

非言语声音影响汉语听者言语声音的知觉^{*}

刘文理 乐国安

(南开大学周恩来政府管理学院社会心理学系, 天津 300071)

摘 要 采用启动范式, 以汉语听者为被试, 考察了非言语声音是否影响言语声音的知觉。实验 1 考察了纯音对辅音范畴连续体知觉的影响, 结果发现纯音影响到辅音范畴连续体的知觉, 表现出频谱对比效应。实验 2 考察了纯音和复合音对元音知觉的影响, 结果发现与元音共振峰频率一致的纯音或复合音加快了元音的识别, 表现出启动效应。两个实验一致发现非言语声音能够影响言语声音的知觉, 表明言语声音知觉也需要一个前言语的频谱特征分析阶段, 这与言语知觉听觉理论的观点一致。

关键词 言语知觉; 听觉理论; 动觉理论; 频谱对比效应; 启动效应

分类号 B842

1 前言

言语知觉主要关心人们如何从连续的、变化的声学信号中感知分离的语言学单元, 即人们如何在头脑中处理声学信号和语音单元之间的映射问题。言语的声学信号存在变异性和非线性。人们产生言语时存在协同发音(coarticulation)效应, 相邻音段的发音动作会存在部分或完全重叠, 如产生一个辅音-元音(consonant-vowel, CV)音节/du/, /d/和/u/的发音嘴型即存在完全的重叠。协同发音导致的一个问题是由此产生的音段的声学信号会受到语境影响, 表现出很大的变异性(Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), 且发音动作的重叠也导致相邻音段间的声学信号存在重叠(非线性)。此外个体间声道形状的差异、语速、韵律、情绪状态等因素也会导致音段的声学信号存在变异。

根据言语声学信号的特点, 研究者提出了截然不同的言语知觉理论, 有影响的理论包括言语知觉的动觉理论和听觉理论。言语知觉动觉理论由Liberman 及其同事(Liberman et al., 1967; Liberman & Mattingly, 1985)提出, 该理论认为声学信号和要感知的音位间缺少对应关系, 音位的不变性存在于发音水平而非声学水平。言语知觉的目标是讲话者

的发音动作表征, 即表征在脑中的恒定的动作命令。动觉理论的第二个假设是人类感知言语的能力主要依赖一个特殊的解码器或特定于言语的模块, 该模块是人类特有的, 且与非言语声音的感知不同。与动觉理论相对立的是言语知觉听觉理论(Diehl, Lotto, & Holt, 2004), 听觉理论认为言语知觉的目标是物理的声学线索, 而非言语产生中的动作事件, 听者可以直接从声学信号中感知语音范畴, 即言语知觉不需要动作表征的中介; 言语声音的感知不具有特殊性, 其与非言语声音的感知有很多类似之处。总之, 两个理论争论的焦点主要围绕两方面展开。首先是言语知觉的目标, 动觉理论认为知觉的目标是发音中的动作事件, 听觉理论认为知觉的目标是声学信号。其次是言语知觉的特殊性和一般性问题。动觉理论认为言语知觉具有特殊性, 其内在机制是非人类的动物和非言语声音知觉所不具备的; 听觉理论认为言语知觉具有一般性, 很多特征是和人类的动物及非言语声音知觉所共享的。

围绕理论争论的问题, 研究者在言语知觉的目标、动物被试的言语知觉能力及非言语声音和言语声音知觉之间的关系等方面开展了大量实验。其中考察非言语声音和言语声音知觉之间的关系有助

收稿日期: 2011-08-24

^{*} 中央高校基本科研业务费专项资金资助项目(NKZXB10152)。

通讯作者: 刘文理, Email: wenli_160@sina.com

于澄清两种理论之间的争论。动觉理论认为言语知觉有一个特殊的模块,该模块只负责言语声音的知觉,非言语声音知觉不具备这种特殊性,两种声音的知觉相互之间没有影响。听觉理论认为言语声音和非言语声音的知觉具有共同性,都依赖听觉皮层对声音信号中声学特征的分析,因此如果非言语声音和言语声音之间具有相同的声学特征,非言语的声音语境有可能影响到言语声音的知觉。

言语知觉中确实存在明显的语境效应,研究者首先发现连续言语中一个音段的知觉会受到周围音段的影响。Mann (1980)将合成的/da-ga/连续体与发音人产生的/al/和/ar/刺激相拼接,要求被试识别/da-ga/连续体。结果发现当前面是/ar/时,被试更可能将/da-ga/连续体上的刺激识别为/da/;而当前面是/al/时,个体更倾向于将刺激识别为/ga/。研究者用协同发音的知觉补偿(perceptual compensation of coarticulation)来解释这种语境效应。单独产生时,/da/音节有更靠前的发音部位(齿龈),/ga/的发音部位更靠后(舌根);而在连续言语中,由于协同发音的影响,/da/和/ga/的发音部位会有一些变化。如在/al/后,与在/ar/后相比,由于/l/发音部位更靠前,导致后接的/d/或/g/的发音部位也更靠前;而在/ar/后,/d/或/g/的发音部位相比在/al/后更靠后。与/r/、/l/语境对/da/和/ga/发音部位的影响相比,其对/da/和/ga/知觉方向的影响正好相反。如在/al/后刺激会有更靠前的发音部位,这与/da/一致,因此听者应该有更多的/da/反应,而事实上听者更多的将刺激知觉为/ga/;在/ar/后听者更多的将刺激知觉为/da/。因此听者的知觉似乎弥补了协同发音所产生的影响。Mann (1980)认为这表明言语知觉考虑了发音的一些知识,如协同发音的影响,这与言语知觉动觉理论的观点相符。之后的研究表明缺乏相应语音范畴经验的日语被试(日语听者分辨英语/r/-/l/对比存在困难)和 4 个月的婴儿也能表现出类似的语境效应(Mann, 1986; Fowler, Best, & McRoberts, 1990),这似乎表明知觉弥补具有普遍性,独立于语言经验和听者的年龄。

然而进一步的实验发现非言语的声音语境也能诱发言语目标知觉中的这种语境效应。Lotto 和 Kluender (1998)用正弦波滑音模拟了/ar/和/al/结尾处的 F3(第三共振峰)过渡,这类滑音为非言语声音。结果发现非言语的滑音影响了/da/-/ga/连续体的识别,影响方向和语音语境(ar/和/al/)完全一致。进一步的实验发现纯音(纯音频率分别接近/ar/和

/al/的 F3 终止频率)也能影响到/da/-/ga/连续体的识别。实验结果表明言语知觉不仅受到语音语境的影响,非言语声音也能影响到言语声音的知觉。研究者认为这种语境效应可以从声学 and 听觉方面进行解释,是频谱对比效应的一个例子,并非反映了听者对协同发音的知觉补偿(Lotto & Kluender, 1998): /al/语境有更高的 F3 终止频率,听者易于将后面的刺激感知为有更低 F3 起始频率的声音,即/ga/音节;而当前面的语境是/ar/时,其有更低的 F3 终止频率,听者易于将后面的刺激感知为有更高 F3 起始频率的声音,即/da/音节。因此听者的知觉表现出一种频谱对比效应,即使前面的语境是非言语声音,只要前后声音的声学特征提供了对比性的频谱成分,就能诱发言语知觉的这种语境效应。

之后更多的实验发现非言语声音语境能够影响到言语声音的知觉。Holt, Lotto 和 Kluender (2000)发现在/bvb/和/dvd/语境中,模仿/b/和/d/的 F2 过渡段的正弦波滑音和纯音能够影响到中间元音的知觉,影响模式类似/b/和/d/语境,表现出频谱对比效应。Coady, Kluender 和 Rhode (2003)发现/e/和/o/语境以频谱对比的方式影响到/ba/-/da/连续体的范畴化,进而带通过滤后的/e/和/o/的 F2 以及/e/和/o/频谱互补的声音(都是非言语声音)也能以频谱对比的方式影响到/ba/-/da/连续体的知觉。Holt(2005)发现即使非言语声音语境(20 个 70ms 的纯音)与语音目标之间的间隔达到 1.3s 或者中间间隔很多中间频率的纯音,非言语声音语境也能以频谱对比的方式影响到/da/-/ga/连续体的知觉。进而,非言语的纯音语境能与言语语境相结合影响语音目标的知觉(Holt, 2006),如果纯音语境与言语语境的频谱相匹配,则能提高语境效应;如果二者冲突,非言语的纯音语境能够抵消甚至反转言语语境的影响。另有研究表明即使非言语声音出现在言语声音后面,也能以频谱对比的方式影响到言语声音的知觉(Wade & Holt, 2005);言语声音的语境也能影响到非言语声音的知觉(Stephens & Holt, 2003)。

除频谱对比效应,最近的研究也发现有相同频谱成分的非言语声音能够加快元音目标的识别,即表现出启动效应。Wallace 和 Blumstein (2009)在一个实验中考察了非言语声音对元音目标的启动效应。目标刺激是元音/i/和/a/,启动刺激是 2 个非言语的纯音复合音,每个复合音由两个纯音组成。其中一个复合音的纯音频率分别与元音/i/的 F1 和 F2 频率相匹配,另一个复合音的纯音频率与元音/a/的

F1 和 F2 频率相匹配。结果发现纯音复合音能够加快与其频率匹配的目标元音的识别,即表现出启动效应,特别是启动刺激的时长较短时(25~100ms)。

以上研究表明当非言语声音和言语声音共享相同的声学特征时,非言语声音的加工能够影响到言语声音的知觉,表现出频谱对比效应或启动效应,这支持了言语知觉听觉理论的观点。听觉理论认为言语知觉没有特殊性,与非言语声音知觉共享类似的过程,可以用一般听觉加工机制的属性进行解释;只要非言语声音与言语声音有匹配的声学特征,二者的加工就存在相互影响的可能。言语知觉听觉理论认为言语知觉的目标是言语产生中的动作表征,具有特殊性;非言语声音没有相应的动作发音表征,不具有这种特殊性,因此非言语声音的知觉和言语声音的知觉不存在交集。而以上实验中发现的频谱对比效应或启动效应表明非言语声音的加工确实能够影响到言语声音的知觉,这为言语知觉听觉理论提供了证据。

以上实验结果主要是在以英语为母语的被试中得到的,其它语言中还缺少类似的实验证据。不同的语言其语音特征存在较大的差别,具体到汉语和英语,汉语属于汉藏语系,英语属于印欧语系(伍铁平, 2006)。两种语言的语音存在很多差别,小到音位的发音方式、声学特征,音位的数量和组合方式,大到语言的韵律特征等等,都存在差别。如汉语与英语的音位数目不同,音位的组合方式也不同,如英语有复辅音(/pl/、/str/等),而现代汉语则没有;汉语有声调的区别,英语则没有。语音特征的差别是否会影响到听者的知觉系统,导致不同语言使用者的言语知觉存在系统的差异?这是值得探讨的一个问题。Huang 和 Holt(2009)考察了汉语声调知觉是否受到非言语语境的影响。目标刺激是声调范畴连续体,变化基频(f_0)的起始频率从一声变化到二声。非言语的启动刺激包括复合音(f_0 和谐音)序列和单纯音(f_0 频率)序列,每类刺激又构成高 f_0 和低 f_0 两种语境条件。结果发现声调连续体的知觉受到非言语语境的影响,不管是复合音还是纯音,高 f_0 语境导致更多的低 f_0 声调反应(二声,起始频率低),低 f_0 语境导致更多的高 f_0 声调反应(一声),即非言语语境以频谱对比的方式影响到汉语声调序列的知觉。这为非言语声音的加工影响汉语听者超音段特征的知觉提供了证据。

目前研究将以汉语大学生为被试,通过启动范式考察汉语听者的音段知觉是否也受到非言语语

境的影响。如果汉语被试的音段知觉也受到频谱构成类似的非言语声音的影响,则表明不同语言使用者的言语知觉具有共性,这为言语知觉的听觉理论提供了更多的证据。除了能够为理论争论提供新的证据外,使用启动范式考察非言语声音对言语声音知觉的影响也为了解言语知觉的内在机制提供了新的途径。通过启动范式,系统的变化启动刺激(非言语声音)和目标刺激(言语声音)之间频谱构成的关系,比较不同条件下的启动效应,可以为了解言语知觉过程中各种声学特征的分析 and 整合机制提供非常方便的手段。

研究包括两个实验。实验 1 考察了非言语声音对汉语辅音范畴连续体知觉的影响。实验 2 考察了非言语声音对汉语元音知觉的影响。如果不同语言使用者言语知觉的共性大于差异性,我们预期在实验 1 和实验 2 中能够发现频谱对比效应或启动效应,即非言语声音的加工能够影响汉语音段的知觉。进而通过非言语声音和言语声音之间的频谱关系和影响方向,我们也可以初步推断音段知觉中频谱分析和整合的机制。

2 实验一: 纯音对汉语辅音范畴连续体知觉的影响

2.1 方法

2.1.1 被试 南开大学本科生 20 名,平均年龄 20.9 岁,其中男生 10 名。视力或矫正视力正常,听力正常。实验后付给报酬。

2.1.2 材料 言语刺激由 Klatt 合成器(SenSyn 1.1)合成,变化 F2 过渡的起始频率产生从/ba/到/da/变化的 8 个刺激,刺激时长 260ms。F2 起始频率从 900Hz(/ba/)以 100Hz 为步长变化到 1600Hz(/da/)。60ms 过渡后 F2 抵达 1200Hz 的稳态频率,稳态部分持续 200ms。F1、F3、F4 和 F5 变化在 8 个刺激中保持一致, F1 在 0~60ms 内从 710Hz 线性上升到 950Hz,之后维持稳态。F3、F4 和 F5 一直处于稳态,频率分别是 2600Hz、3500Hz 和 4400Hz。基频(F_0)从 140Hz 下降到 130Hz,有一个下降的声调。输出刺激以 11-kHz 的抽样率,16-bit 的分辨率数字化存储为波形文件。

非言语刺激是两个正弦波纯音,由声音处理软件 Cool Edit 2.0 生成,抽样率和分辨率与言语刺激相同。纯音频率分别与连续体终端的/ba/和/da/的 F2 起始频率相同,即 900 和 1600Hz,刺激时长 100ms,刺激的起始和结束分别有 5ms 的渐变期。

将纯音和言语声音进行拼接,刺激间间隔 50ms。2 个纯音分别和 8 个 CV 音节配对共产生 16 个声音文件,刺激的简要语图见图 1。另外对纯音刺激和言语刺激的 RMS (root mean square)能量进行了匹配。

2.1.3 程序 实验程序、刺激呈现和数据收集由 E-Prime 1.1 完成。实验在安静的实验室里进行,被试用耳机接收声音文件,音量调整到舒适的水平。被试首先完成练习,16 个声音文件每个呈现 2 次构成练习刺激,要求被试判断纯音后面的 CV 音节是“ba”还是“da”,如果是“ba”,按“Z”键进行反应,如果是“da”,按“M”键进行反应。如果 3s 内被试无反应,继续呈现下一个刺激。被试反应后间隔 1.5s 呈现下一个刺激。正式实验中,16 个声音文件每个呈现 10 次组成刺激项目,分两个 block

呈现,每个 block 包括其中的 5 次呈现(80 个项目)。整个实验约花费 20min。

2.2 结果与讨论

用 SPSS 15.0 对实验数据进行了分析。2 名被试的数据因为程序问题记录不全,没有参与进一步的分析。剩余 18 名被试在两种纯音启动条件下对各言语目标的平均“ba”反应百分比见图 2。首先对被试反应百分比进行了 2(纯音条件) $\times$ 8(言语目标)重复测量的方差分析。纯音启动条件主效应显著, $F(1, 17) = 10.14, p < 0.01$; 1600Hz 纯音启动条件下被试有更多的/ba/反应。言语目标主效应显著, $F(7, 119) = 460.96, p < 0.001$; 随刺激向/da/端变化被试/ba/反应的百分比迅速下降。纯音条件 $\times$ 言语目标交互作用显著, $F(7, 119) = 3.7, p < 0.05$ 。简单

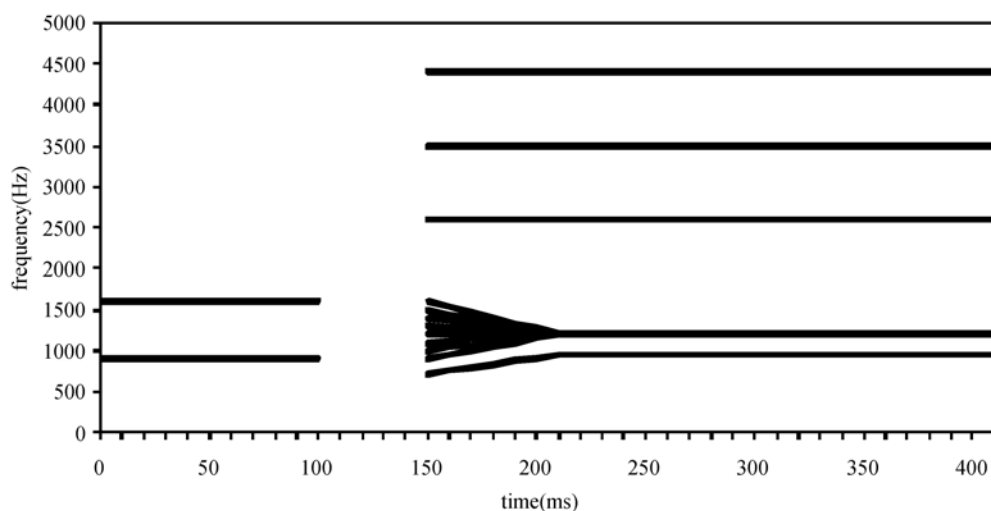


图 1 实验 1 刺激的简要语图。右侧是/ba-da/连续体上的 8 个刺激;左侧是两个纯音启动音。

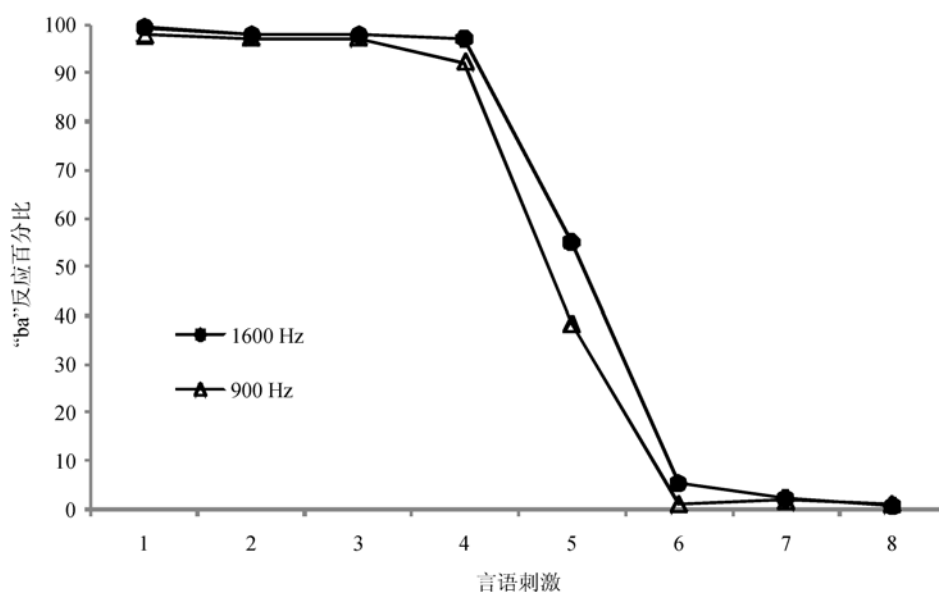


图 2 两种纯音启动条件下被试在各言语刺激上平均的“ba”反应百分比

效应分析表明被试在言语刺激 4 和 5 上的/ba/反应百分比在两种启动条件下差异显著, 刺激 6 的/ba/反应百分比差异边缘显著, 被试在 1600Hz 启动条件下的/ba/反应百分比都高于 900Hz 启动条件: 刺激 4, $F(1, 17) = 4.64, p < 0.05$; 刺激 5, $F(1, 17) = 5.74, p < 0.05$; 刺激 6, $F(1, 17) = 3.68, p = 0.07$ 。

另外对被试的识别数据进行了 probit 分析, 计算出了每个被试在两种启动条件下识别曲线的音位界限值(识别百分比为 50%的刺激值, 即识别曲线的中间值), 1600Hz 纯音条件下的平均界限值(标准差)为 5.05 (0.39), 900Hz 纯音条件下的平均界限值(标准差)为 4.74 (0.42), 这表明 1600Hz 条件下被试识别曲线的音位界限明显向/da/端偏移, 即有更多的/ba/反应。对两种条件下音位界限的配对 t 检验发现差异显著, $t(17) = 3.19, p < 0.01$ 。

结果分析发现非言语纯音影响汉语辅音连续体的识别, 影响方向表现出显著的频谱对比效应: 当纯音频率更高时, 即与连续体上终端刺激/da/的 F_2 起始频率相同(1600Hz), 被试更倾向于将后面的言语刺激识别为有更低 F_2 起始频率的声音, 即/ba/音节; 当纯音频率更低时, 即与连续体上终端刺激/ba/的 F_2 起始频率相同(900Hz), 被试更倾向于将后面的言语声音范畴化为有更高 F_2 起始频率的刺激, 即/da/音节; 特别是对于模糊的言语目标, 即处于范畴连续体中间的刺激, 被试的范畴化显著受到纯音语境的影响。因此, 当前面的非言语声音与后面言语目标的频谱构成一种对比关系时, 汉语听者辅音范畴连续体的知觉也会受到非言语语境的影响, 表现出频谱对比效应, 这与英语被试的研究结果一致, 也与言语知觉听觉理论的观点一致。言语知觉听觉理论认为只要相邻的声音之间构成频谱上的对比关系, 不管前面的声音是言语的还是非言语的, 都会对后面言语声音的加工产生影响, 具体表现为一种对比效应。而动觉理论认为这种语境效应反映了被试对协同发音的知觉补偿, 虽然这可以解释言语语境的效应, 但不能解释非言语语境的效应, 因为非言语声音缺乏动作表征, 不会有协同发音的经验。进而, 动觉理论研究者认为言语语境效应可以用知觉补偿来解释, 非言语语境的效应存在不同的解释机制, 如可能反映了刺激之间的掩蔽效应(Fowler, Brown, & Mann, 2000; Fowler, 2006; Viswanathan, Magnuson, & Fowler, 2010)。但研究表明掩蔽解释不了非言语语境的这种效应, 即使非言语语境与语音目标之间的间隔达到 1.3s 或者中间

间隔很多中间频率的纯音, 频谱对比效应仍然存在(Holt, 2005; Lotto & Holt, 2006)。因而, 听觉理论对非言语和言语语境效应提出了统一而简洁的解释, 而动觉理论对于非言语语境效应始终缺乏合适的解释。

3 实验二: 纯音和复合音对汉语元音知觉的影响

3.1 方法

3.1.1 被试 南开大学本科生 18 名, 其中男生 9 名, 平均年龄 20.6 岁。视力或矫正视力正常, 听力正常。所有被试都没有参加实验 1, 实验后付给报酬。

3.1.2 材料 言语刺激是汉语元音 /i/ 和 /a/, 由 Klatt 合成器合成, 刺激时长 250ms。/i/ 的 5 个共振峰维持稳态, 频率分别是 350Hz、2150Hz、3280Hz、3870Hz 和 4560Hz。/a/ 的 5 个共振峰也维持稳态, 频率分别是 950Hz、1200Hz、2600Hz、3500Hz 和 4400Hz。基频(F_0)从 140Hz 下降到 130Hz。输出刺激以 11-kHz 的抽样率, 16-bit 的分辨率数字化存储为波形文件。

非言语刺激是 2 个纯音和 2 个纯音复合音, 由 Cool Edit 2.0 生成, 抽样率和分辨率与言语刺激相同。2 个纯音频率分别是 1200Hz 和 2150Hz, 即分别与元音/a/和/i/的 F_2 频率相同。每个复合音都由两个纯音混合而成, 其中一个复合音的两个纯音频率是 950Hz 和 1200Hz, 即与元音/a/的 F_1 和 F_2 频率相同; 另一个复合音的两个纯音频率与元音/i/的 F_1 和 F_2 频率相同, 分别是 350Hz 和 2150Hz。纯音和复合音刺激时长 100ms, 刺激的起始和结束分别有 5ms 的渐变期, 另外对非言语刺激和言语刺激的 RMS 能量进行了匹配。实验刺激的语图见图 3。

3.1.3 程序 实验设备、测试环境同实验一。4 种启动条件(2 个纯音和 2 个复合音) $\times$ 2 个言语目标共构成 8 种刺激配对, 其中 4 个配对的频率相匹配, 4 个配对的频率不匹配。8 个刺激配对每个呈现 20 次, 共呈现 160 次, 分两个 block (每 10 次呈现构成一个 block)进行测试。被试首先听到启动音, 间隔 25ms 后呈现目标声音, 要求被试快速判断目标声音是“i”还是“a”。其中一半被试对“i”按“Z”键反应, “a”按“M”键进行反应; 另一半被试的反应选择相反(“a”按“Z”键, “i”按“M”键)。如果 3s 内被试无反应, 继续呈现下一个刺激。被试反应后间隔 1.5s 呈现下一个刺激。正式实验前先进

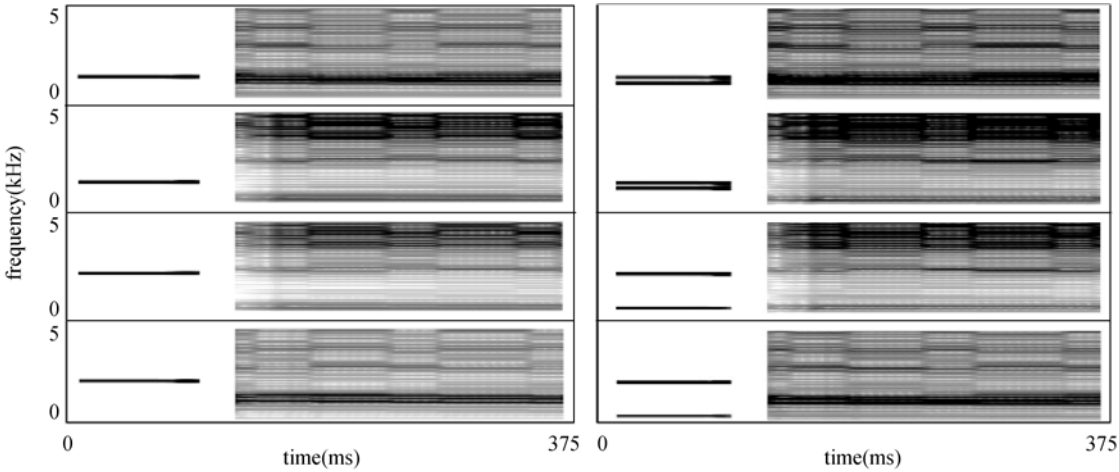


图 3 实验 2 刺激语图。左侧是纯音启动条件，由上到下分别是：启动音 1200Hz，元音/a/(匹配)；启动音 1200Hz，元音/i/(冲突)；启动音 2150Hz，元音/i/(匹配)；启动音 2150Hz，元音/a/(冲突)。右侧是复合音启动条件，由上到下分别是：启动音 950~1200Hz，元音/a/(匹配)；启动音 950~1200Hz，元音/i/(不匹配)；启动音 350~2150Hz，元音/i/(匹配)；启动音 350~2150Hz，元音/a/(不匹配)。

行练习, 8 个刺激对每个呈现 3 次构成练习项目, 被试熟悉实验任务后开始正式实验。

3.2 结果与讨论

用 SPSS 15.0 对实验数据进行了分析。对被试

在每种启动-目标条件下的反应时及错误率进行了计算, 计算反应时时剔除了被试的错误反应及三个标准差以外的数据(少于 4%)。被试在各种启动条件下的平均反应时(标准差)及错误率见表 1。

表 1 被试在各种启动条件下的平均反应时(标准差)及错误率

启动条件	/a/		/i/	
	反应时	错误率	反应时	错误率
1200Hz	407.10(90.45)	3.33	452.69(93.44)	3.89
2150Hz	442.66(60.93)	4.44	418.84(89.87)	2.78
950-1200Hz	419.46(60.41)	1.67	450.28(52.73)	5.56
350-2150Hz	454.79(66.15)	5.56	416.15(57.28)	1.11

首先对纯音启动条件下的数据进行了 2 (启动条件) × 2 (元音目标) 重复测量的方差分析。反应时的分析发现启动条件和元音目标的主效应不显著, 启动条件 × 元音目标交互作用显著, $F(1, 17) = 7.34, p < 0.05$ 。简单效应分析表明 1200Hz 启动条件下元音/a/的反应时显著快于元音/i/的反应时, $F(1, 17) = 5.88, p < 0.05$; 元音/a/在 900Hz 条件下的反应时显著快于 2150Hz 条件, $F(1, 17) = 7.49, p < 0.05$; 元音/i/在 2150Hz 条件下的反应时快于 900Hz 条件, 差异达到边缘显著, $F(1, 17) = 4.31, p = 0.05$ 。错误率分析表明启动条件、元音目标及二者的交互作用都不显著。

对复合音启动条件下的数据也进行了 2 × 2 重复测量的方差分析。反应时分析发现启动条件 × 元

音目标交互作用显著, $F(1, 17) = 23.53, p < 0.001$ 。简单效应分析表明 950-1200Hz 条件下元音/a/的反应时显著快于元音/i/的反应时, $F(1, 17) = 4.49, p < 0.05$; 元音/a/在 950-1200Hz 条件下的反应时显著快于 350-2150Hz 条件, $F(1, 17) = 10.99, p < 0.01$; 350-2150Hz 条件下元音/i/的反应时显著快于元音/a/的反应时, $F(1, 17) = 7.89, p < 0.05$; 元音/i/在 350-2150Hz 条件下的反应时显著快于 950-1200Hz 条件, $F(1, 17) = 12.68, p < 0.01$ 。错误率分析发现启动条件 × 元音目标交互作用显著, $F(1, 17) = 5.0, p < 0.05$ 。简单效应分析发现 950-1200Hz 条件下元音/a/的错误率显著低于元音/i/的错误率, $F(1, 17) = 4.5, p < 0.05$; 元音/i/在 350-2150Hz 条件下的错误率显著低于 950-1200Hz 条件下的错误率, $F(1, 17) =$

4.86, $p < 0.05$ 。

结果分析表明纯音启动条件下目标元音/a/表现出显著的启动效应: 当纯音频率与/a/的 F2 频率匹配时, 被试识别元音/a/的速度显著快于频率不匹配的两种情形(1200Hz-/i/和 2150Hz-/a/)。复合音启动条件下元音/a/和/i/都表现出显著的启动效应: 反应时方面, 当复合音的频谱与目标元音的共振峰频率相匹配时, 元音的识别速度要显著快于频谱不匹配的条件; 错误率方面, 频谱匹配时被试识别目标元音的错误率也低于频谱不匹配的情形。因此, 非言语声音也能影响到汉语元音的知觉, 这与英语被试的实验结果一致(Wallace & Blumstein, 2009)。实验2的结果从汉语元音角度进一步支持了言语知觉的听觉理论, 当非言语声音的频谱和目标元音的频谱相匹配时(在纯音条件下甚至仅与目标元音单一的共振峰频率相匹配), 非言语声音的加工能够影响到目标元音的识别。而动觉理论认为言语知觉具有特殊性, 非言语声音不可能影响言语声音(不管是元音还是辅音)的加工, 这与实验结果相矛盾。

4 综合讨论

当前研究以汉语听者为被试探讨了非言语声音是否会影响汉语音段的知觉。实验1通过操纵纯音频率和辅音范畴连续体的频谱之间的关系, 考察了非言语声音对辅音知觉的影响。实验2通过操纵纯音和复合音的频率和目标元音的频谱之间的关系, 考察了非言语声音对元音知觉的影响。实验1的结果分析发现纯音影响到辅音范畴连续体的知觉曲线, 表现出典型的频谱对比效应; 实验2发现当纯音或复合音的频谱与目标元音的频谱相匹配时, 元音的识别表现出启动效应。两个实验结果一致表明当非言语声音的声学特征与言语声音的声学特征具有类似性时, 汉语听者音段(辅音和元音)的知觉也会受到非言语声音的影响, 这与英语国家被试的实验结果一致。因此非言语声音能够影响到言语声音的知觉似乎具有跨语言的一致性, 不受不同语言各具特色的语音特征的影响, 这种跨语言的一致性为言语知觉听觉理论提供了更多的证据: 言语声音的知觉不具有特殊性, 言语知觉的很多特征都是和非言语声音的知觉所共享的, 都需要大脑听觉皮层对声音信号中所包含的声学特征(时域、频域、强度等)进行分析和整合; 这种分析和整合机制在言语和非言语声音知觉中可能是重叠的, 只有共享相同的加工机制, 当非言语声音和言语声音有类

似的声学特征时, 非言语声音的加工才有影响言语声音加工的可能。而言语知觉动觉理论从根本上排除非言语声音加工影响言语知觉的可能性, 认为二者存在截然不同的加工机制, 这与两个实验的结果不一致。

实验1考察了纯音对辅音知觉的影响, 发现存在频谱对比效应。实验2考察了纯音和复合音对元音知觉的影响, 发现存在启动效应。实验1和实验2的整体范式类似, 都是一种启动范式。为什么两个实验发现了不同的效应模式? 有几种可能的原因。一是两个实验的整体范式虽然类似, 但在言语目标的设置和因变量的选择方面有些差异。实验1的言语声音是一个辅音范畴连续体, 因变量是被试的反应百分比。纯音语境主要影响连续体中间模糊刺激的知觉比率, 其对明确的终端刺激的影响并不显著。实验2的言语声音是明确的元音范畴, 因变量是被试的反应时。事实上, 英语被试的研究表明如果将言语声音设置为元音范畴连续体, 邻近声音为纯音或滑音, 也可以发现元音知觉中的频谱对比效应(Holt et al., 2000)。另一个可能的原因是元音和辅音知觉的具体机制存在一定的差异。元音一般由相对稳态的几个共振峰组成, 频谱特征相对稳定, 持续的时间也相对长; 辅音持续时间比较短, 频谱变化更加迅捷。已有研究表明听觉系统中可能存在平行的加工通道, 一个负责加工瞬时的信息, 对言语信号中快速变化的频谱特征敏感, 主要整合20~50ms范围内的信息; 另一个通道对言语中相对稳态的部分敏感, 负责整合150~300ms范围内的信息, 这两个通道可能分别负责了连续言语中辅音和元音的加工(Cohen & Grossberg, 1997; Boemio, Fromm, Braun, & Poeppel, 2005; Poeppel, Idsardi, & van Wassenhove, 2008)。由于元音频谱相对稳定, 元音知觉中可能更容易发现启动效应, 甚至启动音是单一的纯音。而辅音频谱时间短、变化快, 因此辅音知觉中可能更难发现启动效应。而另一方面, 辅音知觉中可能更容易发现频谱对比效应, 由于负责辅音加工的瞬时通道对言语信号中快速变化的频谱特征敏感, 而频谱对比效应主要由相邻声音的频谱存在对比(快速的频谱变化)所诱发。Takayama (2004)曾考察了辅音知觉中的启动效应。言语目标是 CV 音节/ba/和/ga/, 启动音是/ba/和/ga/的正弦波模拟音(用纯音模拟 CV 音节最初三个共振峰轨迹)。结果发现正弦波模拟音(被试难以将之感知为言语时)没有表现出启动效应。因此即使模拟了辅音三

个共振峰变化轨迹的复合音对辅音目标的知觉也没有启动效应。

启动效应和频谱对比效应从表面上看对言语目标知觉的影响似乎是相反的。启动效应中启动音提高了与之频谱相匹配的目标音的识别速度和正确率, 频谱对比效应中启动音提高了与之频谱相对立的目标音的反应百分比。但二者潜在的机制可能不像行为反应这么对立。启动效应在认知心理学的多个领域都已经得到证实, 在知觉中也普遍存在。刺激重复会导致行为反应成绩的提高, 但与之相应的神经活动会下降(重复抑制), 即下降的神经反应与提高的行为成绩相关, 神经适应是一种可能的解释 (Grill-Spector, Henson, & Martin, 2006; Bergerbest, Ghahremani, & Gabrieli, 2004)。神经适应的另一个重要意义可能是对比提高, 即适应提高了大脑皮层对新异刺激的反应强度, 行为上表现为对新异刺激的偏爱(Ulanovsky, Las, & Nelken, 2003; Ulanovsky, Las, Farkas, & Nelken, 2004)。如 Ulanovsky 等(2004)的研究表明对前面呈现的纯音的适应使得初级听觉皮层神经元对后面不同频率的纯音有不同强度的反应, 两个纯音频率差异越大, 神经反应越强, 差异越小, 适应越明显。Brosch 和 Scheich (2008)的研究也表明哺乳动物听觉皮层加工双纯音序列时同时存在重复抑制和对比提高的效应, 即与启动音频率相同或相近的目标音的神经反应会下降, 但与启动音频率波段不同的目标音(相差一个倍频)的神经反应会提高。因此, 启动效应和频谱对比效应可能都与神经适应有密切关系, 在神经反应上可能分别与重复抑制和对比提高相关。当目标刺激清晰明确时, 与之相同或相似的启动刺激可能会提高目标刺激的行为反应(启动效应), 特别是对于相对稳态的元音刺激, 尽管相应的神经反应会下降; 而当目标刺激相对模糊, 处于两个范畴中间时, 对与某一范畴相同或相似的启动刺激的适应可能会提高对立范畴的神经和行为反应(对比效应), 如对比辅音/ba/的 F2 起始频率相同的纯音的适应可能会使得偏向/da/端的 F2 区域有更强的神经激活, 从而导致模糊刺激有更多的/da/反应。

考察频谱构成相对简单的非言语声音对言语声音知觉的影响也为探讨言语知觉的内在机制提供了机会。当前研究中纯音或复合音的频谱都与目标言语的一部分频谱相重叠: 实验 1 的纯音频率与/ba/或/da/的 F2 过渡的起始频率相同; 实验 2 的纯音频率与目标元音/i/和/a/的 F2 频率一致, 复合音

由与元音/i/和/a/的 F1 和 F2 频率一致的两个纯音组成。两个实验也都发现纯音或复合音能够影响言语声音的知觉(对比效应或启动效应)。这表明言语声音知觉包括一个前言语的频谱特征分析阶段, 这个阶段在言语和非言语声音加工中是共享的; 且不同频率波段的信息在这个阶段的加工可能是平行的, 以至于分离的纯音频率能够影响言语目标中相对应的频率范围内信息的加工, 或者是重复抑制, 或者是对比提高, 初级听觉皮层在这个过程中可能承担关键角色。大量研究已经表明从听神经到初级听觉皮层的加工通路存在音频拓扑结构(tonotopy), 即声音信号加工时解析为不同的频率波段, 不同频段的信息由不同的神经元组进行加工, 且从听神经到初级听觉皮层保持对应(Eggermont, 2001; Schreiner & Winer, 2007; Young, 2008)。超越初级听觉皮层后, 不同频段的信息可能在更高级的次级听觉皮层内实现整合, 包括受到非言语语境影响的频段信息(其神经反应或提高或下降), 以形成完整的知觉。这与很多言语知觉的理论模型也是一致的, 这些模型也建议言语知觉存在声学特征分析和声学特征整合阶段(McClelland & Elman, 1986; Pisoni & Luce, 1987; Massaro, 1989)。因此, 当前研究中非言语声音对言语声音的影响可能发生在前言语的频谱分析阶段, 期间不同频段的信息加工是平行的、相对独立的。

5 结论

本研究采用启动范式考察了非言语声音是否会影响汉语听者的音段知觉。实验结果发现: (1) 纯音能够影响汉语被试对辅音范畴连续体的知觉, 影响方向表现为频谱对比效应; (2) 非言语声音(纯音和复合音)也能影响元音的知觉, 当非言语声音的频谱和元音的频谱相匹配时, 被试识别元音表现出启动效应。

参 考 文 献

- Bergerbest, D., Ghahremani, D. G., & Gabrieli, J. D. E. (2004). Neural correlates of auditory repetition priming: Reduced fMRI activation in the auditory cortex. *Journal of Cognitive Neuroscience*, 16, 966-977.
- Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8, 389-395.
- Brosch, M., & Scheich, H. (2008). Tone-sequence analysis in the auditory cortex of awake macaque monkeys. *Experimental Brain Research*, 184, 349-361.
- Coady, J. A., Kluender, K. R., & Rhode, W. S. (2003). Effects of contrast between onsets of speech and other complex

- spectra. *Journal of the Acoustical Society of America*, 114, 2225–2235.
- Cohen, M. A., & Grossberg, S. (1997). Parallel auditory filtering by sustained and transient channels separates coarticulated vowels and consonants. *IEEE Transactions on Speech and Audio Processing*, 5, 301–318.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149–179.
- Eggermont, J. J. (2001). Between sound and perception: Reviewing the search for a neural code. *Hearing Research*, 157, 1–42.
- Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, 68, 161–177.
- Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid co-articulatory influences on following stop consonants. *Perception & Psychophysics*, 48, 559–570.
- Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids consonants on stop identification in humans. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 877–888.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Science*, 10, 14–23.
- Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, 16, 305–312.
- Holt, L. L. (2006). Speech categorization in context: Joint effects of nonspeech and speech precursors. *Journal of the Acoustical Society of America*, 119, 4016–4026.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, 108, 710–722.
- Huang, J., & Holt, L. L. (2009). General perceptual contributions to lexical tone normalization. *Journal of the Acoustical Society of America*, 125, 3983–3994.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*, 68, 178–183.
- Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60, 602–619.
- Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407–412.
- Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition*, 24, 169–196.
- Massaro, D. W. (1989). Multiple book review of speech perception by ear and eye: A paradigm for psychological inquiry. *Behavioral and Brain Sciences*, 12, 741–755.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- Pisoni, D. B., & Luce, P. A. (1987). Acoustic phonetic representations in word recognition. *Cognition*, 25, 21–52.
- Poeppel, D., Idsardi, W., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Science*, 363, 1071–1086.
- Schreiner, C. E., & Winer, J. A. (2007). Auditory cortex mapmaking: Principles, projections, and plasticity. *Neuron*, 56, 356–365.
- Stephens, J. D. W., & Holt, L. L. (2003). Preceding phonetic context affects perception of non-speech. *Journal of the Acoustical Society of America*, 114, 3036–3039.
- Takayama, T. (2004). Priming effects in speech and nonspeech modes of perception. *Acoustical Science and Technology*, 25, 196–202.
- Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience*, 24, 10440–10453.
- Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, 6, 391–398.
- Viswanathan, N., Magnuson, J., & Fowler, C. A. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1005–1015.
- Wade, T., & Holt, L. L. (2005). Effects of later-occurring non-linguistic sounds on speech categorization. *Journal of the Acoustical Society of America*, 118, 1701–1710.
- Wallace, A. B., & Blumstein, S. E. (2009). Temporal integration in vowel perception. *Journal of the Acoustical Society of America*, 125, 1704–1711.
- Wu, T. P. (2006). *General Linguistics* (2nd ed). Beijing: Higher Education Press.
- [伍铁平. (2006). *普通语言学概要* (第2版). 北京: 高等教育出版社.]
- Young, E. D. (2008). Neural representation of spectral and temporal information in speech. *Philosophical Transactions of the Royal Society B: Biological Science*, 363, 923–945.

The Non-speech Sounds Affect the Perception of Speech Sounds in Chinese Listeners

LIU Wen-Li; YUE Guo-An

(Department of Social Psychology, Zhou Enlai School of Government, Nankai University, Tianjin 300071, China)

Abstract

A long-standing debate in the field of speech perception concerns whether specialized processing mechanisms are necessary to perceive speech sounds. The motor theory argues that speech perception is a special process and non-speech sounds don't affect the perception of speech sounds. The auditory theory suggests that speech perception can be understood in terms of general auditory process, which is shared with the perception of non-speech sounds. The findings from English subjects indicate that the processing of non-speech sounds affects the perception of speech sounds. Few studies have been administered in Chinese. The present study administered two experiments to examine whether the processing of non-speech sounds could affect the perception of speech segments in Chinese listeners.

In experiment 1, speech sounds were a continuum of synthesized consonant category ranging from /ba/ to /da/. Non-speech sounds were two sine wave tones, with frequency equal to the onset frequency of F2 of /ba/ and /da/, respectively. Following the two tones, the /ba/-/da/ series were presented with a 50ms ISI. Undergraduate participants were asked to identify the speech sounds. The results found that non-speech tones influenced identification of speech targets: when the frequency of tone was equal to F2 onset frequency of /ba/, participants were more likely to identify consonant series as /da/; when the frequency of tone was equal to F2 onset frequency of /da/, participants had more /ba/ responses, especially for ambiguous intermediate stimuli. The responses of participants showed a kind of spectral contrast effect.

In experiment 2, speech sounds were two synthesized vowels—/i/ and /a/. Non-speech sounds were two tones and two tone complexes. The frequency of two tones was equal to the F2 frequency of /i/ and /a/ respectively, and the complexes were composed of two tones at the frequencies of the first two formants of the vowel /i/ and /a/ respectively. The ISI between non-speech primes and speech targets was 25ms. Participants were asked to label vowel with /a/ or /i/ quickly. The results found that in the tone-prime condition the identification of vowel /a/ was significantly faster with the frequency-matched prime than with the frequency-clashed prime; in the complex-prime condition, the identification of vowels /a/ and /i/ both showed priming effect.

The results of experiments 1 and 2 confirmed that the processing of non-speech primes affected the perception of speech sounds (including consonants and vowels) in Chinese listeners. It provided cross-language evidence for the auditory theory of speech perception, and indicated that speech perception experiences a process of pre-lexical spectral analysis and the process is shared with the perception of non-speech sounds.

Key words speech perception; the auditory theory; the motor theory; spectral contrast effect; priming effect