Cognition 127 (2013) 177–202

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Vocal imitation of song and speech

James T. Mantell *, Peter Q. Pfordresher

Department of Psychology, University at Buffalo, The State University of New York, United States

ARTICLE INFO

Article history:
Received 5 November 2010
Revised 7 December 2012
Accepted 21 December 2012
Available online 1 March 2013

Keywords:
Vocal imitation
Song
Speech
Modularity
Encapsulation
Domain specificity

ABSTRACT

We report four experiments that explored the cognitive bases of vocal imitation. Specifically, we investigated the accuracy with which normal individuals vocally imitated the pitch-time trajectories of spoken sentences and sung melodies, presented in their original form and with phonetic information removed. Overall, participants imitated melodies more accurately than sentences with respect to absolute pitch but not with respect to relative pitch or timing (overall duration). Notably, the presence of phonetic information facilitated imitation of both melodies and speech. Analyses of individual differences across studies suggested that the accuracy of imitating song predicts accuracy of imitating speech. Overall, these results do not accord with accounts of modular pitch processing that emphasize information encapsulation.

1. Introduction

Speech and song are forms of vocal communication. Each of these behaviors requires the coordinated use of the respiratory system, the larynx, and the vocal tract to provide variation in vocal intensity, pitch, and phonetic variation (Sundberg, 1999; Welch, 2005). In this context, it is not surprising that the distinction between speech and song is often blurred in practice, as in German sprechgesang and sprechstimme (sung speech and rhythmically heightened speech, respectively, which are utilized in certain operatic performances), and in the Japanese narrative forms of Nohgaki and Shinnai (Feld & Fox, 1994; List, 1963; Welch, 2005). Further, there is evidence to suggest that the perceptual identification of a vocal sequence as speech or song is plastic. Deutsch, Henthorn, and Lapidis (2011; see also Deutsch, Lapidis, & Henthorn, 2008; Falk & Rathcke, 2010) recently found that repeatedly presenting a spoken phrase causes that phrase to sound more like song; this suggests that context can influence the identification of a vocal sequence as speech or song. Yet, there are ways in which speech and song differ. For example, speech is a form of linguistic communication but song can serve as linguistic and/or musical communication. In everyday life, situational context underscores the distinction between speech and song. Individuals use speech when conversing but song is reserved for special occasions including celebration events, religious activities, and some social interactions (e.g., with young children). Some researchers have argued for shared processing of music and language (Koelsch, 2011; Patel, 2008; Sammler et al., 2009), some have emphasized that these modalities simultaneously present shared and distinct characteristics (Jackendoff, 2009; Jackendoff & Lerdahl, 2006), and some have suggested that music and language processing occur in separate cognitive modules (Peretz & Coltheart, 2003).

1.1. Modularity and music

The concept of modularity has been vigorously debated by scientists and philosophers since Fodor's (1983) landmark publication. In his essay, Fodor argued that mental input systems could be described as modules based on their possession of most or all of nine properties. For Fodor (1983, 2000), the single most important of these

* Corresponding author. Address: 207A Park Hall, North Campus, University at Buffalo, The State University of New York, Buffalo, NY 14226, United States. Tel.: +1 716 645 0225.

E-mail address: jtm29@buffalo.edu (J.T. Mantell).

0010-0277/$ - see front matter
http://dx.doi.org/10.1016/j.cognition.2012.12.008

characteristics is information encapsulation, the notion that not all information available to an organism informs operation of a modular system. Information encapsulation can be clearly imagined via a flowchart: when a researcher draws boxes to distinguish components of a processing system, it becomes clear that "only the inputs and outputs of functionally individuated systems can mediate their information exchanges" (Fodor, 1983, p. 87). Fodorian modularity gained early support by researchers. For example, Peretz and Morais (1989) argued that tonal encoding of pitch is accomplished by a cognitive processor that meets several of Fodor's modularity properties, including domain specificity (processing applies only to music), automaticity (operation is mandatory, given the input), and information encapsulation. However, several researchers (e.g., Pinker, 1997; Tooby & Cosmides, 1992, p. 113) explored the possibility that most or all of our mental faculties are evolutionarily adapted, domain specific, information processing modules; in so doing these researchers sought to expand the notion of modularity in ways that Fodor (1983) suggested were untenable. This approach, still under development today, is known as 'massive modularity' (Carruthers, 2006b).

Other researchers have eschewed Fodor's primary criterion, information encapsulation, in favor of another of Fodor's modularity characteristics, domain specificity. Coltheart (1999) proposed that a processing system is modular if it responds only to a particular class of stimuli (i.e., it is domain specific). However, Fodor (2000) rejected Coltheart's (1999) definition of modularity based on domain specificity (p. 113). For Fodor (1983), information encapsulation is "perhaps the most important aspect" (p. 37), "the essence" (p. 71), and "the key" (p. 98) to modularity. Other massive modularity theorists have dismissed the primacy of information encapsulation (Barrett & Kurzban, 2006, pp. 631–633; Carruthers, 2006a, pp. 12, 57–59). Barrett and Kurzban (2006) proposed a broad modularity based on functional specialization; their approach blends formal computationalism and evolutionary psychology. The authors assert that "Only information of certain types or formats will be processable by a specialized system... domain specificity is a necessary consequence of functional specialization" (p. 630).

But there is a problem with a modularity based only on domain specificity, and several researchers have recognized it (Besson & Schön, 2011; Fodor, 1983, 2000; Gibbs & Van Orden, 2010; Prinz, 2006). The problem is that declaring domain specificity as the essential quality of modularity trivializes the concept. In other words, a modularity based on specificity of input does not say anything useful about what modules do (see Fodor, 2000, p. 113; Prinz, 2006, p. 34). Instead, it posits a single characteristic as the definition of modularity and then points as "evidence" to the abundant cognitive systems that conform to this property. In line with Prinz's (2006) critique, Barrett and Kurzban appear to tacitly accept that most or all of the systems in the brain are modular (p. 630), writing "...whether an information-processing system "is or is not" modular is not useful. There is little doubt that different kinds of information are handled by different systems in the brain." This is probably what Fodor (1983, 2000) had in mind when he rejected domain specificity as the primary characteristic of a module. Today, modularity as a concept and a term continues to be debated (e.g., see the discussion between Carruthers, 2008 and Cowie, 2008; Machery, 2008; and Wilson, 2008), and it is clearly the case that neither massive modularity nor Fodorian modularity has been accepted by all researchers (Robbins, 2010).

The concept of cognitive modularity has not been decisively defined but there is considerable agreement that the specific information processing components that characterize modular processes must be information encapsulated, domain specific, or both. Thus, we have framed the empirical discussion within this paper around these two information processes. It is our hope that expanding knowledge of these two information processing characteristics will contribute to the debate on modularity in the cognitive processing of language and music. One modular model is particularly relevant to the current research because it makes empirical predictions about the performance and processing overlap between language and music. Peretz and Coltheart (2003) proposed a modular model of music processing based primarily on case studies of individuals with brain damage who together represent doubly dissociated music and language deficits. In their model, information from an initial acoustic analysis module is sent to specialized pitch, time, and speech modules. Separate modules facilitate the analysis of pitch, and of these distinct processors, one in particular—tonal encoding—is domain specific because it only accepts musical pitch information and likewise encapsulated to speech because phonological information cannot enter the module to influence pitch processing.[1] If a tonal encoding module exists as depicted in the model, it should handle tonality processing without access to phonological or linguistic information. Tonality is an informational property of music and not language; it is what determines why a single tone may sound good in one musical context and terrible in another (Krumhansl & Kessler, 1982). According to Patel (2008, p. 201), "At present there is no evidence of anything resembling scales or pitch hierarchies in speech melodies."

Evidence on the domain specificity and encapsulation of speech and song processing is mixed. Recent imaging research revealing substantial overlap in brain activations associated with speaking and singing (Callan et al., 2006; Saito, Ishii, Yagi, Tatsumi, & Mizusawa, 2006; Schön et al., 2010; Özdemir, Norton, & Schlaug, 2006) suggests that vocal processing may not be domain specific. However, these studies have also revealed non-overlapping areas with some exclusively right hemispheric activation for song tasks, indicating that there is something special about song. Moreover, the link between neural activations and modules is not entirely clear in part due to the fact that current imaging technology may not be capable of revealing the fine detail of adjacent neural networks (Peretz,

[1] It is conceivable that phonetic information could influence pitch processing, or that pitch information could influence phonetic processing, but neither of these possibilities is represented in the model in its current form. This is likely because no neuropsychology data has been collected to support such claims.