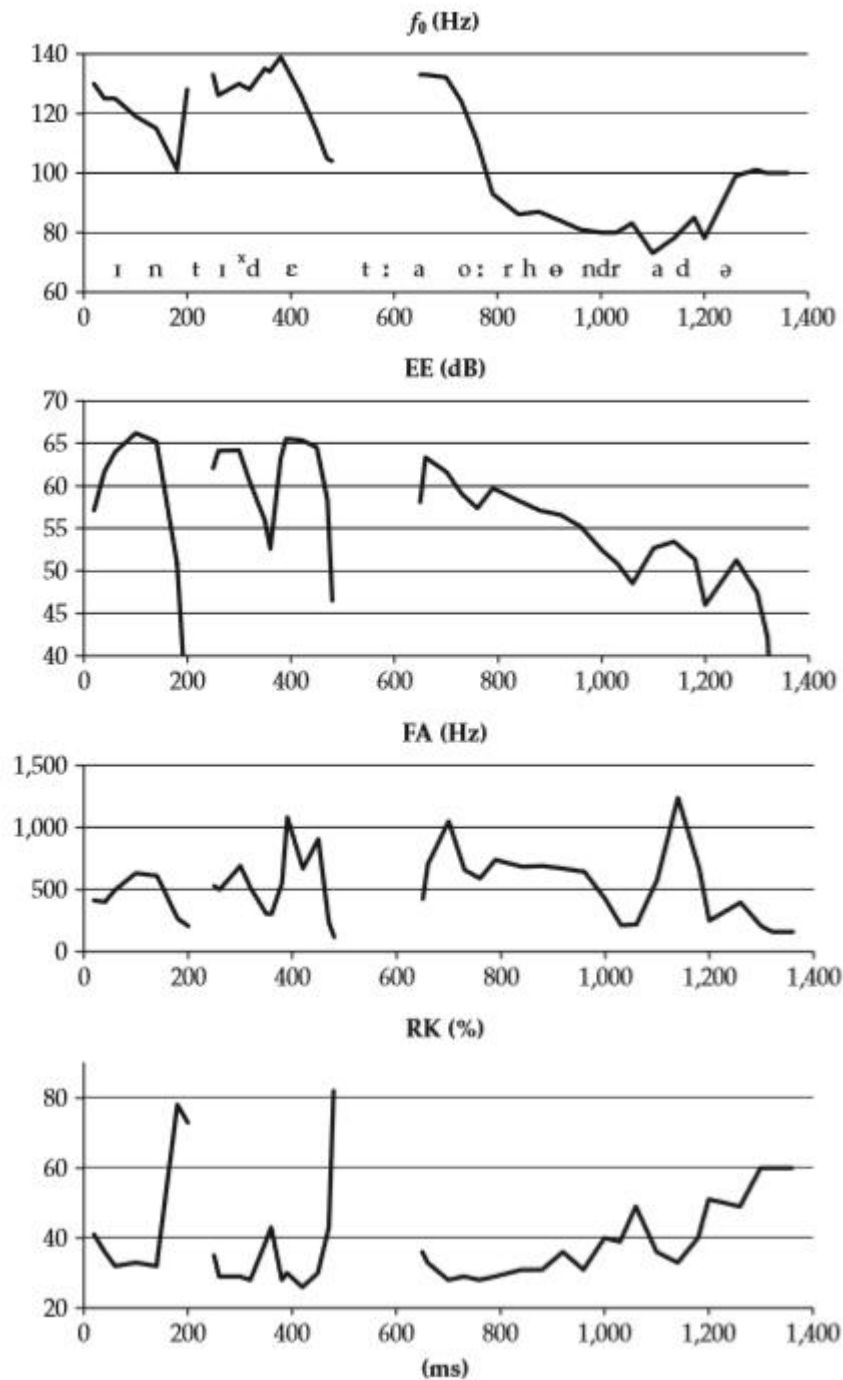


Figure 11.13 Voice source dynamics in the Swedish utterance “Inte i detta århundrade” [Inti xdet:a o:rhøndradə].

声源特征也有助于——重音——音节的突出。图 11.14 中的片段说明了瑞典语语调短语“stor sal”[stu:r sA:l]中最后两个单词在 f_0 、EE、FA 和 RG 中的变化，其中核心重音在“sal”上。“请注意，重读音节[sA:l]的突出程度不仅与元音前期的突出低音有关，而且显然与发音的紧张程度有关，我们可以从声源参数的组合中推断出来；高激发强度(EE)，高声源成分的相对增强(高 FA)，声门频率的增加(高 RG)。值得注意的是，注意，高次谐波的增强与重音的相关性已经在其他一些研究中被提出(如，Sluijter & van heuven, 1996; Campbell & beckman, 1997; Fant 等, 2002; 爱普斯坦, 2003)。

Gobl(1988)对焦点重音作了进一步的观察，其中分析了一个单词在话语的焦点、焦点前和焦点后位置上的源特征。当单词处于焦点位置时，声源激发(EE)的动态范围比在其他环境中要大得多，元音较强，而周围的浊辅音较弱。这有效地增强了重读音节中元音/辅音的区别。这种增强可能比声源激发的绝对水平更重要。

一些研究集中在高声调和低声调可能被音高以外的声源特征所区别的可能性上。Pierrehumbert(1989)报道了高音高发音/重音和高音高之间的声源相关关系，而没有发现其他与高音高相关的声源，但是最近，epstein(2003)没有发现高音高和低重音之间一致的声源差异。这个问题是复杂的，因为在任何情况下都存在 f_0 和其他声源参数之间的交互，对此我们仍不清楚。

应该强调的是，除了在短语边界处出现挤喉音的音段外，这里提到的与语调类别相关的各种声源不会被听者作为音质的变化而听到。很简单，在任何使用正常嗓音的话语中，作为话语韵律表达的一部分，都存在着对声源参数的广泛调制。

本质上这是韵律中一个基本未被研究的方面，但随着在倪 Chasaide 和 Gobl (2004 a, 2004 b), 整体韵律分析方法势在必行，语言韵律——声调或语调——作为整个声源信号动态的瞬时性的变化，而不只是第一次谐波 f_0 的频率。

图 11.13 在瑞典话语声音源动力?我强度 detta ? rhundrade ?[印梯 xDEt: o: rh ? ndr@d]。

4.2 语音声源变异的副语言学层面

说话者在音质上使用时间转换来影响语音信号，比如他们的情绪和情感，就像他们与对话者和互动内容的态度一样。在口语互动中，可以被广泛地观察到在两个层面上传输：第一，以言语内容的形式，可以在文本中显示更多或更少；第二，在文本中无法表达的声音的声调中。声音的音调可以改变，也可以与口头内容相反。比如，说话者可能听起来很感兴趣或者漠不关心，这是他们真实态度的反映，但同样的，他们可能会用一种语气来表示对某个他们觉得非常无聊的话题感兴趣，或者假装漠不关心，掩盖一个非常真实但可控的兴趣、愤怒或其他情绪。值得记住的是，声音质量的改变与极端的情感联系在一起像这样，它们是超语言的并且可能是普遍的，不属于传统的符号系统。毫无疑问，声音的非指向性调音来表达情感，虽然往往植根于非自愿的情感表达的相关关系，但却更容易在语言或文化群体中被约定俗成，因而不大可能具有普遍性。

文学中的印象式观察传统上将特定的音质与特定的情感状态联系起来，例如，对于说英语的人来说，挤喉音被认为与厌烦有关，气嗓音与亲密有关，耳语的声音与机密性有关，尖锐的/刺耳的 harsh 声音与愤怒有关(参见，例如，laver, 1980)。关于口语交际这一重要方面的数量信息很少，这一事实主要反映了此类研究存在的许多技术和方法上的困难。

作为声源的可靠声学测量结果，该领域的大多数研究都倾向于关注 f_0 、音强和时长的最容易度量的参数。Scherer 和他的同事对情绪发声进行了广泛的研究，为 f_0 的变化(平均值

的变化、范围、轮廓类型、变异性)、音强(平均值的变化)和时间模式提供了丰富的信息。(见其他相关研究, 马苏契案, 1998 年, Carlson et al., 1992 年)当这些研究揭示了许多重要的见解时, 我们仍然缺乏对音质的关键方面的基本知识。一项对一些情感状态的声源相关性的初步研究强调了音质的根本重要性, 并说明许多声源参数在相当复杂的方面存在差异。

关于语音样本的问题应该进一步进行分析。到目前为止, 大多数研究都集中于愤怒、恐惧、快乐和悲伤等一些强烈情绪, 并且涉及了一些“语义上中性”的语句。这些研究表明, 个人可能不仅在行为动作/ 描述的风格上不同, 而且在他们模拟情感的能力也不同。并且, 说话者的描述虽然对他们自己感兴趣, 但很容易与自发的、自然出现的情感不同。与典型范例不同的是, Campbell 和 Mokhtari(2003)使用了一个非常大的语料库来记录一个人多年来的自发讲话。NAQ 的度量(见第 2 节) 说明声源的系统变异取决于与对话者(母亲、婴儿、朋友等)的关系以及他们之间互动的性质。这项工作突出了音质的作用, 为言语内容提供了提供明确、人性化的维度。

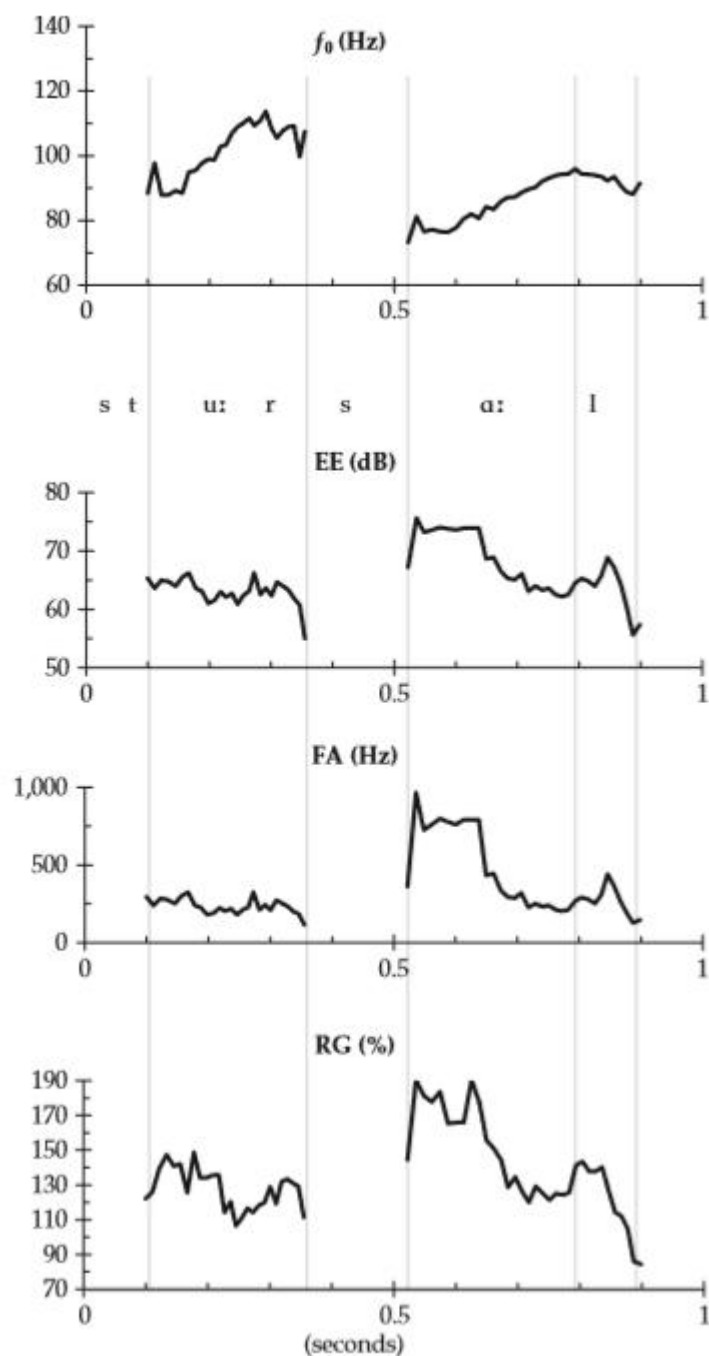


Figure 11.14 Variation in the voice source parameters EE, RA, and RG for the Swedish utterance fragment “stor sal” [stʊ:r sɑ:l].

图 11.14 瑞典语 “stor sal” (stʊ:r sɑ:l) 的语音来源参数 ee、ra 和 rG 的变化。

在一种截然不同的方法中，作者和他们的同事还使用了另一种截然不同的方法，即通过合成和感知实验来探索音质的效应投射。采用参数声源滤波器合成方法，对不同的音质产生不同的刺激。对基本方法的全面描述，见 Gobl & ni Chasaide, 2003b。一系列的知觉测试引出了不同的声音品质是否以及在多大程度上唤起了不同的情感色彩。每个单独的测试测试一对相反的情感属性，例如，无聊/感兴趣。采用 7 分制，0 表示没有情感色彩，0 (从-3 到 +3) 的偏差表示任何感知情感的强度。测试的全部属性对包括放松/紧张、内容/愤怒、友好/敌对、悲伤/快乐、无聊/感兴趣、亲密/正式、胆怯/自信和害怕/不害怕。

最初的实验结果如图 11.15 所示 (Gobl & ni) (Chasaide, 2003b), 并展示每个测试对每个音质测试的价值评级。它们说明了, 仅仅改变声音质量就能给原本中性的话语赋予不同的情感色彩。此外, 在为传统智慧提供广泛支持的同时, 研究结果还提出了一些关于音质和情感关联的改进建议。举个例子, 无聊程度最高的声音不是挤喉音, 而是松挤喉音; 同样的音质在这里比气嗓音更能唤起亲密感。他们还指出, 一种特定的音质往往与一组情感状态 (虽然不一定是相关的, 但往往是相关的) 联系在一起, 而不是通常所暗示的影响语音质量的一对一映射。所以, 举例来说, 松挤喉音不仅无聊度和亲密度高, 而且放松度和内容相关度也高。

进一步的实验 (Gobl 等, 2002; Yanushevskaya 等人 (2005) 研究了音质和 f0 在暗示影响 cueing affect 中的结合。音高调制器对声音的声调也有很大的贡献, 就像上面提到的, 在 f0 与某些情感的关联上有一个相当大的文学样式。在这些实验中, 除了早期测试的音质刺激 (VQ 刺激), 还包括进一步的刺激集, 包括第一个实验中使用的不同或进一步修改的话语。一组 (f0 刺激) 包括不同的 f0 轮廓, 还有正常嗓音。使用的 f0 轮廓是基于 Mozziconacci (1995) 中提供的与影响相关的轮廓。另一组刺激包括与第一组相同的音质数组, 但这一次, 每个情感 f0 轮廓线都与一个或另一个音质组合 (f0 + VQ 刺激)。

这些实验的结果表明, 与音质操纵或音质与 f0 变异性的结合这两者相比, 单独的 f0 刺激在表达爱意时无效的。这些结果可能解释了为什么试图通过将产出研究中描述的 f0 的变异结合对于 f0 合成中复制情感语言的尝试没有成功。

这种相同的实验范式也曾被用于从声音到情感的映射的跨语言/文化差异, 对来自不同语言背景的主体使用相同的测试, 包括英语和日语 (见 Yanushevskaya et al., 2006)。虽然在许多情况下, 特定刺激与特定影响之间的联系对于不同的语言组来说是相当相似的, 但确实出现了一些显著的差异。例如, 英语受试者的愤怒情绪与紧音有关, 并且这种紧音的 f0 较高且呈动态变化, 而日语受试者的这种音质与亲密度有关。

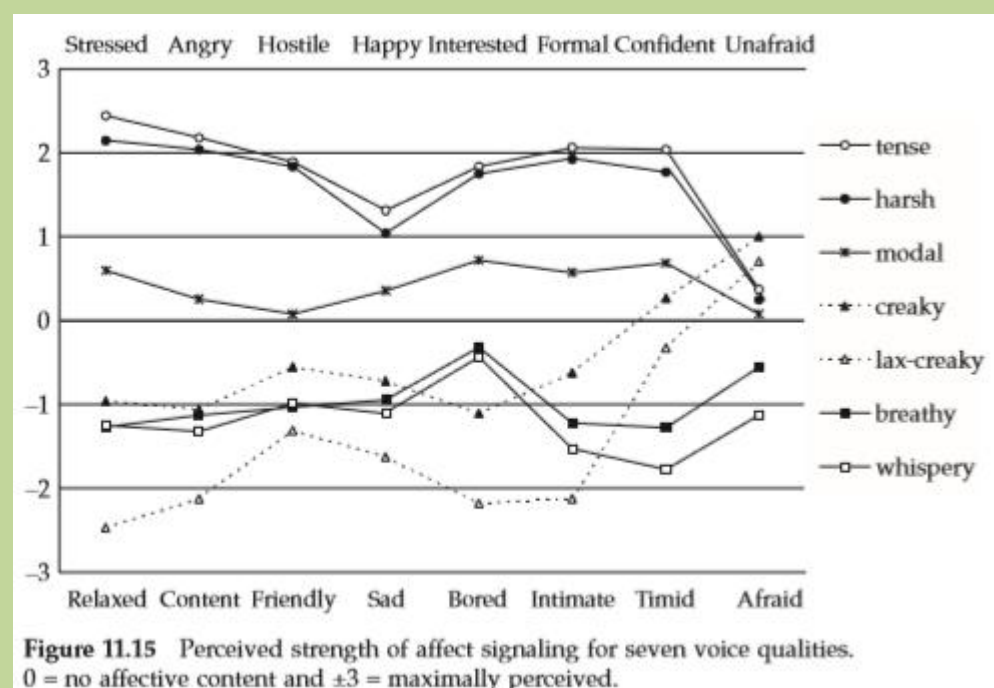


图 11.15 对 7 种声音质量的感知信号强度。0 = 不内容和 \pm affective 3 = maximally perceived

4.3 社会语言学和跨语言变异

音质也可能具有社会语言学的功能，并为不同的语言、区域和社会群体服务。咽上特征和发音特征可用于此目的。对于那些练习过外语的人来说，音质的跨语言差异的确是一种口音的重要方面，但很难教授，因为他们从来没有在语言或应用语言文学中被描述过。这可能会导致对文化的误解，就像一位母语演讲者很容易使用他/她自己的行为信号的副语言系统来扮演外国人的音质一样。上面提到的日语和英语听众之间的差异正好说明了这种误解是如何产生的。

在特定的语言或方言群体中，音质特征可以用来划分社会子群体。在爱丁堡英语中，挤喉音的发生率越高，社会地位越高，而儿语发音和刺耳的声音则与较低的社会地位有关(esling, 1978)。格拉斯哥英语中也有类似的发现(Stuart-Smith, 1999)。在诺维奇，工薪阶级口音和中产阶级口音的区别是建立在习惯性发音和咽上设置的基础上的(trudgill, 1974)。其他社会群体也可能倾向于使用不同的音质。rose(1989)认为，吴语镇海方言中所发现的极端刺耳的“咆哮”发声方式，在说话者的性别和年龄上存在差异，女性最不刺耳，男性最刺耳。在研究与性别和年龄相关的差异时，区分真正的社会语言学标记物和喉部解剖造成的差异当然是重要的。

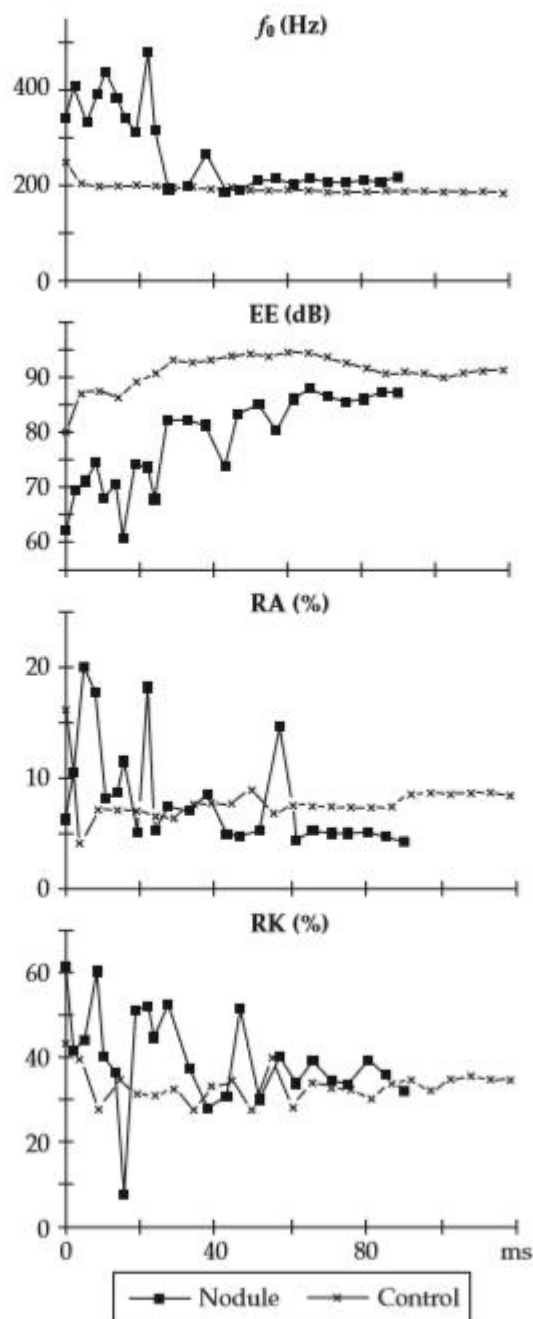


Figure 11.16 Values of f_0 , EE, RA, and RK compared for a speaker with vocal fold nodules and a normal speaker, from the initiation of voicing (following stop release) in the nonsense word *baa*.

图 11.16 f_0 、ee、ra 和 rK 对一个声音折叠和正常扬声器的比较价值

4.4 声源的超语言决定因素

还有其他因素决定了声音的质量，其中许多因素是说话人无法控制的。喉结构的大小、形状和肌张力的差异起着重要作用。男性、女性和儿童的声音主要反映了解剖学上的差异，尽管基于解剖学的内在特征可能会随着社会文化背景的不同而增强或减弱。例如，在男性环境中工作和竞争的女性可能会选择一种更接近男性的发音方式。我们对女性和儿童声音的理解落后于男性声音。关于女性声音的一些描述见 Karlsson (1992)，hanson (1995, 1997)，holmberg 等人 (1988)，关于儿童和成年男性声音的差异见 Gobl (1988)。

音质也受到个人身心健康的影响。已经有相当多的研究表明,某些精神疾病与声学(主要是 f_0 和音强)相关,如抑郁症和精神分裂症。关于抑郁症的声音指标的研究综述见 Scherer(1987)。喉部结构的病理学也会影响音质,许多关于声音的研究都是由医学推动的。图 11.16 显示了患有声带结节的女性说话者的许多源参数(f_0 、EE、RA、RK),并将之与年龄、性别和口音匹配的正常说话者相对比。这种病理现象似乎与发声之初特别相关,振动周期在许多方面受到严重干扰。 f_0 很高,声门激发较弱(EE 低),有相当程度的动态泄漏(RA 高),脉冲形状一般较对称(RK 高)。可能与实际值同样重要的是声门脉冲在初始时间间隔内的不稳定性,这从所有参数从一个脉冲到另一个相邻脉冲的相当大的波动可以看出。在第 10 个和第 20 个声门循环之间的某个时间点,发音模式突然发生了变化。 f_0 降低 1 个倍频程直至正常值,其他声源参数显示声门脉冲更加正常稳定。关于这项研究的更多细节,请参见 Kane 和 ni Chasaide(1992)。

除了目前提到的语言和非语言因素之外,音质还承载着独特的个人信息,在帮助我们识别说话者并将他们区分开来方面发挥着重要作用。

5 未来研究

人类的声音进化成一种工具,用于传达许多不同类型的信息,而人类的听觉器官已经发展出检测非常小、非常微妙的音质变化并解释它们功能的能力。尽管近年来描述和模拟声源变化的技术有所改进,但我们的能力远远落后于人类的耳朵。在这一阶段,我们只需要了解研究的规模和复杂性,就能充分理解语音在言语交际中的应用。迄今为止的大多数研究,无论是在分析的数据量方面,还是在采取的各种声源措施方面,都非常有限。

因为声源同时许多不同的层次上传递信息,所以我们对任何一个方面的理解都取决于我们对整体的把握,这样我们就可以正确地解释一个维度相对于另一个维度的变化。举个明显的例子,说话人之间的差异通常是非常显著的,这可能比说话人内部(有时是微妙的)传递语言或副语言信息的变异类型更大。

或许不那么明显的是,我们需要的研究跨越了目前看来可能是独立的研究领域。我们认为,语言韵律和受韵律影响的副语言信号是一个系统的两个方面/一体两面。虽然,像往常一样,我们在这里把它们分别放在不同的标题下讨论,但我们认为,只有相对于彼此,才能正确地理解它们。目前,口语交际的这两个方面是作为两个不同的实体来研究的,主要是在不同的学科中。在语言学中,重点主要放在音位上,语调的对立层面。很少有人关注语调在情感信号中的作用,尽管这在早期的语言学研究中被视为语调的一个核心功能(如 o' Connor & arnold, 1961)。情感的声音表达研究主要是在心理学领域进行的,对语言韵律的研究较少。我们认为,影响涉及的副语言信号不是一个单独的系统,而是对语言韵律的基本要素进行系统的修改和增强。

理解韵律的这两个维度如何协作的关键可能是提供对语音动态的更全面的描述。在当前的大多数文献中,流行的观点是,语言韵律与 f_0 动态有关,副语言信号在很大程度上取决于音质的变化。如上所述,声源参数对韵律的语言和副语言层面都有重要影响。同样, f_0 的动态变化对两者都有重要影响。理想情况下,我们需要在描述框架中包括整个声源信号(f_0 和其他重要的声源参数)如何随时间动态变化。正如 ni Chasaide 和 Gobl (2004a)所指出的那样,这种全面的方法有望通过提供对语言韵律工作的更全面的理解,为更好地理解副语言韵律奠定基础。通过消除语言学和副语言学、 f_0 和音质之间毫无帮助的二分法,我们应该能够实现语音韵律的统一。

对韵律作如此全面的解释的进展将不可避免地取决于是否有适当的分析工具。第 2 节概述的手工交互技术允许对资料来源进行细粒度的分析,但由于其劳动密集的性质,不适用于在这一领域取得进展所理想需要的大规模研究。因此,研究议程不仅必须针对描述性研究,

而且必须针对设计新的技术或加强现有的技术,以便自动取得准确的声源数据。许多新的自动算法已经提出了,虽然它们可能对估计声源的某些全局的、长期的方面很有用,但是它们还不能生成所需的精确度,以提供用于更多目的的更细致的分析。到目前为止,大多数技术都集中在信号处理解决方案上:我们相信,为了在声源过滤分解领域取得重大进展,将需要在更大程度上以语音生成知识为指导的新方法。在语音研究的其他领域,声学分析必须辅以生理实验,以阐明潜在的生产过程,例如使用高速数字成像的声带振动,如hirose(本卷)所述。

更好地理解声源信号的各个方面以及它在语音中的变化将为许多应用打开大门。最直接的应用是在语音合成中提供更具表现力和潜在的更灵活的声音,这将大大提高基于合成的设备的可接受性。我们可以设想,在未来的某一天,为有声音障碍的人量身定制的语音助手的可能性,旨在匹配用户的原始音质,并允许调制影响传递的语音。此外,当有相当准确的自动分析程序可用时,我们可以设想在说话人识别和验证等领域的许多其他应用。随着对正常和病理嗓音变异范围和类型认识的加深,这些技术也可能促进更先进的声学筛查方法和声音障碍诊断程序。