Journal of Phonetics 53 (2015) 46–54

Contents lists available at ScienceDirect

Journal of Phonetics

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

Role of imitation in the emergence of phonological systems

Noël Nguyen [a,*], Véronique Delvaux [b,c]

[a] Aix Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France
[b] Institut de Recherche en Sciences et Technologies du Langage, Université de Mons, Mons, Belgium
[c] Fonds de la Recherche Scientifique, FNRS, Belgium

ARTICLE INFO

Article history:
Received 8 December 2014
Received in revised form
20 July 2015
Accepted 25 August 2015
Available online 26 September 2015

Keywords:
Imitation
Emergence
Phonological systems

ABSTRACT

The issue we address in this review paper is to what extent mutual adaptation plays a role in the emergence and evolution of phonological systems.

Adaptation to the interlocutor has been shown to take many forms and to embrace all the levels of spoken language, from adjustments in vocal intensity to changes in word forms over the course of a conversational exchange, as well as lexical and syntactic alignment across speakers, to name but a few examples. Phonetic convergence, that is, the tendency for two speakers engaged in a conversational exchange to sound more like each other, is one important aspect of between-speaker adaptation. Empirical evidence has recently accumulated that shows that phonetic convergence is a recurrent phenomenon in mature speakers. This phenomenon relies on sensory-motor abilities that infants may already possess at birth. Phonetic convergence affects the way in which both speakers speak after their interaction has ended, and may build up over long periods of time. It may also be a driving mechanism in the acquisition of the phonology and phonetics of a second language.

In this paper, (i) we outline the role of imitation in modern speech and language; (ii) we review the evidence provided by experimental and modeling studies for the potential role of imitation in the emergence and evolution of phonological systems; and (iii) we discuss how the resulting hypotheses could be tested in the framework provided by the multi-agent computational COSMO model.

1. Introduction

In their target paper for the present Journal of Phonetics Special Issue, Moulin-Frier, Diard, Schwartz, and Bessière (this volume) present an ambitious and far-reaching account of how phonological systems may emerge from a small number of general principles governing interactions between agents in a speech communication task. One key feature of the COSMO model relates to the mechanism that allows agents to converge towards a common set of wordlike forms for referring to external objects. In COSMO, these forms gradually arise through a sequence of pairwise communications that take place between agents assigned as speakers and agents assigned as listeners. Each communication entails the speaker's designating an object, and results in the adjustment of both the speaker's motor prototypes and listener's auditory prototypes associated with that object. Importantly, each instance of communication in COSMO is asymmetrical: the speaker speaks, the listener hears that speaker. Which agent is the speaker and which is the listener is subject to random changes across time. As a result, while the speaker's produced speech pattern has an impact on the auditory prototypes associated with the object in the listener, there is no reciprocal influence of the listener on the speaker and, more generally, no interaction between the two agents, inasmuch as an interaction entails an influence of both agents on each other. In that respect, COSMO may be seen as a non-interactional account of the emergence of a speech-mediated code.

COSMO's approach is at least in part in accord with other influential theoretical frameworks such as the one developed by Oudeyer (2005) for example. For Oudeyer, it is possible to simulate how a speech-mediated code may form in a society of agents without having to endow these agents with the capacity to interact with each other. According to an assumption made by Oudeyer, agents do not communicate, in the sense of intentionally conveying meaning to one another, and do not necessarily make the distinction between the speech signals they produce and those produced by others. This approach stands in sharp contrast with models of the emergence of phonological systems in which reciprocal influences of agents upon each other play a central role. In

* Corresponding author.

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.08.004

De Boer's (2000) model, for example, the formation of a phonological system is the collective and cumulative by-product of a large number of local, pairwise interactions between agents one of which aims to imitate the speech pattern the other agent has produced.

Adaptation to the interlocutor has been shown to take many forms and to embrace all the levels of spoken language, from adjustments in vocal intensity (Natale, 1975) to changes in word forms over the course of a conversational exchange (Fowler, 1988), as well as lexical and syntactic alignment across speakers (Garrod & Pickering, 2004), to name but a few examples. Phonetic convergence, that is, the tendency for two speakers engaged in a conversational exchange to sound more like each other, is one important aspect of between-speaker adaptation. Empirical evidence has recently accumulated that shows that phonetic convergence is a recurrent phenomenon in mature speakers (e.g., Babel, 2011). This phenomenon relies on sensory-motor abilities that infants may already possess at birth (Meltzoff & Moore, 1977; but see Jones, 2009). Phonetic convergence affects the way in which both speakers speak after their interaction has ended (Pardo, 2006), and may build up over long periods of time (Harrington, Palethorpe, & Watson, 2000; Pardo, Gibbons, Suppes, & Krauss, 2012). It may also be a driving mechanism in the acquisition of the phonology and phonetics of a second language (Lewandowski & Dogil, 2009; Sancier & Fowler, 1997).

The issue we propose to address in this paper is to what extent does between-speaker phonetic convergence play a role in the emergence and evolution of phonological systems. In Moulin-Frier and colleagues' COSMO model, as pointed out above, the focus is placed on reference, i.e. the setting up of a common speech-mediated code for designating external objects, to a much greater extent than on mutual adaptation. Because information passes in one way only (from the speaker to the listener), it appears to us that phonetic convergence can only occur in an indirect and delayed manner (and only in the sensory-motor version of the model), through the change that the speaker's produced speech pattern induces in the listener's auditory prototypes, as these are later brought into play by the listener's own speech production system when that listener becomes a speaker. By contrast, in other theoretical frameworks such as De Boer's (2000), imitation is consubstantial to the way in which speakers are assumed to interact with each other. In the following, we outline the role of imitation in modern speech and language (Section 2), we review the experimental evidence that may exist for a potential role of imitation in the emergence and evolution of phonological systems (Section 3), and we discuss how the resulting hypotheses could be tested in the framework provided by COSMO (Section 4).

2. Role of imitation in speech and language

In human beings, vocal imitation is a behavior that manifests itself over the lifespan. While some empirical studies have reported mimicry of sounds as early as 2–6 months of age (e.g. Gratier & Devouche, 2011; Kuhl & Meltzoff, 1996; Kokkinaki & Kugiumutzakis, 2000), a phenomenon that could be facilitated by audio-visual congruence in the model (Legerstee, 1990), other data suggest that the ability to imitate does not fully develop until the second year of life (Jones, 2009). Imitation involves a variety of perceptuo-motor, cognitive and social skills which makes it one of the “building blocks from which spoken language develops in typical development” (Charman, 2006:106). In particular, early vocal imitation has been found to positively correlate with later lexical development (e.g. Masur & Eichorst, 2002). The acquisition of L2 phonetics and phonology is also considered to be partly grounded on the ability to reproduce foreign speech sounds, so that individual differences in “speech imitation ability” (Reiterer, Hu, Sumathi, & Singh, 2013) or “phonetic compliance” (Delvaux, Huet, Piccaluga, & Harmegnies, 2014) may result in behavioral foreign accent differences in late L2 learners. Elderly speakers also show some ability to reproduce unfamiliar speech sounds (Delvaux, Huet, Piccaluga, & Harmegnies, 2013).

Regardless of its role in language development and second language acquisition, imitation of speech sounds is typically observed in mature speakers, i.e. in speakers who have reached the full mastery of their native language. Empirical evidences of phonetic convergence have accumulated over the last decade, whether in laboratory settings exposing a speaker to another individual's speech productions (Babel, 2010, 2012; Babel & Bulatov, 2012; Dufour & Nguyen, 2013; Delvaux & Soquet, 2007; Goldinger & Azuma, 2004; Gentilucci & Bernardis, 2007; Honorof, Weihing, & Fowler, 2011; Lelong & Bailly, 2011; Lelong, 2012; Mitterer & Ernestus, 2008; Miller, Sanchez, & Rosenblum, 2013; Namy, Nygaard, & Sauerteig, 2002; Nielsen, 2011; Nguyen, Dufour, & Brunellière, 2012; Shockley, Sabadini, & Fowler, 2004; Sato et al., 2013; Yu, Abrego-Collier, & Sonderegger, 2013), or in actual conversational interactions (Aubanel, 2011; Kim, Horton, & Bradlow, 2011; Pardo, 2006; Pardo, Cajori, & Krauss, 2010) but the exact role of phonetic convergence in speech and language remains an open question. Still, phonetic convergence may inform us on how speech sounds are dealt with, i.e. how they are structurally organized, cognitively processed and socially used.

First, phonetic convergence resides in the active exploitation of an effective sensory-motor link in processing speech sounds. Adapting to the interlocutor's speech initially requires the speaker to be able to make a cross-modal correspondence between the sounds he has just perceived (in the auditory domain) and the sounds he is about to produce (using motor commands), independently of the utterances they encode. As recalled by Moulin-Frier and colleagues (Section 1.2), recent findings on the existence of a mirror system in humans from which Broca's area may have evolved (Arbib, 2005a) are valuable in the search for a sensory-motor association system. Note, however, that while sound imitation can not be achieved in the absence of such a “parity” mechanism, the reverse is not true: the existence of a sensory-motor link does not imply per se that it will be exploited to support the imitation of the interlocutor's vocal productions. After all, mirror neurons have been first discovered in macaque monkeys, a species with poor imitation skills (Kopp, Wachsmuth, Bonaiuto, & Arbib, 2008; but see Kumashiro et al., 2003), and their potential role in supporting action understanding (through mental simulation) and imitation is still under debate (Hickok, 2010).

In COSMO, sensory-motor agents do not imitate each other, although they are attuned to their environment in that they update their motor and/or auditory prototypes following each deictic game. Phonetic convergence effects indicate that mature human