Brain & Language 118 (2011) 9–14

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Vocal pitch discrimination in the motor system

Alessandro D'Ausilio [a,b,*], Ilaria Bufalari [a], Paola Salmas [b], Pierpaolo Busan [a], Luciano Fadiga [a,b,*]

[a] DSBTA, Section of Human Physiology, University of Ferrara, Ferrara, Italy
[b] IIT, The Italian Institute of Technology, Genova, Italy

ARTICLE INFO

Article history:
Accepted 27 February 2011
Available online 31 March 2011

Keywords:
Motor cortex
Motor theories of perception
Speech comprehension
TMS
Motor resonance
Larynx motor area
Phonation
Vocal pitch

ABSTRACT

Speech production can be broadly separated into two distinct components: Phonation and Articulation. These two aspects require the efficient control of several phono-articulatory effectors. Speech is indeed generated by the vibration of the vocal-folds in the larynx (F0) followed by "filtering" by articulators, to select certain resonant frequencies out of that wave (F1, F2, F3, etc.). Recently it has been demonstrated that the motor representation of articulators (lips and tongue) participates in the discrimination of articulatory sounds (lips- and tongue-related speech sounds). Here we investigate whether the results obtained on articulatory sounds discrimination could be extended to phonation by applying a dual-pulse TMS protocol while subjects had to discriminate F0-shifted vocal utterances [a]. Stimulation over the larynx motor representation, compared to the control site (tongue/lips motor cortex), induced a reduction in RT for stimuli including a subtle pitch shift. We demonstrate that vocal pitch discrimination, in analogy with the articulatory component, requires the contribution of the motor system and that this effect is somatotopically organized.

1. Introduction

Sensori-motor integration requires specific brain circuits subserving coordinate transformation. The speech perception and production system is a particularly important and integrated instance of this kind. The intuition that articulatory goals may mediate perception was initially proposed long ago by the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) as well as the theory of analysis by synthesis (Stevens & Halle, 1967). The main reason for suggesting articulatory constraints for speech perception was the apparent lack of invariant cues in the acoustic signal to specify our phenomenal experience (Galantucci, Fowler, & Turvey, 2006). Generally speaking, both theories embrace a constructivist approach. Both of them indeed maintain that speech perception is mediated by constraints imposed by a sensori-motor model (via an inferential processes or internal simulation), mapping sensory input on the speech production system. A slightly different approach is the direct realist theory (Fowler, 1986). This theory proposes that, although there are no acoustic features that invariantly specify the units of speech, there are invariant properties in sensory stimulation that unambiguously specify the articulatory gestures, responsible for production, in a direct manner. This model, in fact, does not require any inferential process. According to this approach what we perceive is not sensory in nature but directly relates to the articulatory gesture (Callan, Callan, Gamez, Sato, & Kawato, 2010).

Both the constructivist and the direct realist approach stress the role that phono-articulatory gestures might have for both production and perception of speech. Indeed a central theoretical tenet of both approaches is that speech classification is ultimately the recognition of the phono-articulatory gestures produced by the sender. At the same time several competing theories have emerged in the past decades suggesting that an exclusive sensory analysis is sufficient for classification (Diehl, Lotto, & Holt, 2004). Both views may enumerate a large number of evidences accumulated in a 50 years long debate regarding the supremacy of a sensori-motor or a purely sensory account. However, recent trends of research may now be added to the discussion by offering a renewed support for the former view.

In fact, on the computational side a recent and growing trend in automatic speech recognition literature acknowledge the beneficial role of articulatory features in improving phoneme/word classification (King et al., 2007). Furthermore, a simple prediction is the recruitment of the motor system during speech perception tasks. Many neuroimaging and neurophysiological studies have indeed showed that motor and premotor cortices become active while listening to speech sounds (Binder, Liebenthal, Possing, Medler, & Ward, 2004; Boatman & Miglioretti, 2005; Callan, Callan, Honda, & Masaki, 2000; Callan, Jones, Callan, & Akahane-Yamada, 2004;

* Corresponding authors at: DSBTA – Section of Human Physiology, University of Ferrara, Via Fossato di Mortara, 17/19 – 44100 Ferrara, Italy. Fax: +39 0532 455242.
E-mail addresses: alessandro.dausilio@unife.it (A. D'Ausilio), luciano.fadiga@unife.it (L. Fadiga).

0093-934X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2011.02.007

Callan et al., 2003, 2006, 2010; Londei et al., 2007, 2009; Möttönen, Jarvelainen, Sams, & Hari, 2004; Nishitani & Hari, 2002; Ojanen et al., 2005; Pulvermüller, Shtyrov, & Ilmoniemi, 2003; Pulvermüller et al., 2006; Shahin, Bishop, & Miller, 2009; Skipper, Goldin-Meadow, Nusbaum, & Small, 2007; Skipper, Nusbaum, & Small, 2005; Wang, Sereno, Jongman, & Hirsch, 2003; Wilson & Iacoboni, 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004; and many others). Therefore, both computational and neuroimaging data seems to converge on the suggestion that motor computations/centers might take part in the speech classification process.

However, neuroimaging and most neurophysiological techniques use a correlational approach whereas transcranial Magnetic Stimulation (TMS) and direct cortical stimulation can directly demonstrate the activation of motor areas during perception of speech sounds. For instance, the discrimination of stop consonant voicing, place-of-articulation, syllable final vowel, steady-state vowels, pure tones, and frequency-modulated tones, is interfered by the direct electrical stimulation over superior temporal gyrus (Boatman, 2004). However, phoneme processing tasks such as phoneme monitoring, rhyming, and phoneme identification are interfered also by stimulation over the Inferior Frontal Gyrus. Interestingly, the inferior frontal gyrus might be recruited in syllable discrimination in patients with impaired comprehension under degraded listening conditions (Boatman & Miglioretti, 2005). On the other hand, studies employing TMS demonstrated that listening to tongue-produced speech sounds indeed activates the tongue motor representation (Fadiga, Craighero, Buccino, & Rizzolatti, 2002), whereas lips sounds activates the mouth motor area (Roy, Craighero, Fabbri-Destro, & Fadiga, 2008; Watkins, Strafella, & Paus, 2003). Furthermore TMS may be used to alter activity in motor centers and measure their causal involvement in perceptual tasks (D'Ausilio et al., 2009; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Möttönen & Watkins, 2009; Sato, Tremblay, & Gracco, 2009). In a recent study we applied online focal TMS to the lips or tongue motor regions while subjects discriminated tongue- or lips-produced phonemes ([b] and [p] vs. [d] vs. [t]). TMS stimulation facilitated the discrimination of concordant phonemes with respect to the discordant items, thus showing that this effect is somatotopic and causally associated to the motor system (D'Ausilio et al., 2009). However, all these studies explored the motor somatotopy for speech perception focusing only on the articulatory component and neglecting the phonatory side of speech perception. This is mainly due to the fact that Liberman's theory derives from the observation that no invariant cues in the acoustic speech signal are present (i.e. coarticulation phenomena). However, this is not the case for vocal pitch processing on individual vowel, with pitch information clearly present in the acoustic signal. At the same time phonation is intimately linked to articulation at the production level.

In fact, phono-articulatory gestures include several articulators that might be controlled independently, with very different functions. Phonation is the control of laryngeal musculature in order to have the vocal-folds produce a quasi-periodic vibration (F0). Articulation instead, is the control of the tongue, lips, jaw, and other speech organs in order to filter resonant frequencies out of the F0 component. A speech signal, such as a vowel, is thus created by the combination of phonation and articulation. Typically, the larynx produce a fundamental frequency that is locally weighted by the resonances caused by the vocal tract. The F0 is the primary cue for pitch perception whereas subsequent formants (F1, F2, F3, etc.) characterize the vowel category (Ghazanfar & Rendall, 2008). Therefore, a strong associative link must exist between phonation, articulation and respiration, at the level of motor control. Here we seek to demonstrate whether such associative link translate into the involvement of larynx motor cortex in a vocal pitch discrimination task.

In this study we investigate the contribution of the motor system in the perception of phonated signals. Since laryngeal function in speech production has a central importance specifically in determining vocal pitch, we investigated if the larynx representation of the motor cortex plays a role in vocal pitch discrimination. We used an approach similar to our previous TMS study on articulation (D'Ausilio et al., 2009), where we applied online TMS stimulation in order to experimentally alter activity in the motor system during a discrimination task. In the present study, subjects were required to discriminate whether pairs of vocal utterances [a] were the same or not. The two stimuli could parametrically differ in the F0 component height. We predict that if motor centers, and the larynx motor area specifically, have some role in discriminating vocal pitch, subjects' performance must be significantly different when TMS is applied.

2. Methods

2.1. Subjects

Ten healthy right-handed (measured with the Oldfield handedness questionnaire) subjects volunteered after giving informed consent and were compensated for their participation (mean age: 22.9, SD: 0.99; 7 females). None had any history of neurological, traumatic or psychiatric diseases and all of them had normal hearing. Procedures were approved by the local ethical committee in agreement with the Declaration of Helsinki.

2.2. Stimuli

Sound stimuli were recorded with a semi-professional microphone (AKG, C1000S) in a silent chamber and delivered to subjects via in-ear headphones. Stimuli were vocal recordings of a male actor producing a vowel sound [a]. The sound was cut and edited to last 600 ms with a roughly constant 105 Hz fundamental frequency. This vocal utterance was then pitch-shifted via Praat software (Version 5.1.05) by selectively moving the F0 formant. This procedure leaves unaltered all other formants and induces a perceptual pitch shift. F0 was initially shifted to ±3%, ±6%, ±9% and ±12%.

We then run a pilot experiment on nine subjects to select a subset of suitable pitch shifts. In fact, our aim was to avoid ceiling effects in behavioral recognition performance and to define intervals that were either very easy (close to 100%), very difficult (close to chance level) and an interval between them. Finally we had to consider that typical TMS experiments are relatively short and a limited number of magnetic pulses can be administered for safety reasons (Rossi et al., 2009). Therefore, we paired the original file with all other pitch-shifted samples and subjects had to decide whether the two differed or not. The ±3% interval lead to an average of ~61% of correct responses, ±6% allowed ~85%, ±9% enabled a very high performance of ~93% and ±12% with ~98% was the easiest discrimination. We thus decided to use only the ±3%, ±6% and ±9% (7 samples in total, 1 original and 6 pitch-shifted) since the ±12% contrast was too easy to discriminate and could not add much information to the experiment.

2.3. Task and procedure

Subjects were comfortably seated on a reclinable armchair and the audio stimuli were presented via headphones. Subject' responses were acquired by a custom-made response pad and both stimuli presentation and data recording were controlled by an E-Prime (Psychology Software Tools, Inc.) script. The correct synchronization between auditory stimuli and TMS occurrence was