·海外新译·

从语音到歌声的感知错觉转换

[美] 戴安娜·多伊奇，[美] 特雷弗·翰瑟，[美] 瑞秋·莱皮迪斯（著）[1]；
王博涵[2]，李小诺[3]（译）

（1. 加州大学 圣地亚哥分校，美国 加利福尼亚州；2. 华东师范大学 教育学部；
3. 上海音乐学院 音乐学系，上海 200031）

摘要： 文章针对听觉系统对音乐与语音之间的交互处理方式进行了研究。研究摘取一段语音中的一个短句，将其进行多种声音技术处理，通过两个实验探讨了语言与歌唱感知之间的错觉转换关系。实验一，首先将选取的语音短句进行两种处理，之后将原句及这两种处理呈现给被试，运用任务分级评价的测量方法获取被试对其感知的判断；经统计表明，在原句经过多次重复时，被试逐渐将其判断成歌唱而不是说话的错觉转换才会产生。实验二在实验一的基础上，设计四种刺激条件，收取被试代表听辨之后的复述呈现，运用声学技术和统计学运算，得出其与语音、歌唱等之间的数据关系；并由此对这一错觉转换的神经学基础、音乐与语言处理的脑机制及相关理论进行了深入讨论。

关键词： 语音（语言）；歌声（歌唱）；错觉转换；感知；脑机制

DOI：10. 3969/j. issn. 1008－7389. 2016. 04. 014

中图分类号： J60－051 **文献标识码：** A **文章编号：** 1008－7389（2016）04－0133－13

本文探讨了一个错觉现象，就是仅用多次重复的办法人们就会对一个言语短句（a spoken phrase）的感知从语音转化为歌声的声音。在实验Ⅰ中，被试重复听 10 遍短句，然后对它进行判断，在 1 分至 5 分分别为“完全是语音”和“完全是乐音”的五分制选项中打分。最初和最后的两次短句表述完全相同，当介于中间的几次表述也相同时，被试的判断结果呈现出从语音变化成歌声的稳定趋势，而当中间表述音调轻微调整或者音阶打乱顺序时，被试的判断就不会发生变化。在实验Ⅱ中，短句播放 1 次或 10 次，被试在最后一次听完后口头重复该短句。在听完一次表述之后，被试所重复的内容为说话的音调，然而在听完 10 次后，他们重复的内容是唱歌的音调。并且，听完 10 次后被试重复的内容更接近假设的音调旋律，而不是原语音短句。

收稿日期： 2016－09－07

基金项目： 教育部人文社科研究基金一般项目“音乐认知的理论与实践”（11YJA760039）；上海高校高原学科“艺术学理论－音乐艺术本原与当代音乐文化批判”。

作者简介： 1. [美] 戴安娜·多伊奇，[美] 特雷弗·翰瑟，[美] 瑞秋·莱皮迪斯：美国加州大学圣地亚哥分校心理学系，戴安娜·多伊奇（1938－）教授为本文通讯作者，主要从事音乐听觉错觉和音乐记忆等研究；2. 王博涵（1990－），女，华东师范大学教育学部硕士研究生，苏州工业园区青剑湖学校教师；3. 李小诺（1968－），女，上海音乐学院研究员，音乐学系书记、副系主任，华东师范大学心理学博士后。

原文 Diana Deutsch，Trevor Henthorn，and Rachael Lapidis. “Illusory transformation from speech to song” **发表于** *Acoustical Society of America*（**美国声学杂志**），Volume，129（4） April 2011.

一、引 言

最近，人们对音乐与语音之间的关系，尤其是听觉系统对这两种交流形式的处理方式方面的研究兴趣日渐高涨（cf. Zatorre *et al.*，2002；Koelsh *et al.*，2002；Koelsch and Siebel，2005；Zatorre and Gandour，2007；Schon *et al.*，2004；Peretz and Coltheart，2003；Patel，2008；Hyde et al. 2009；Deutsch，2010）。在探讨这类问题时，人们一般根据其听觉特点判断一个短句是语音还是基于这一语音声学特性的歌唱。语音是由频率急剧滑动构成的，这种滑动往往是振幅和频率的迅疾转换。相反，大部分歌声都是由大量离散的音高组成（即每个音高之间不是滑动联接的），每个音高都持续一定的时间，并且彼此以较小的音高距离联接。在现象学层面上讲，语音就像是一连串辅音和元音音色的快速变化呈现，在语音中音高轮廓是大致呈现的（至少在非声调语言中）。相反，歌唱听起来主要是一连串音高明确的音符（尽管也有辅音和元音），这些音符联接起来构成明确的音高关系和节奏型。然而，言语与非言语的物理特性的区分并不明确。有研究表明，特定的非言语声音可以在长期训练后（Remez *et al.*，1981，Mottonen *et al.*，2006）或在口头背景时（Shtyrov *et al.*，2005）以语音形式表达（be interpreted）。

人们普遍认为，语音与音乐可以通过声学特性来区分，这一观点在针对其知觉特征与神经学的基础研究中有所体现。对于语音来说，研究者关注的是诸如快速迁移的共振峰和发音的起始时间（Diehl et al.，2004）之类的特点，而对于音乐来说，研究者探究的是音高序列处理、乐器音色和节奏模式等问题（Stewart et al.，2006）。

使用不同物理特征的声音信号来独立研究乐音与语音是必要的，但是，如果发现它们的处理方式上有区别，那么其原因既有可能由于运用了不同的信号，也有可能由于信号是通过不同的神经通路处理的（Zatorre and Gandour，2007）。相反，本文描述和研究的是一种将口语短句转换成听起来像歌唱的（而不是说话的）感知错觉。产生这一错觉不需要信号的任何转变，不需要训练，不需要其它声音提供任何背景，仅仅是将短句重复数次后的结果。因此，本文提供了对于语音与音乐处理差异方面的见解，不需要再刻意借用不同的信号参数或者不同的声音背景。

该错觉研究首次发表在 Deutsch（2003）作为演示范例的光碟上。光碟上首先播放一个口语句子，接下来重复插入于句子中的一个短句。大多数人听到短句重复后会感觉自己听到的变成了歌唱的旋律，如图 1 所示。本文为该错觉的第一个正式研究和描述，并对其可能存在的潜在基理展开讨论。

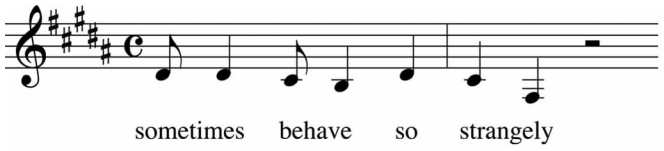

图 1　像唱歌一样的口语断句（感知为乐音的短句），来自 Deutsch（2003）。

本研究包括两个实验，实验Ⅰ运用任务分级评价的测量方法，对该错觉的研究进行规则性限制。只有对原句中的某一短句进行准确重复时，这种错觉才会出现；而当重复短句有轻微的（音高）移位或者音阶错乱时，这种错觉就不会出现。在实验Ⅱ中，通过分别让被试在听一次和听十次以后复述出所听到的短句，在细节上探讨了这一感知转换的特点。我们假设，在处理重复短句的过程中，音高联接形成短句的感知显著性不断增加，并且，重复短句发生感知变化形成了音调旋律。实验中的发现为我们的假设提供了证据。最终，我们对这一错觉的神经学基础提出假设，并对语音与音乐关系之间的普遍含义进行了讨论。

二、实验Ⅰ

（一）方法

1. 被试

实验共有54名至少有5年音乐训练经历的被试有偿参与。他们共分为3组，每组18人，每组应对不同的情况。第一组的被试（3名男性，15名女性）平均年龄21.7岁（年龄范围，18－33岁），平均接受10.2年的音乐训练（接受范围，6－14年）。第二组的被试（4名男性，14名女性）平均年龄22.4岁（年龄范围，18－29岁），平均接受10.6年的音乐训练（接受范围，6－15年）。第三组的被试（3名男性，15名女性）平均年龄20.3岁（年龄范围，18－28岁），平均接受10.0年的音乐训练（接受范围，6－14年）。被试中没有绝对音感者。经过听力测试鉴定，均有250Hz至6kHz的正常听力范围，并且他们对实验目的与错觉本质都只有非常天真单纯的认识（naïve concerning）。

2. 刺激编码（stimulus pattern）与实验程序

实验在一个安静的房间内进行。刺激编码取自多伊奇（Deutsch 2003）制作的光碟第22条的一个句子，表述如下 "The sounds as they appear to you are not only different from those that are really present，but they sometimes behave so strangely as to seem quite impossible."该句在各种条件（condition）下均出现，一个2300毫秒的停顿间隔之后，该句中的短句 "sometimes behave so strangely" 重复10遍，每遍中间的停顿间隔为2300毫秒。在每遍停顿时，被试要对所听短句状况作出判断，在一个1分至5分分别是 "完全是讲话" 和 "完全是歌唱" 的五分制选项中打分。

在所有条件中，句子第一次与最后一次播放的都是没有改变的原句，而中间插入的短句则是根据条件的不同而有所变化：在对照组环境下（In the untransformed condition），中间的短句不进行变化。在实验组环境下（In the transposed condition），中间的短句有轻微的音调变化，而其频率共振峰保持不变。按播放的顺序，中间呈现的断句转换的程度，分别是$+\frac{2}{3}$半音；$-1\frac{1}{3}$半音；$+1\frac{1}{3}$半音；$-\frac{2}{3}$半音；$+1\frac{1}{3}$半音；$-1\frac{1}{3}$半音；$+\frac{2}{3}$半音；$-\frac{2}{3}$半音。作为一个有序混合的整体时，中间的短句是不变调，但是其中的音节是打乱顺序播放的。短句包含7个音节（1 = some；2 = times；3 = be；4 = have；5 = so；6 = strange；7 = ly；），在中间重复时音节的播放顺序分别是6、4、3、2、5、7、1；7、5、4、1、3、2、6；1、3、5、7、6、2、4；3、6、2、5、7、1、4；2、6、1、7、4、3、5；4、7、1、3、5、2、6；6、1、5、3、2、4、7；2、5、4、3、7、1、6。

最后，所有被试完成一份调查其年龄与音乐训练的问卷。

3. 设备与软件

原始句子来自多伊奇（Deutsch，2003）制作的光碟第22条，上传至Power Mac G5计算机存为AIF文件，采样频率44.1kHz。使用BIAS PEAK PRO 4.01版本的软件包创制所有条件下使用的刺激物，并且创制出语句在原有有序混合整体情况下音节被打乱顺序后的音响。使用软件包PRAAT 4.5.06版本（Boersma and Weenink，2006）创制变换音调位置的短句，使用音高同步叠加法。随后重组信号保存至光碟。被试听到的声音使用DenonDCD－815光碟播放器，其输出声音经过Mackie CR 1604－VLZ混频器，通过两个DynaudioBM15A扩音器播放，被试听到的声音大约70dB的声压级（SPL）。

（二）结果

图2表示的是在第一遍、最后一遍，以及在上述三种条件下短句感知情况的平均比率。第一遍、最后一遍，不变音调位置、变化音调位置和打乱顺序重新组合三种情况下短语感知情况的平均比率。被试以

五分制形式对所听短语打分，1 分、5 分两端分别是“完全是说话”，“完全是歌唱”。我们可以看出，短句第一遍播放的感知结果为语音，但是最后一遍播放的感知结果取决于中间播放短句的特点。当中间短句音调没有改变时，最后一遍的结果为乐音；但是，如果中间短句音调变化了，最后短句的结果为语音，但是位置略偏向乐音；当中间短句顺序打乱时，最后结果为语音。

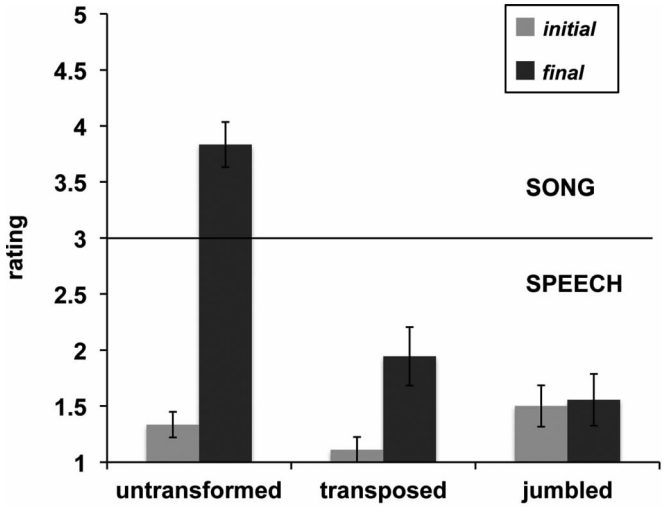

图 2 三种条件下短句感知情况的平均比率

为了对不同条件下被试做出的判断进行数据比较，我们做了一个 2 × 3 方差分析，将第一遍和最后一遍播放的短句作为被试内因子（within – subjects factor），将三种条件（不变音调位置、变化音调位置和打乱顺序重新组合）作为被试间因子（between – subjects factor）。第一遍与最后一遍播放间差异显著 [F（1，51） = 62. 817；p < 0. 001]，三种条件影响显著 [F（2，51） = 16. 965；p < 0. 001]，播放与条件之间的交互作用显著 [F（2，51） = 25. 593；p < 0. 001]。

在得到如上发现之后，我们做了更深一步的方差分析，把三种条件下、第一遍和最后一遍被试的判断分开来进行比较。在第一遍播放时，三种条件下被试的判断没有显著差异 [F（2，51） = 1. 912；p > 0. 05]。但是，在最后一遍播放时，中间插入短句的影响是非常明显的 [F（2，51） = 27. 317；p < 0. 001]。因此，我们在最后一遍播放的三种条件下展开比较。我们发现，被试在不变音调位置条件下和在变化音调位置条件下作出的判断是有明显差异的（p < 0. 001），并且不变音调位置和打乱顺序重新组合条件下作出的判断的差异也是具有显著性的（p < 0. 001）。而被试在变化音调位置条件下与在打乱顺序重新组合条件下的判断的差异不具有显著性（p > 0. 05）。

（三）讨论

本此实验中，被试对实验目的只是有着天真单纯的认识，挑选被试的条件也仅仅是具有至少五年的音乐训练经历，在实验中我们发现，被试在听到重复播放的短句后会认为听到的是歌唱而不是说话。但是，当中间短句有些轻微音调变化或者打乱音节顺序后，上述感知错觉就不会出现。因此，这一错觉的原因不会是因为该短句音高轮廓的重复或甚至不会是因为确定的旋律音程的重复，因为这些在变调中都是保持不变的。更进一步来讲，当中间的词组音调位置变化时，信号的时间维度关系没有改变，感知错觉也没有发生，所以原因也不会是短句确定的时间关系的重复。另外，当打乱音节顺序呈现时该错觉也没有发生，所以打乱音节顺序的重复也不会造成这一结果。因此，该错觉只是在短句不改变音调位置、并以相同音节顺序重复时出现。

实验Ⅱ运用行动任务（production task）的方法，在更深入的细节上研究这一错觉转换问题。在整句播放之后，其中被选出的短句播放 1 遍或 10 遍，要求被试在听完后将他们所听到的准确重复出来。随后我们对被试在两种条件下重复内容的差异进行分析。

该实验激发了两个假设。首先，与歌唱相反，语音的音高特点在感知上是不突出的，目前出现的感知错觉的一个重要特点是对音节音高的突显感知在重复过程中被持续增强，因此我们假设，在反复听到短句之后，被试的音高显著性的感知会增强，这就导致被试模仿的声音音高和原短句更接近，进而取得主体间的一致性。第二，一旦当音节听起来具有凸显的音高结构，它们就会发生感知变化，与貌似正确的旋律表述相一致，具体来说，如图 1 所示，被试对语句音高的假设。

三、实验Ⅱ

（一）方法

1. 被试

共有 31 名被试有偿参与到实验中来，分为 3 组。第一组有 11 名被试，平均年龄 23.8 岁（年龄范围，19－35 岁），平均接受 11.2 年的音乐训练（接受范围，5－15 年）。在加入实验之前，他们都已经听过句子以及随后重复的短句，并都表示他们听到的语音短句转化成为歌唱。第二组也有 11 名被试，平均年龄 18.9 岁（年龄范围，18－20 岁），平均接受 8.9 年的音乐训练（接受范围，6－12 年）。在加入实验之前，他们没有听过刺激音组。第三组有 9 名被试，平均年龄 19.2 岁（年龄范围，18－22 岁），平均接受了 8.0 年的音乐训练（接受范围，4－11 年）。所有被试不具备绝对音感；由听力测试测定，在 250Hz－6Hz 范围内全部听力正常，全部对实验目的与错觉本质都只有非常天真单纯的认识。

2. 实验条件与程序

实验共分四种情况。

第一，“有语言重复（repeat speech condition）”条件，其刺激音组与实验Ⅰ相同，其中的短句“sometimes behave so strangely”接连播放 10 遍，和实验Ⅰ唯一的区别就在于这里重复表述的停顿时间为 780ms；第二，“无语言重复（nonrepeat speech condition）”条件，这与“有语言重复”条件的刺激音响是相同，区别在于“无语言重复”条件中的短句只重复一遍；第三，“无歌唱重复”（nonrepeat song condition）条件中，刺激音响只包括由本文其中的一位作者（R. L.）在听过多次重复后演唱的该短句歌声般的录音。在这三种条件下，被试都被要求在听到刺激音组后将听到的声音准确重复三至四遍；第二遍被提取出来作为分析使用。第一组被试参与“有语言重复”条件的实验，第二组参与“无语言重复”条件，第三组参与“无语言重复”条件以及紧接其后的“无歌唱重复”条件的实验。

最后，以三种听辨条件为内容进行评价，我们提取“有语言重复”和“无语言重复”条件下两组共 22 名被试（每组 11 名）所重复的声音，并展示给第三组被试。这些短句以随机顺序播放，每个之间间隔 8 秒，跟随每次停顿及声音呈现期间，被试要求把对上述短句听起来是说话还是歌唱的判断填写到相对的表格中。

3. 仪器与软件

被试的这一从听辨到口头呈现的实验是在安静房间内以个人为单位单独进行的。此处我们用来播放刺激音组的仪器与实验Ⅰ所用相同。被试声音以 44.1kHz 采样频率收录在 Edirol R－1 24 位录音机上。我们使用 AKG C1000s 麦克风制作录音，麦克风放置在距离被试嘴部 8 英尺的位置。这些声音文件传输到 iMac 电脑上，存为 AIF 文件，采样频率 44.1K。随后，我们在每一个声音文件中，提取出被试对句子的第二遍复述，存为单独文件，使用软件包 BIAS PEAK PRO5.2. 版本将振幅做归一化处理（normalized for amplitude）。在 5 毫秒中间间隔时，我们使用软件包 PRAAT5.0.09 版本获得被试发音的 F0 估值（F0 estimates）（自相关法 autocorrelation method）。随后对每一个音乐文件，将 F0 估值平均分布在音阶上；这样一来，沿对数频率联接，就计算出一个短句的平均 F0 估值。需要说明的是，每一个短句都被分割成几个音节（some，times，be，have，so，strange 和 ly），每一个单独音节的 F0 估值分别计算。

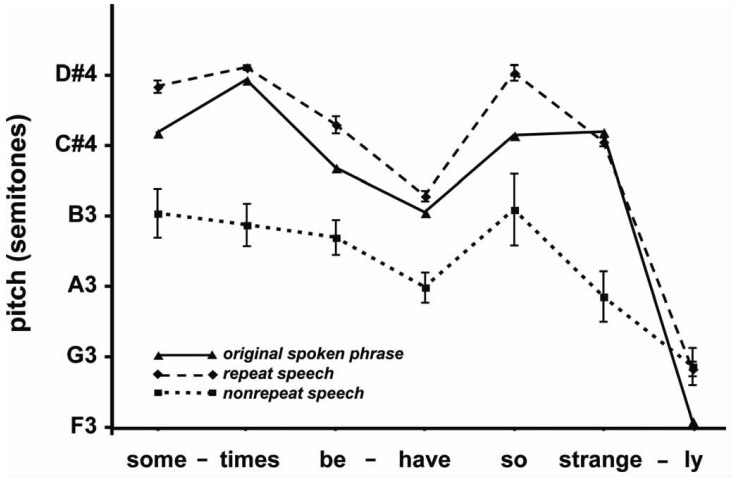

图 3　三角形代表原始短语中每个音节的平均 F0 估数

菱形代表所有被试在有语言重复条件下发音中每个音节的平均 F0 估数；正方形代表所有被试在无语言重复条件下发音中每个音节的平均 F0 估数。F3 = 174. 6 Hz；G3 = 196. 0 Hz；A3 = 220 Hz；B3 = 246. 9Hz；C#4 = 277. 2 Hz；D#4 = 311. 1 Hz。

（二）结果

被试的判断经过评价显示，无语言重复条件下的发音听起来像说话，而有语言重复条件下的发音听起来像歌唱。具体来说，在被试 198 次判断中上述结果的准确率为 97. 5%。该结果符合实验 I 的发现；在实验 I 中，被试认为原短句第 1 遍播放的听起来像说话，最后一遍听起来像歌唱。

我们具体分析了被试发音的音高模式，以此总结被试在反复听到原短句后的判断和发音变化。图 3 表现的是原短句在所有音节以及被试在有语言重复条件和在无语言重复条件下发声的平均 F0s。作为进一步的分析，图 4 展示了原始短句以及 4 名被试在有语言重复条件、4 名在无语言重复条件下的音高轨迹（pitch tracing）。这几条音高轨迹图概括了全部被试在几种条件下的发音。

我们得到两个发现，这两个发现来源于我们对音高突出性随重复而增强的假设。这体现出，相比无语言重复条件，在有语言重复条件下短句整体以及其中每个单独音节的平均音高在被试间都更持续，和原始短句也更接近。

首先，相比无语言重复条件，有语言重复条件下 F0 平均值的被试间方差是相当低的。以被试对完整短句发音的 F0 平均值为例，它的方差差异十分显著［F（10，10）= 5. 62，$p < 0.01$］。而对短句中每个音节展开比较时，这种模式依然持续：some，F（10，10）= 19. 72，$p < 0.0001$；times，F（10，10）= 69. 22，$p < 0.0001$；be，F（10，10）= 6. 2，$p < 0.01$；have，F（10，10）= 9. 71，$p < 0.001$；so，F（10，10）= 22. 68，$p < 0.0001$；strange，F（10，10）= 35. 76，$p < 0.0001$；ly，F（10，10）= 12. 71，$p < 0.001$。

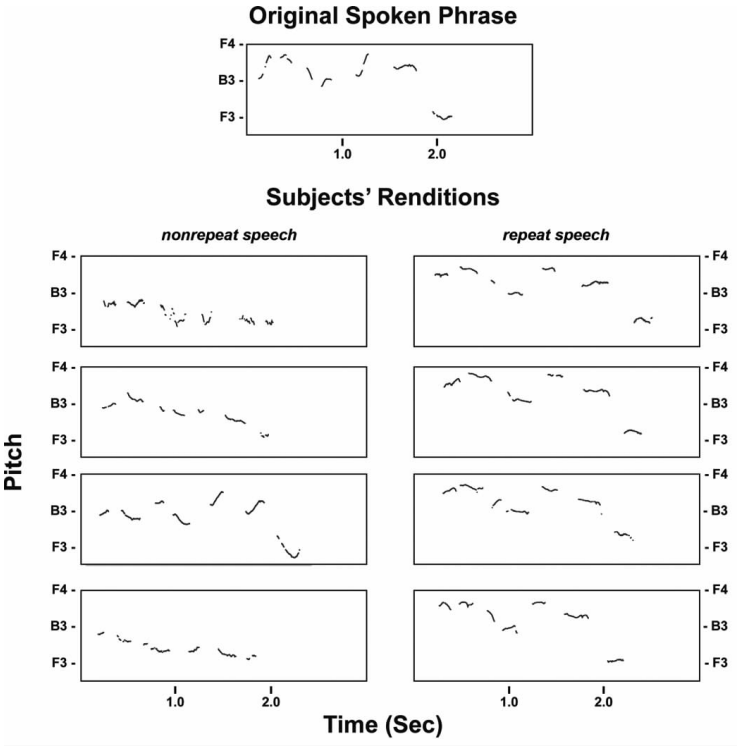

图 4　原始短语以及被试代表 4 名在有语言重复条件

4 名在无语言重复条件条件下的音高轨迹图。F3 = 174. 6 Hz；B3 = 246. 9Hz；F4 = 349. 2 Hz。

其次，我们发现，相比在无语言重复条件下的 F0 **平均值，有语言重复条件下的** F0 **平均值和原短句的** F0 **平均值要更接近。为了求出这一表现的数值，我们计算了每位被试短句发音的** F0 **平均值与原始短句** F0 **平均值的差数**（a difference score）。**我们使用独立样本** T **检验，并假定不对称样本方差，发现被试在听到** 10 **遍重复短句后的差数要明显低于听到** 1 **遍短句后的** [t（13. 34） = −4. 03；p < 0. 01]。**这一模式适用于单独拆分开的每一个音节，除了最后一个**：some：t（11. 01） = 5. 37，p < 0. 001；*times*：t（10. 29） = 7. 68，p < 0. 0001；*be*：t（13. 14） = 6. 46，p < 0. 0001；*have*：t（12. 04） = 6. 28，p < 0. 0001；*so*：t（10. 88） = 4. 23，p < 0. 01；*strange*：t（10. 73） = 6. 07，p < 0. 001；**对于** ly，**效果并不明显** [t（11. 19） = 1. 34；p = 0. 23] **呈现出相同的趋势**。

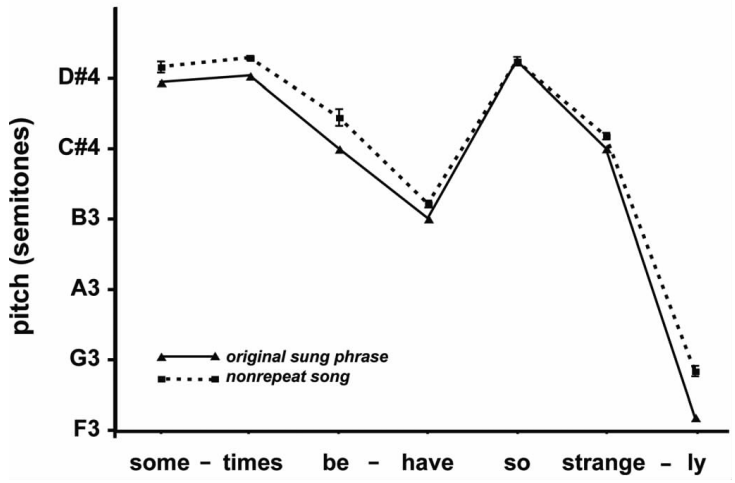

图 5　原始乐句中每个音节的 F0 **平均值（用三角形表示）以及全部被试在无歌唱重复条件下发音的每个音节的** F0 **平均值（用正方形代表）**。F3 = 174. 6 Hz；G3 = 196. 0 Hz；A3 = 220 Hz；B3 = 246. 9Hz；C#4 = 277. 2 Hz；D#4 = 311. 1 Hz。

为了确定我们发现的上述区别不是因为简单重复的结果，我们在无语言重复和无歌唱重复两种条件之间作了比较，这两种条件下被试都只接受到一次刺激音组。图 5 表现了原歌唱短句中所有音节的平均 F0 值以及无歌唱重复条件下所有被试发音的 F0 平均值。我们可以看出，被试在这种条件下的发音彼此间十分相近，和原歌唱短句也十分相近。作为进一步的阐释，图 6 展示了原歌唱短句以及有代表性的被试在无歌唱重复条件下的音高轨迹（这里的音高轨迹选自图 4 中无语言重复条件下被试的音高轨迹）。由此我们可以发现，相比无语言重复条件下被试的发音与原语音短句的关系，在无歌唱重复条件下的复述呈现在被试间更持续，并且与原歌唱短句更相近。

针对这两种条件下的复述呈现，我们作了两种形式的数据比较。

首先，在无歌唱重复条件下被试间方差的 F0 平均值要明显低于在无语言重复条件下。得出整句短句发音的 F0 平均值后，我们发现方差的差异十分显著 [F（10，10）= 7.39，$p < 0.01$]，当每个音节单独计算时也是这样：*some* [F（10，10）= 19.89，$p < 0.0001$]；*times* [F（10，10）= 97.66，$p < 0.0001$]；*be* [F（10，10）= 5.31，$p < 0.01$]；*have* [F（10，10）= 21.63，$p < 0.0001$]；*so* [F（10，10）= 60.51，$p < 0.0001$]；*strange* [F（10，10）= 66.06，$p < 0.0001$]；*ly* [F（10，10）= 17.74，$p < 0.0001$]。

其次，计算完整短句发音的 F0 平均值与原始歌唱短句 F0 平均值的差分（difference score），得出每个被试在无歌唱重复条件下的差分。我们使用相关样本 T 检验，发现在无歌唱重复条件下得出的差分要明显低于在无语言重复条件下的得出的 [t（10）= 3.31；$p < 0.01$]。当我们把七个音节分开来计算时，除了最后一个音节，在前六个音节测量中该模式依然适用：*somet*（10）= −4.16，$p < 0.01$；*timest*（10）= −7.56，$p < 0.0001$；*bet*（10）= −5.69，$p < 0.001$；*havet*（10）= −6.12，$p < 0.001$；*sot*（10）= −2.41，$p < 0.05$；*strange t*（10）= −6.6，$p < 0.0001$；对于 ly，效果并不明显 t（10）= −1.37，$p < 0.200$。

我们之前假设，反复聆听原语音短句会使得音高显著性增强，因而被试认为他们听到的更像歌唱而非说话，以上发现与假设相符。与无语言重复条件相比，被试在有语言重复条件下的复述呈现与原语句更接近，在被试内更一致。但是，被试在无歌唱重复条件下的复述呈现更接近原语句，更具有被试内一致性。

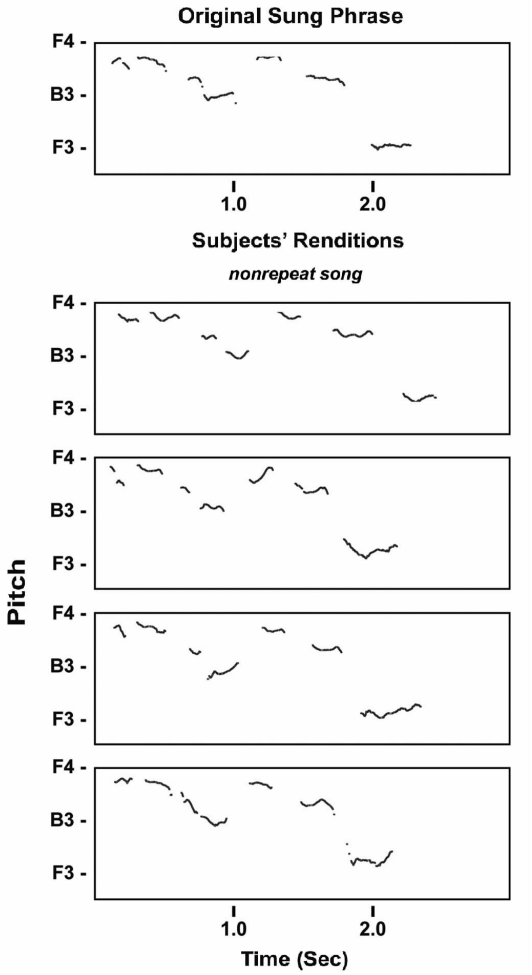

图6　原始歌唱短句、以及被试代表在无歌唱重复条件下的音高描轨迹图

这四名被试就是图4中无语言重复条件下选择的被试 F3 = 174.6 Hz；B3 = 246.9Hz；F4 = 349.2 Hz。

现在讨论我们在前文所预测的——一旦组成原语句的音节的音高显著性增强，人们就会在感知上将其转变以符合音调旋律。具体来说，我们预测，被试在有语言重复条件下的复述呈现符合图1模式，因此也就符合0，－2，－2，＋4，－2，－7（半音）的区间序列。因此我们假设，被试相比原口语短句中形成的音程结构区间序列，被试在在有语言重复条件下的复述呈现的音程结构区间更符合旋律性的表述。

为了验证假设，我们计算了原语音短句形成的六个旋律音程（melodic intervals），同样也计算了被试在有语言重复条件和无语言重复条件下复述呈现的旋律区间，基本方法为使用每个音节的F0平均值。接下来，针对每种条件，我们计算了六个音程的两组差分：（1）被试的复述呈现与原语音短句的音程之间的差分；（2）被试的复述呈现与基于假设的旋律表述音程之间的差分。随后我们对这两组差分作了数据上的比较。

计算结果可以在表1中看到。正如我们预期，被试在无语言重复条件下复述呈现的差分在两类比较中差异不显著。但是，与被试在有语言重复条件下复述呈现更接近的不是原始的语音短句，而是假设的旋律表述。具体来说，在有语言重复条件下，对于六个音程来说，被试的复述呈现与假设旋律表述之间的区别要小于被试复述呈现与原语音短句之间的区别。两类比较有显著的数据差异（$p < 0.016$，单尾检验，二项式测验）。该结果和我们假设相一致——被试的复述呈现在重复听短句后受到假设的音调旋律感知表述的影响。

表 1 被试复述呈现的音程与以下二者之间的差分（a）原语音短句，（b）基于假设的旋律表述（取被试平均数）。

	平均差（半音）	
	（a）原语音短句	（b）基于假设的旋律表述
有语言重复条件		
some 至 *times*	0.89	0.75
times 至 *be*	0.63	0.41
be 至 *have*	0.68	0.43
have 至 *so*	1.42	0.55
so 至 *strange*	2.04	0.51
strange 至 *ly*	1.41	0.72
无语言重复条件		
some 至 *times*	1.85	1.53
times 至 *be*	1.63	1.31
be 至 *have*	0.91	1.20
have 至 *so*	2.16	2.53
so 至 *strange*	2.35	1.68
strange 至 *ly*	5.06	4.06

表 2 被试复述呈现的音程与以下二者之间的差分（a）原歌唱短句，（b）基于假设的旋律表述（取被试平均数）。

	平均差（半音）	
	（a）原歌唱短句	（b）基于假设的旋律表述
无歌唱重复条件		
some 至 *times*	0.46	0.52
times 至 *be*	0.40	0.43
be 至 *have*	0.35	0.35
have 至 *so*	0.39	0.43
so 至 *strange*	0.40	0.31
strange 至 *ly*	0.82	0.36

表 2 显示了同样形式的计算结果，这里的计算基于原歌唱短句，比如无歌唱重复条件。我们可以发现，这里的差分数值很小，并且不会因为与原歌唱短句或与假设旋律表征之间的比较结果而有所变化。这一结果基于我们的假设——原歌唱短句本身就受到由歌者建构的旋律表征的强烈影响。

四、讨 论

当考虑到这种转换结果的可能原因时，我们注意到，语音的元音成分（vowel components）由泛音列（harmonic series）组成，因此一个人可能会希望能够清晰地感知到音高，即使已经不再重复了。但是和歌唱相反，语音的音高特点并不具备感知显著性。因此我们假设，在听到正常表达的语音时，潜藏在音高显著性下的神经元回路通过某种方式被抑制了，从而或许能够使听者将注意力集中在言语流的其它特点上，而这些特点对于语义来说是具有本质意义的，比如辅音和元音。同时我们假设，对短句的准确重复解除了对神经元回路的抑制，从而强化了听者感知到的音高显著性。考虑到我们的研究可能会涉及到脑领域，我们进行了脑成像研究，识别出颞叶双边区域，初级听皮层的前外侧，这一区域会对声音音高显著性优先反应（Patterson et al.，2002；Penagos et al.，2004；Schneider et al.，2005）。这就引起我们的推测：对短句的重复收听强化了这一脑区域的活性。

从语音短句转化为结构较好（well – formed）的音调旋律的感知转化过程一定是非常复杂的，包括多个层次的抽象活动（Deutsch，1999）。旋律音程在最低层次形成（Deutsch，1969；Demany and Ramos，

2005）。这一过程包括颞叶区域，这是一个进一步远离初级听皮层、侧重右半球发挥主要功能（Patterson et al.，2002；Hyde et al.，2008；Stewart et al.，2008）的过程。另外，听者要在感知上将语音短句转换成音调旋律，必须利用他们对熟悉音乐的长期记忆。这就需要他们将音高信息投射到曾强化学习过的范围内（Burns，1999）并运用我们声调系统中更深入的规则导向特点（Deutsch，1999；Deutsch and Feroe，1981；Lerdahl and Jackendoff，1983；Krumhansl，1990；Lerdahl，2001）因此需要音乐语法参与处理。因此，一旦短句的感知发生转化，更深层的脑区域就也参与其中。脑成像研究显示，两个半脑的大脑额叶区域，尤其是布洛卡区及其同族体，参与到音乐句法的处理中（Patel et al.，1998；Maess et al.，2001；Janata et al.，2002；Koelsch et al.，2002；Koelsch and Siebel，2005）。进一步来讲，人们已经发现，顶叶区域，尤其是左缘上回，参与到音乐曲调的短期记忆中（Schmithorst and Holland，2003；Vines et al.，2006；Koelsch et al.，2009），同样参与的还有其它周质层，比如叶上回（Janata et al.，2002；Koelsch et al.，2002；Schimithorst and Holland，2003；Warrier and Zatorre，2004）。因此，我们假设，当把语音短句转化感知为歌唱时，上述脑区域会发生深层活性活动。

应该指出的是，目前研究中的被试均有音乐训练经历。人们发现，相比没有音乐训练经历的听者，有过音乐训练经历的听者对音高结构更加敏感（Schneider et al.，2002；Thompson et al.，2004；Magne et al.，2006；Musacchia et al.，2007；Wong et al.，2007；Kraus et al.，2009；Hyde et al.，2009），所以，如果被试为未经过音乐训练的听者，可能无法得到目前实验所达到的清晰程度。

最后，目前的发现对有关乐音与语音感知基底的一般理论发生影响。正如 Diehl et al.（2004）和 Zatorre 与 Gandour（2007）所评论的那样，这一领域大部分研究都由两个恰好相反的理论所推动。“特定领域”（domain－specifically）理论认为，语言的声音与音乐的声音是由特定系统进行处理的，这一系统只针对某一声音，排斥其他声音（Liberman and Mattingly，1985；Peretz and Coltheart，2003）。与此相反，“线索基础”（cue－based）理论认为，判断一个刺激音是说话、音乐或是其它声音，主要依据它的声学特点，没有必要去确定某一机制是特定处理语音还是音乐的（Diehl et al.，2004；Zatorre et al.，2002）。目前我们的发现无法由上述两种强大的理论体系容纳在内，因为上述我们所用条件并无疑点，而该条件下同一刺激音组可以在某种情况下是被认为是说话，而在另外情况下被认为是音乐。因此有人提出相反的建议（Zatorre 和 Gandour 提出了类似的建议，2007），说话和音乐在很大程度上是由同一种神经通路处理的，但是在最终感知时，针对语音或音乐的特定神经元回路会发挥作用。

五、结 论

总之，我们描述并探索了一个新的感知转换效果，仅用多次重复的办法将人们对一个口语短句的感知从说话转化为歌唱。因为听者在重复收听后将所听到的表述出来，而他们所表述的内容音高发生了转变以形成一段结构良好的旋律。我们还需要做进一步的研究并总结产生这种的错觉口语短句的特点，记录其神经基础以及理解其发生原因是十分必要的。该研究将会对有关语音与歌唱的脑部机制方面提供相关信息。

致谢： 感谢 Adam Tierney，David Huber，和 Julian Parris 提出的讨论；感谢 Stefan Koelsch 和匿名审稿人对本文早期草稿给予的有益建议。

参考文献：

[1] Boersma，P.，and Weenink，D.（2006）. “Praat：Doing phonetics by computer（version 4. 5. 06），” http：//www. praat. org/（Last viewed December 8，2010）.

[2] Burns，E. M.（1999）. “Intervals，scales，and tuning，” in *The Psychology of Music*，2nd ed.，edited by D. Deutsch（Aca－

demic Press, New York), pp. 215 – 258.

[3] Demany, L., and Ramos, C. (2005). "On the binding of successive sounds: Perceiving shifts in nonperceived pitches," J. Acoust. Soc. Am. 117, 833 – 841.

[4] Deutsch, D. (2010). "Speaking in tones," Sci. Am. Mind 21, 36 – 43.

[5] Deutsch, D. (2003). *Phantom Words, and Other Curiosities* (Philomel Records, La Jolla) (compact disc; Track 22).

[6] Deutsch, D. (1999). "Processing of pitch combinations," in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic Press, New York), pp. 349 – 412.

[7] Deutsch, D. (1969). "Music recognition," Psychol. Rev. 76, 300 – 309.

[8] Deutsch, D., and Feroe, J. (1981). "The internal representation of pitch sequences in tonal music," Psychol. Rev. 88, 503 – 522.

[9] Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). "Speech perception," Annu. Rev. Psychol. 55, 149 – 179.

[10] Hyde, K. L., Peretz, I., and Zatorre, R. J. (2008). "Evidence for the role of the right auditory cortex in fine pitch resolution," Neuropsychologia 46, 632 – 639.

[11] Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., and Schlaug, G. (2009). "Musical training shapes structural brain development," J. Neurosci. 29, 3019 – 3025.

[12] Janata, P., Birk, J. L., Horn, J. D. V., Leman, M., Tillmann, B., and Bharucha, J. J. (2002). "The cortical topography of tonal structures underlying Western music" Science 298, 2167 – 2170.

[13] Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. D. (2002). "Bach speaks: A cortical 'language – network' serves the processing of music," Neuroimage 17, 956 – 966.

[14] Koelsch, S., and Siebel, W. A. (2005). "Towards a neural basis of music perception," Trends Cogn. Sci. 9, 578 – 584.

[15] Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Muller, K., and Gruber, O. (2009). "Functional architecture of verbal and tonal working memory: An fMRI study," Hum. Brain Mapp. 30, 859 – 873.

[16] Kraus, N., Skoe, E., Parbery – Clark, A., and Ashley, R. (2009). "Experience – induced malleability in neural encoding of *pitch, timbre and timing*," Ann. N. Y. Acad. Sci. 1169, 543 – 557.

[17] Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch* (Oxford University Press, New York), pp 1 – 318.

[18] Lerdahl, F. (2001). *Tonal Pitch Space* (Oxford University Press, New York), pp. 1 – 411.

[19] Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music* (MIT Press, Cambridge, MA), pp. 1 – 368.

[20] Liberman, A. M., and Mattingly, I. G. (1985). "The motor theory of speech perception revised," Cognition 21, 1 – 36.

[21] Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). "Musical syntax is processed in Broca's area: An MEG study," Nat. Neurosci. 4, 540 – 545.

[22] Magne, C., Schoön, D., and Besson, M. (2006). "Musician children detect pitch violations in both music and language better than nonmusician chil – dren: Behavioral and electrophysiological approaches," J. Cogn. Neurosci. 18, 199 – 211.

[23] Mottonen, R., Calvert, G. A., Jaaskelainen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., and Sams, M. (2006). "Perceiving identical sounds as speech or non – speech modulates activity in the left posterior superior tem – poral sulcus – ," Neuroimage 30, 563 – 569.

[24] Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). "Musicians have enhanced subcortical auditory and audiovisual processing of speech and music," Proc. Natl. Acad. Sci. U. S. A. 104, 15894 – 15898.

[25] Patel, A. D., (2008). *Music, Language, and the Brain* (Oxford University Press, Oxford), pp. 1 – 513.

[26] Patel, A. D., Peretz, I., Tramo, M. J., and Lebreque, R. (1998). "Processing prosodic and musical patterns: A neuropsychological investigation," Brain Lang. 61, 123 – 144.

[27] Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). "The processing of temporal pitch and melody information in au – ditory cortex," Neuron 36, 767 – 776.

[28] Penagos, H., Melcher, J. R., and Oxenham, A. J. (2004). "A neural represen – tation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging," J. Neurosci. 24, 6810 – 6815.

[29] Peretz, I., and Coltheart, M. (2003). "Modularity of music processing," Nat. Neurosci. 6, 688 – 691.

[30] Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). "Speech perception without traditional speech cues," Science 212, 947 - 949.

[31] Schmithorst, V. J., and Holland, S. K. (2003). "The effect of musical train - ing on music processing: A functional magnetic resonance imaging study in humans," Neurosci. Lett. 348, 65 - 68.

[32] Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). "Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians," Nat. Neurosci. 5, 688 - 694.

[33] Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., Dosch, H. G., Bleeck, S., Stippich, C., and Rupp, A. (2005). "Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch percep - tion preference," Nat. Neurosci. 8, 1241 - 1247.

[34] Schon, D., Magne, C., and Besson, M. (2004). "The music of speech: Music training facilitates pitch processing in both music and language," Psycho - physiology 41, 341 - 349.

[35] Shtyrov, Y., Pihko, E., and Pulvermuller, F. (2005). "Determinants of domi - nance: Is language laterality explained by physical or linguistic features of speech?" Neuroimage 27, 37 - 47.

[36] Stewart, L., Von Kriegstein, K., Warren, J. D., and Griffiths, T. D. (2006). "Music and the brain: Disorders of musical listening," Brain 129, 2533 - 2553

[37] Stewart, L., Overath, T., Warren, J. D., Foxton, J. M., and Griffiths, T. D. (2008). "fMRI evidence for a cortical hierarchy of pitch pattern proc - essing," PLoS ONE 3, e1470.

[38] Thompson, W. F., Schellenberg, E. G. and Husain, G. (2004). "Decoding speech prosody: Do music lessons help?" Emotion 4, 46 - 64.

[39] Vines, B. W., Schnider N. M., and Schlaug, G. (2006). "Testing for causality with transcranial direct current stimulation: Pitch memory and the left supramarginal gyrus," NeuroReport 17, 1047 - 1050.

[40] Warrier, C. M., and Zatorre, R. J. (2004). "Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task," Brain 127, 1616 - 1625.

[41] Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). "Musical experience shapes human brainstem encoding of linguistic pitch patterns," Nat. Neurosci. 10, 420 - 422.

[42] Zatorre, R. J., and Gandour, J. T. (2007). "Neural specializations for speech and pitch: Moving beyond the dichotomies," Philos. Trans. R. Soc. Lon - don Ser. B 2161, 1 - 18.

[43] Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). "Structure and function of auditory cortex: Music and speech," Trends Cogn. Sci. 6, 37 - 46.

【责任编辑：杨正君】