Role of imitation in the emergence of phonological systems

Noël Nguyen, Véronique Delvaux

▶ **To cite this version:**

Noël Nguyen, Véronique Delvaux. Role of imitation in the emergence of phonological systems. Journal of Phonetics, 2016, 53, pp.46-54. 10.1016/j.wocn.2015.08.004 . hal-01394207

HAL Id: hal-01394207
https://hal.science/hal-01394207

Submitted on 8 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*Manuscript
Click here to view linked References

Role of imitation in the emergence of phonological systems

Noël Nguyen[1] and Véronique Delvaux[2, 3]

1. Aix Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France
2. Institut de Recherche en Sciences et Technologies du Langage, Université de Mons, Mons, Belgium
3. Fonds de la Recherche Scientifique, FNRS, Belgium

Abstract

The issue we address in this review paper is to what extent does mutual adaptation play 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 modelling 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 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 et al., 2000; Pardo et al., 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. Kuhl & Meltzoff, 1996; Kokkinaki & Kugiumutzakis, 2000; Gratier & Devouche, 2011), 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 et al., 2013) or "phonetic compliance" (Delvaux et al., 2014) may result in behavioral foreign accent differences in late L2 learners. Elderly speakers also show some ability to reproduce unfamiliar speech sounds (Delvaux et al., 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 (Namy et al., 2002; Goldinger & Azuma, 2004; Shockley et al., 2004; Delvaux & Soquet, 2007; Nielsen, 2011; Babel, 2010, 2012 ; Babel & Bulatov, 2012; Gentilucci & Bernadis, 2007; Honorof et al., 2011; Miller et al., 2013; Mitterer & Ernestus, 2008; Yu et al., 2013; Dufour & Nguyen, 2013; Lelong & Bailly, 2011; Lelong, 2012; Nguyen et al., 2012; Sato et al., 2013), or in actual conversational interactions (Pardo, 2006; Pardo et al., 2010; Kim et al., 2011; Aubanel, 2011) 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 et al., 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 speakers do not only use their sensory-motor representations to ensure speech perception, but that they actively exploit them to drive their own productions (towards their interlocutor's) during a specific conversational interaction. Besides the need for these mental representations to be rich and flexible in structure, it is not yet fully understood how the routine imitative process affects functional aspects of speech-related representations, i.e. how they are built in, stored, retrieved, and updated (for a review, see Nielsen, 2011).

Second, phonetic convergence is only an aspect of the phenomenon of adaptation to the interlocutor that has been extensively documented over the years, in particular in the sociolinguistics literature in relation with the concepts of interpersonal accommodation (Giles et al., 1991; Gallois et al., 2005; Trudgill, 2008) and style-shifting behaviour (e.g. Eckert & Rickford, 2001). Well-known occurrences of adaptation to the interlocutor include infant-directed speech, clear speech and foreign-directed speech (Uther et al., 2007; Beckford-Wassink et al., 2007; Smiljanic & Bradlow, 2009).

Convergence between interlocutors concerns postures, mannerisms, facial expressions (the "chameleon effect": Chartrand & Bargh, 1999), as well as linguistic form at virtually all levels of linguistic hierarchy along a variety of parameters: prosody (pause frequency, pause duration (Gregory & Hoyt, 1982); overall intensity (Natale, 1975); fundamental frequency (Gregory, 1990); speaking rate (Webb, 1970)), lexicon (lexical choice: Brennan & Clark, 1996), syntax (syntactic structure: Pickering & Garrod, 2004; Branigan et al., 2000, 2007), and, of course, phonetics and phonology (vowel formants, vowel duration, VOT, etc.). Moreover, phonetic convergence arises in conversational interactions although there is no explicit instruction to imitate (e.g. Aubanel, 2011; Pardo, 2006), as well as in minimally interactional experimental settings involving exposure to another individual's speech (Delvaux & Soquet, 2007; Nielsen, 2011; Yu et al., 2013), including shadowed speech (Namy et al., 2002; Goldinger, 1998;

Shockley et al., 2004; Babel, 2010, 2012; Babel & Bulatov, 2012; Honorof et al., 2011; Miller et al., 2013; Mitterer & Ernestus, 2008). Besides, shadowers have been shown to align to a model's spoken words whether those words are presented auditorily or visually, i.e. in a lip-reading task (Gentilucci & Bernardis, 2007; Miller et al., 2010; Sanchez et al., 2010). Altogether, empirical evidence thus points out to a widespread behavior which relies on processes that may be consubstantial to human spoken interaction, or at least to the cognitive processing of speech whenever the perception of another's speech accompanies one's own productions.

The extensive evidence on linguistic convergence should not, however, obscure the fact that convergence effects are typically subtle ("perfect" imitation never occurs, actually the size of the effects is typically small) and variable across social and situational factors. It is not a purely automatic process. In particular, phonetic convergence has been shown to be mediated by (i) linguistic, system-internal, factors (e.g. American English speakers exhibited deferred, post-exposure, imitation of extended VOT but not of reduced VOT in word-initial voiceless stops, presumably because reducing VOT could hold the risk of confusion with voiced stops (Nielsen, 2011); see also Olmstead et al., 2013, on the effect of native language on patterns of VOT imitation); (ii) speaker-specific, situation-independent factors which are related with personality ("openness" (Yu et al., 2013), one of the Big 5 personality traits (Costa & McCrae, 1992)), cognitive functioning (in particular: the ability to allocate attention resources, more than their availability as measured by working memory capacity; Yu et al., 2013), and attitude (towards race, nationality and dialect; Babel, 2010, 2012); (iii) situation-specific factors, such as attitude towards the talker (Yu et al., 2013; Pardo et al., 2012), gender of the pair of talkers and talker's role in conversation (Pardo, 2006; Pardo et al., 2010) and the degree of social context embedded in the situation (Babel, 2012).

Altogether, the reviewed evidence suggests that phonetic convergence in conversational interactions is grounded on a low-level cognitive process involving a strong sensory-motor association, hence its pervasiveness in minimally interactive designs. However, phonetic convergence does not result from a purely reflexive process, since it is selective (and as such, requires selective attention to the fine-grained phonetic details of the imitated sounds and some kind of higher-level matching process between production and perception: Babel, 2011, 2012; see Studdert-Kennedy (2000a) for a view of imitation as a "purely formal process") and is typically modulated by a variety of psychological and social factors which are partly under the control of the talker.

Actually, communication accommodation theory first developed in the 1970s by Giles and colleagues (Giles, 1973; Giles et al., 2001; Gallois et al., 2005) claims that linguistic convergence and its opposite, divergence, are strategically used by speakers engaged in spoken interactions in order to respectively minimize or maximize the social distance with their interlocutors, so as to reinforce their own social identity. Whether driven by social identity matters and available for conscious manipulation (Eckert, 2001), or largely automatic and "deterministic" (Trudgill, 2004, 2008), linguistic accommodation is typically considered in sociolinguistics as one of the mechanisms potentially playing an influential role in channelling linguistic variation towards dialect formation, and ultimately language change. That is, local,

short-term accommodation in repeated conversational interactions could lead, at the individual level, to long-term accommodation, and at the community level to the propagation and diffusion of innovative variants (Auer & Hiskens, 2005), especially if they are associated with perceived prestige (Lev-Ari & Peperkamp, 2014), thus leading to sound change. Depending on the structure of the social network disseminating the innovations, accommodation could account for both dialect levelling and dialect formation (e.g. along highly inner-connected but loosely outer-connected subparts of the network).

Two types of empirical studies support this scenario. First, convergence effects elicited in the laboratory have been shown to persist several days after exposure (Goldinger & Azuma, 2004), while long-term accent changes can usually be retraced to extensive contact with the accommodated dialect (Harrington, 2006; Munro et al., 1999; Evans & Iverson, 2007), laying the foundation for a link between short-term and long-term accommodation within individuals. Second, recent sociolinguistic studies have used computer simulations on a social structure and game theory to test scenarios of language change (Ke et al., 2008; Fagyal et al., 2010; Mühlenbernd & Quinley, 2013). Fagyal and colleagues (Fagyal et al., 2010) have shown that in scale-free networks the process of an innovative variant's spread, stabilization, and gradual replacement by a new one can be modelled providing that (i) the network includes both strongly influential, high-connected agents who support the propagation of new variants, as well as peripheral, low-connected agents whose conservative variants may constitute a reservoir for further innovations, and (ii) agents select which variants to adopt in their one-to-one interactions based on their interlocutor's social status, defined as the agent's number of outgoing connections.

To sum up, imitation of speech sounds is a pervasive, multifaceted phenomenon in speech and language, which has been claimed to play a role in language development, second language acquisition, conversational interactions and language variation and change. It is not clear however, whether all the phenomena that may be regrouped under this label of "imitation of speech sounds" are actually different aspects of a single, unified process. In particular, phonetic convergence effects might result from automatic attunement to ambient speech via the constant, real-time updating of sensory-motor representations, as well as from controlled, deliberate imitation of one's interlocutor for social purposes. The former could be considered as one of the learning mechanisms supporting language development, that would remain in mature speakers as an automatic process of sensorimotor recalibration primarily facilitating speech *perception* through comparison between one's own productions and external speech inputs provided by others (Garnier et al, 2013). The latter could be described as a complex cognitive ability acquired through practice and learning, that would involve coordinating action and perception based on cognitive processes such as "conscious maintenance and recall of past auditory or vocal episodes, selective attention to subcomponents of experienced and produced sounds, identification of specific goals of reproducing certain acoustic features, and awareness of possible benefits that can be attained through successful sound reproduction" (Mercado et al., 2014:10).

3. **Role of imitation in the emergence of phonological systems**

In a COSMO deictic game, the listener does not respond to the speaker. An important consequence of this is that reciprocal adaptation of the speaker and listener to each other does

not occur. In particular, phonetic convergence, or imitation, of the two agents towards each other, is not deemed to significantly contribute to the building up of a phonological system. As underlined above, there is in that respect a marked difference between COSMO and other theoretical frameworks which, on the contrary, make imitation a driving force in the emergence of language and its phonological component.

The discovery of the mirror system in area F5 of the premotor cortex in the macaque's brain, and the role this system has been established to play in both the execution and recognition of grasping actions, has given rise to the idea that imitation, taken as a general sensory-motor ability, is one of the factors that, in the course of evolution, made our brains "ready for language" (Arbib, 2002, 2005b; Rizzolatti & Arbib, 1998). Relying on the fact that area F5 in the monkey brain is regarded as an homologue of Broca's area in the human brain, Arbib has proposed an evolutionary scenario in which human language arose from a primitive system of communication based on manual gesture. Although the phonological component of language is not its primary focus, two key features of this theoretical framework are particularly relevant to the issues discussed here. First, the mirror system is said to be at the origin of parity, defined as the fact that "what counts for the speaker must (approximately) count for the hearer" (Arbib, 2005b), a property central to human language and one of the basic requirements for speech communication to be successful in Moulin-Frier and colleagues' COSMO model. Second, in Arbib's proposed framework, imitation is not a behavior that merely mirrors that observed in another individual. Rather, it involves mapping the action to be imitated onto a repertoire of already available motor schemas that may then be recomposed, fractionated and/or tuned in a novel way. In other words, while it allows sets of actions to be passed on from one individual to another through observation, imitation is also conducive to the emergence of behavioral patterns that deviate from those they are based upon, and can therefore be a factor of innovation and change.

That imitation contributes to accounting for the building up of phonological systems is a central tenet of Studdert-Kennedy's approach to language (e.g., Studdert-Kennedy, 2000b, 2005). According to Studdert-Kennedy, human language developed from an early form of facial and vocal imitation in Homo erectus, with a mirror system akin to that found in the monkey as possible neural substrate. Like Arbib, Studdert-Kennedy holds that imitation is much more than mere behavioral echo, as the perceived action is assumed to undergo parsing into a set of elementary components that are then reassembled. It is because of the conjunction of these two major phenomena, the propensity shown by humans to imitate each other, and the decomposition/recomposition process through which imitation is performed, that language came to be endowed with the form it has today. In the course of its phylogenetic trajectory, the vocal apparatus has evolved towards differentiation into a set of functionally (semi-)independent articulatory organs, the lips, tongue tip, body and root, velum, and larynx. In Studdert-Kennedy's view, imitation operates through the mediation of this discrete and finite set of independently controllable organs, and this is what caused discreteness and combinatoriality to emerge in speech patterns. Speech development in children follows a course that to a certain extent parallels this phylogenetic scenario, and leads children to gradually confer a discrete, segmental structure to the quasi-continuous speech flow directed to them as they endeavor to imitate that

speech flow. As in Arbib's account, imitation is for Studdert-Kennedy a selective mechanism, which subjects a perceived action to conversion into a set of component gestures, and may generate novel patterns in the process.

Articulatory Phonology has also assumed that gestural structures may emerge as the by-product of imitation between speakers. For example, Browman and Goldstein (2000) conducted a computational experiment in which speakers were expected to conform to an "accommodation condition", which entailed their wanting to act like each other. As these simulated interactions in dyads of speakers developed, phase relationships between gestures arose that were both stable and shared across speakers, by virtue of what Browman and Goldstein regard as a self-organization process. In a follow-up to this work, Goldstein (2003; see also Goldstein & Fowler, 2003, and Goldstein et al., 2006) asked whether the establishment of gestural contrasts associated with a single vocal organ, can also be modeled as resulting from interactions between speakers in an imitation task. The results showed that, subject to certain conditions (see Goldstein, 2003, and Goldstein & Fowler, 2003, for detail), mutual imitation allowed computational agents to converge towards a partitioning of a gestural continuum into a number of discrete intervals.

As recalled above, imitation between speakers is a central mechanism in de Boer's (2000) computational model of the emergence of vowel systems. The model is made up of a population of agents that are each equipped with an articulatory synthesizer, a perceptual device that allows distances between vowels to be computed in the formant space, and a memory in which both the articulatory and acoustic properties associated with vowel prototypes are stored. These agents engage in a sequence of pairwise interactions that are referred to as imitation games and which, in essence, consists for one of two agents (the imitator) to imitate the vowel produced by the other agent (the initiator). Quite importantly, imitation is only approximate, in the sense that it is achieved through the filter of the vowel prototypes already available to the imitator. More specifically, the perceived vowel sound is mapped onto the closest imitator's vowel prototype, and is reproduced by the imitator as this vowel prototype. The initiator's and imitator's repertoires of vowel prototypes are then updated in a way which depends on whether the game is found successful or unsuccessful by both agents.

Thus, whether in Browman and Goldstein's (2000), Goldstein's (2003), or de Boer's (2000) modelling enterprise, the computational agents' goal is to sound more like each other. In COSMO's deictic games, the agents' goal is of a quite different nature: to use the same word forms when referring to the same objects. Agents interacting with each other may end up sounding more alike but only as an indirect and delayed consequence of this primary goal. The word form employed by Agent A as speaker to designate a particular object has an impact on the auditory prototype associated with that object in Agent B as listener. This may affect the way in which Agent B will in turn refer to the object when this agent is called on to assume the role of speaker and that object presents itself again. However, before this particular conjunction of events happens, and as the sequence of dyadic interactions unfolds, Agent B may be exposed as listener to a potentially high number of word forms produced by other speakers in relationship with that same object. Thus, Agent A's produced word form can influence Agent B's

own produced form only by percolation and through what may be a long series of dyadic interactions.

However, the imitation-based models, COSMO, and other models of the emergence of phonological systems such as Oudeyer's (2005) have one important commonality: they all attribute an essential role to the links between perception and action. Imitation obviously requires that speakers be able to match perceived speech forms to those they in turn produce. In their target paper, Moulin-Frier and colleagues convincingly demonstrate that the sensory-motor version of COSMO does a better job of simulating how phonological systems may emerge than both the motor-only and auditory-only versions. Thus, in our view, the difference between imitation-based models and sensory-motor models such as COSMO centers on how these models answer the following question: in the emergence of phonological systems, is it necessary to assume that adaptation between partners of an interaction is reciprocal and, more specifically, that each partner overtly or covertly endeavors to mimic the speech patterns produced by the other partner?

An important contribution towards answering this question can be found in the experimental work that, in parallel with the development of multi-agent computational models, has been recently conducted in an attempt to reproduce in the laboratory the conditions that may have presided over the emergence of language (Scott-Phillips & Kirby, 2010). Most of this work has been concerned with the emergence of graphical communication systems, and there are yet very few experimental studies on how acoustic communication systems may have formed. In a recent study, however, de Boer and Verhoef (2012, see also Verhoef et al., 2011) examined how a repertoire of vowel-like sounds gradually took shape as groups of participants were successively asked to reproduce these sounds and transmit them to the next group, along what was made to look like a generational chain. The vowel sounds were associated with as many visual forms constructed by systematic combinations of different shapes and colors, and were originally synthesized from random patterns in the F1-F2 formant space. Within each generational group, participants learned to reproduce the vowel sound they heard in association with a particular visual form, by drawing a trajectory on a computer screen, which was then converted into a time series of F1 and F2 frequency values themselves passed on to a speech synthesizer. The authors' hypothesis was that, as the correspondences between vowel sounds and visual forms were transmitted across generational groups, a combinatorial structure would gradually emerge within the vowel sounds, on par with increased learnability. The results did not conform to these predictions, possibly because the learning task was too difficult (see Verhoef et al., 2011, for detailed discussion).

In yet a more recent work, Verhoef et al. (2014) used another task which, instead of vowel sounds reproduced through a graphical medium, involved learning an artificial whistle language. Verhoef et al. (2014) had four parallel chains of ten participants memorize and reproduce twelve whistle sounds of a variety of forms (as freely produced and recorded by different people prior to the experiment), using a slide whistle. Directly relevant to the present paper is the fact that the whistle sounds did not refer to anything, and that participants were simply instructed to imitate these sounds as best they could. The experiment showed that, in the course of being

transmitted from one participant to the following one along each chain, the whistle sounds came to be broken down into smaller components that were then reused in combination with each other. Quantitative analyses revealed that the sounds became both increasingly structured and increasingly learnable. According to Verhoef et al. (2014), the cognitive and sensory-motor constraints associated with the transmission task caused a combinatorial structure to form because it made the sounds easier to memorize and reproduce. Thus, this study demonstrates that in an experimental setting where each individual is asked to learn through imitation a repertoire of sounds produced by another individual, along a linear transmission chain, combinatorial properties emerge in that repertoire of sounds that resemble those found in phonological systems.

The linear transmission chain paradigm employed by Verhoef et al. (2014) makes it possible to experimentally pinpoint the role of learning by imitation in the structuration of a communication system as information is transmitted in a one-way fashion, from the sender to the receiver, with no information passing back from the receiver to the sender. Other studies (e.g., Galantucci, 2005; Garrod et al., 2010) have resorted to experimental paradigms in which participants engage in pairwise interactions. These studies focus on graphical communication systems, and whether their findings can be extended to acoustic communication systems remains to be empirically determined. What they show, however, is that pairwise interactions have a significant influence on how a communication system takes shape. Thus, Garrod et al. (2010) showed that pairwise interactions, in a Pictionary drawing task, led to the development of a simpler, more abstract, and more easily identifiable set of graphical signs compared with the one obtained at the outset of a one-way, linear transmission chain (referred to as a diffusion chain). Similarity of the drawings also increased to a larger extent within interacting pairs than between members of the diffusion chain. These results suggest that the feedback signals that participants were allowed to exchange in pairwise interactions made it easier for them to converge towards a set of common signs. This is consistent with the finding that verbal information is more accurately transmitted along a linear transmission chain when participants in adjacent positions along that chain are permitted to interact with each other than in the absence of such interactions (Tan & Fay, 2011).

It is of course difficult to reproduce the conditions conducive to the emergence of a communication system in a laboratory setting, in particular because this usually involves calling on adult human participants whose already acquired language knowledge can influence their behavior. While this caveat must be borne in mind, the experimental evidence now available suggests that both learning by imitation and pairwise interactions have an important role to play in the process that allow human participants to jointly build up a repertoire of acoustic or graphical forms endowed with combinatorial properties.

4. Discussion

The work we have reviewed in Section 2 strongly suggests that imitation between individuals is pervasive in spoken language, at every stage of life from infants to mature speakers. It is believed to play an important role in both speech development and L2 acquisition. Phonetic

convergence from one speaker towards another speaker's voice has been found to occur to a limited yet systematic extent in both non-interactive and interactive settings. In our view, the evidence now available suggests that imitation in speech may in fact be constituted by the combination of a low-level component with a high-level one. The low-level component involves sensory-motor integration processes that allow speakers to establish a correspondence between the speech signal they are exposed to and their own repertoire of speech motor commands. It is triggered in an automatic manner and represents one instance of the links that are more generally formed between perception and action. The high-level component operates under the control of a variety of linguistic factors (phonetic convergence can occur inasmuch as phonological contrasts are preserved), individual, situational and social factors. At this high level, imitation is a strategy deliberately employed by speakers to modulate the position they occupy with respect to each other in a social space, as their interaction unfolds. It is one of the ways that speakers have to indicate to each other to what extent they are like each other, and to set up a conversational common ground. It may be hypothesized that the strategic, socially-driven component of imitation develops during ontogenesis on top of the sensory-motor component as children acquire the skills that are required for them to engage in social interactions.

As we have turned to the potential role of imitation in the emergence of phonological systems (Section 3), we have found that there is compelling evidence for the fact that imitation, far from being a mere behavioral echo, is selective, compositional, and may lead to innovation and change. It involves the decomposition of the perceived speech signal into smaller units that are then recombined with each other by the listener's own speech production system. Both multi-agent computational models, such as those developed by Browman and Goldstein (2000), Goldstein (2003), and de Boer (2000), and experimental studies suggest that with such characteristics, imitation may provide a strong contribution to the building up of a phonological system.

In an in-depth theoretical and methodological discussion, Oudeyer (2005) advocates what we would refer to as a minimalist approach to the emergence of language, which causes him to attribute no prior communicative or social skills to the agents in his computational model. In Oudeyer's view, his model is in that respect sharply different from one like de Boer's (2000), in which agents aim to imitate each other and are also assumed to jointly establish whether imitation has been successful or not through non-verbal feedback. Because this presupposes that agents already have basic communicative and social skills, Oudeyer claims that de Boer's model offers an account of the cultural evolution of language rather than of the emergence of language. It may be debated, however, whether the use of non-verbal feedback or that of pre-established communicative knowledge necessarily imply that agents already share a full-fledged linguistic system. In this regard, note that, in COSMO, it is assumed that each agent has an internal model of the entire communication situation, although COSMO presents itself as a model of the emergence rather than the evolution of phonological systems. What Oudeyer demonstrates, in quite an elegant and parsimonious manner, is that it is *possible*, for a society of artificial agents, through a self-organization process, to develop a stable and shared repertoire of speech sounds, even though this repertoire is entirely devoid of referential or social

function. Whether this is how language has actually formed is, however, an empirical question. In our own view, a model of the emergence of phonological systems does not necessarily have to be based on the assumption that agents start with a minimal set of cognitive, communicative and social skills. In what follows, we focus on the difference between COSMO and imitation-based models.

One major characteristic of COSMO is the emphasis put on deixis and reference. Note that, in the current version of COSMO, there are no actual objects in the agents' virtual world and which agents would have the capacity to detect, recognize and jointly designate. In fact, objects do not seem to have an independent reality, outside the speech code itself, and look functionally identical to abstract labels for sets of motor and auditory patterns. However, this is due to the way in which COSMO has been implemented so far, rather than a feature of COSMO itself, whose future extension towards a yet more comprehensive model that will include genuine deictic and reference mechanisms can easily be envisioned. In imitation-based models, by contrast, it is in our view the social dimension of inter-individual interactions that is dominant. Whether in Browman and Goldstein's (2000), Goldstein's (2003), or de Boer's (2000) model, agents seek to resemble each other to a greater extent by sounding more like each other. It is unclear, however, upon which grounds one should consider in an a priori manner that either the referential or social dimension of language is predominant. We suggest that computational modeling should aim to investigate what the respective contribution of either dimension may be to the emergence of phonology, as opposed to taking this contribution or absence of contribution as a given.

COSMO provides a computational model that allows studying the emergence of phonological systems as a self-organization process arising from local deictic games between agents. In particular, the architecture of the computational model, which draws from a full internalization of the communication situation, allows for an explicit testing of competing hypotheses about speech-related representations in line with motor vs. auditory vs. sensory-motor theories of speech communication. In our opinion, a valuable extension of the work carried out by Moulin-Frier and collaborators would involve using the COSMO framework to study the potential role of *imitation* in the emergence of phonological systems. To our knowledge, no computational framework has yet explicitly assessed how varying the rules of the interaction games played by agents may affect the sound systems emerging from these games.

As stated above, phonetic convergence effects in modern speech may be considered as resulting from a variety of cognitive abilities and associated behaviors ranging from loose attunement to ambient speech in minimally interactive settings to deliberate imitation allowing for reciprocal adaptation between communicating partners. As a consequence, testing for the potential role of imitation in the emergence of phonological systems may be best achieved by considering several steps on a "minimally-to-fully-imitative" continuum. A first step on this continuum could be implemented by varying the modalities of the procedure for representations' updating which is currently embedded in COSMO. For example, one could introduce a coupled "recency-decay" function allowing for recently experienced realizations to be favoured in the

directly following games while older, for long inactive, realizations would gradually see their probability to be used again tend to zero.

A second step would consist in simulating a minimal interaction in that, for each naming game of a given object o_i,, there would be a two-stage deictic game (keeping C=1 throughout the game): (i) agent A is the speaker; agent B is the listener; both agents update their knowledge respective to their experience; (ii) agent B is the speaker; agent A is the listener; both agents update their knowledge respective to their experience. In other words, agents listen to each other and adapt their auditory prototypes accordingly, but their behavior are still fully supervised by the shared-attention mechanism. A third step would introduce imitation of the interlocutor along with actual perception behavior. At stage 1, the listener (agent B) uses its auditory prototypes to perceive the speaker's production (i.e. to associate it with an object o_i), and in return (stage 2), it performs a speech production based on its closest motor prototype for that object o_i, then updates its motor knowledge accordingly. Agent A then updates its own auditory knowledge only if its inference of the object o_i based on the reproduction from agent B (at stage 2) matches the object agent A has itself initially named (at stage 1). These rules of interaction may be compared to the simulation used by Goldstein and colleagues (Goldstein, 2003; Goldstein & Fowler, 2003), to the extent that agents update their "knowledge" based on an internal assessment of the success of the interaction game. A fourth step would include feedback (in line with the simulations carried out by de Boer, 2000), so that agent A would pass on the result of its assessment of the imitation to agent B (C=1 vs. C=0), which would only then (i.e., at stage 2) update its own motor representations accordingly.

Moreover, in our opinion, valuable insight on the role of imitation in shaping phonological systems could be gained if its effects were assessed in heterogeneous populations of agents. For example, one may carry out simulations in which, due to inter-individual variation, some agents are more prone to convergence than others, and some are even prone to divergence. In the third-step game outlined above, it would mean that agent B would perform the production task by selecting *any* of its motor prototypes associated with the object o_i, or even by selecting the prototype which is the furthest away from (vs. the closest to) the initial production of agent A. The proportion of 'divergent' vs. 'indifferent' vs. 'convergent' agents could be a variable manipulated in the simulations.

On a more structural level, a comparison between minimally-to-fully imitative interaction types may be even more insightful if simulations were carried out on a population of agents exhibiting horizontal as well as vertical structuring. By 'horizontal structuring', we refer here to networks of variously interconnected agents aimed at simulating social structure. In these social networks, the probability of occurrence of specific one-to-one interactions is modulated by the structure of the network itself, and both high-connected and low-connected agents may play an influential role on the specifics of the emerging shared system (Fagyal et al., 2010). As to 'vertical structuring', it could be implemented by gradually introducing newly-initialized agents ('newborns') as well as removing mature agents ('deceased') as simulations accumulate, so as to ensure generational replacement in the long run.

Admittedly, both horizontal and vertical structuring of the agents' population may be more related to the propagation and transmission of the specifics of a shared repertoire of sounds than to its proper building, i.e. it may enlighten us on the cultural evolution, rather than on the emergence, of phonological systems. However, since culture preceded (modern: symbolic, compositional) language in human evolution, it may be unreasonable to simulate the emergence of phonological systems while ignoring the context in which this emergence process took place. In that sense, it may not be considered as an asset of the computational models studying the emergence of phonology, that the simulations carried out with these models invariably lead to a shared sound systems which, once established, is absolutely stable over time. A working hypothesis is that imitative interaction games, together with network structuring, although not absolutely necessary for the emergence of stable shared repertoire of sounds, do not prevent them to appear, but have the advantage of carrying the seeds of further, cultural evolution through the propagation of innovative variants, i.e. they allow to model both the stability of shared sound systems and their potential for sound change.

Acknowledgments

The writing of this paper was partially supported by the Brain and Language Research Institute at Aix-Marseille University (Labex BLRI, ANR-11-LABX-0036).

References

Arbib, M. A. (2002). The mirror system, imitation, and the evolution of language. In: Imitation in animals and artifacts, ed. C. Nehaniv & K. Dautenhahn, pp. 229–80. MIT Press.

Arbib, M. A. (2005a). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics*. Behavioral and Brain Sciences*, 28, 105-167.

Arbib, M. (2005b). The mirror system hypothesis: How did protolanguage evolve. In: Tallermann, M., ed. Language Origins: Perspectives on Evolution. Oxford, 21-47.

Aubanel, V. (2011). Variation phonologique régionale en interaction conversationnelle *Unpublished doctoral dissertation*, Université d'Aix-Marseille.

Auer, P. & Hinskens, F. (2005). The role of interpersonal accommodation in a theory of language change. In Auer, P., Hinskens, F., & Kerswill, P. (Eds.), *Dialect change. The convergence and divergence of dialects in contemporary societies* (pp. 35-57). Cambridge: Cambridge University Press.

Babel, M. (2010). Dialect convergence and divergence in New Zealand English. *Language in Society*, 4(3), 437-456.

Babel, M. (2011). Imitation in speech. *Acoustics today*, *7*(4), 16-23.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics,* 40, 177–189. doi: 10.1016/j.wocn.2011.09.001

Babel, M., & Bulatov, D. (2012). The Role of Fundamental Frequency in Phonetic Accommodation. *Language and Speech*, 55(2), 231-248.

Beckford-Wassink, A., Wright, R.A., & Franklin, A. D. (2007). Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics*, 35, 363–379.

De Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*(4), 441-465.

De Boer, B., & Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Advances in Complex Systems*, *15*(03n04).

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, 75, B13–B25.

Branigan, H. P., Pickering, M. J., McLean J. F., & Cleland AA. (2007). Syntactic alignment and participant role in dialogue. *Cognition,* 104, 163–197.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology, Learning, Memory, and Cognition,* 22, 1482-1493.

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. Les Cahiers de l'ICP. Bulletin de la communication parlée, (5), 25-34.

Charman, T. (2006). Imitation and the development of language. In Williams, J. & Rogers S. J. (Eds.), *Imitation and the social mind: autism and typical development* (pp. 96-117). New York: Guilford Press.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*. Odessa: Psychological Assessment Resources.

Delvaux, V., Huet, K., Piccaluga, M., & Harmegnies, B. (2013). Capacité d'apprentissage phonique et troubles du langage à étiologie cérébrale In Sock, R., Vaxelaire, B., & Fauth C. (Eds.), *La voix et la parole perturbées* (pp. 259-274), coll. Recherches en PArole n°2, Mons: CIPA.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64, 145–173.

Delvaux, V., Huet, K., Piccaluga, M., & Harmegnies, B. (2014). Phonetic compliance: a proof-of-concept study. *Frontiers in Psychology*, 5, Article 1375.

Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Frontiers in psychology*, *4*, Article 346.

Eckert, P. (2001). Style and social meaning. In Eckert, P., & Rickford, J. R. (Eds.), *Style and sociolinguistic variation* (pp. 119–126). Cambridge: Cambridge University Press.

Evans, B. G., & Iverson., P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*, 121, 3814–3826.

Fagyal, Z., Swarup, S., Escobar, A. M., Gasser, L., & Lakkaraju, K. (2010). Centers, peripheries, and popularity: The emergence of norms in simulated networks of linguistic influence. *Penn Working Papers in Linguistics,* 15(2), 81-90.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, 31(4), 307-319.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737-767.

Gallois, C., Ogay, T., & Giles, H. (2005). Communication Accommodation Theory: A look Back and a Look Ahead. In Gudykunst, W. B. (Ed.), *Theorizing About Intercultural Communication* (pp. 121–148). Thousand Oaks, CA: Sage.

Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and imitation of speech. *Frontiers in Psychology*, 4(600).

Garrod, S., Fay, N., Rogers, S., Walker, B., & Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols?. *Interaction Studies*, *11*(1), 33-50.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy?. *Trends in Cognitive Sciences*, *8*(1), 8-11.

Gentilucci, M., & Bernardis, R. (2007). Imitation during phoneme production, *Neuropsychologia*, 45, 608–615.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, 15, 87–105.

Giles, H., Coupland, J., & Coupland, N. (1991). Accommodation Theory: Communication, Context, and Consequence. In Giles, H., Coupland, J., & Coupland, N. (Eds.), *Contexts of Accommodation* (pp.1-68). New York, NY: Cambridge University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279. doi: 10.1037/0033-295X.105.2.251

Goldinger, S. D., & Azuma, T. (2004). Episodic memory in printed word naming. *Psychological Bulletin Review*, 11, 716–722.

Goldstein, L. (2003, August). Emergence of discrete gestures. In Proceedings of the 15th International Congress of Phonetic Sciences (pp. 85-88).

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to language via the mirror neuron system*, 215-249.

Goldstein, L. & Fowler, C.A. (2003). Articulatory phonology: A phonology for public language use. In Schiller, N.O. & Meyer, A.S. (eds.), *Phonetics and Phonology in Language Comprehension and Production*, pp. 159-207. Mouton de Gruyter.

Gratier, M., & Devouche, E. (2011). Imitation and repetition of prosodic contour in vocal interaction at 3 months. *Developmental Psychology,* 47(1), 67-76. doi:10.1037/a0020722

Gregory, S. W., & Hoyt, B. R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psychological Research*, 11, 35–46.

Gregory, S. W. (1990). Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior*, 14, 237–251.

Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English?. *Nature*, *408*(6815), 927-928.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, 34, 439–457.

Hickok, G. (2010). The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes*, 25, 749–776.

Honorof, D. N., Weihing, J., & Fowler, C. A. (2011). Articulatory events are imitated under rapid shadowing. *Journal of Phonetics*, 39, 18-38. doi:10.1016/j.wocn.2010.10.007

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528), 2325-2335.

Ke, J., Gong, T., & Wang W. S-Y. (2008). Language change and social networks. *Communications in Computational Physics*, 3(4), 935–949.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2, 125–156. doi: 10.1515/labphon.2011.004

Kopp, S., Wachsmuth, I., Bonaiuto, J. & Arbib, M. (2008). Imitation in embodied communication–from monkey mirror neurons to artificial humans. In Wachsmuth, I., Lenzen, M., & Knoblich, G. (Eds.), *Embodied Communication in Humans and Machines* (pp.357–390). Oxford: Oxford University Press.

Kuhl, P. K. and Meltzoff, A. N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 2425-38.

Kokkinaki, T., & Kugiumutzakis, G. (2000). Basic aspects of vocal imitation in infant-parent interaction during the first 6 months. *Journal of Reproductive and Infant Psychology*, 18, 173–187.

Kumashiro, M., Ishibashi, H., Uchiyama, Y., Itakura, S., Murata, A. & Iriki, A. (2003). Natural imitation induced by joint attention in Japanese monkeys. *International Journal of Psychophysiology,* 50, 81-99.

Lelong, A. (2012). Convergence phonétique en interaction, *Unpublished doctoral dissertation*, Grenoble University.

Lelong, A., & Bailly, G. (2011). Study of the phenomenon of phonetic convergence thanks to speech dominoes. In Esposito, A., Vinciarelli, A., Vicsi, K., Pelachaud, C., Nijholt, A. , eds., *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues* (pp. 273-286). Springer, Berlin.

Lev-Ari, S., & Peperkamp, S. (2014). An experimental study of the role of social factors in language change: The case of loanword adaptations. *Laboratory Phonology*, *5*(3), 379-401.

Lewandowski, N., & Dogil, G. (2009). Perception- production loop in native–non- native dialogs: Phonetic convergence. *The Journal of the Acoustical Society of America*, *125*(4), 2768.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W. J., & Marchal, A., eds., *Speech Production and Speech Modelling* (pp. 403-439). Springer Netherlands.

Masur E. F., & Eichorst D. L. (2002). Infants' spontaneous imitation of novel versus familiar words: Relations to observation and maternal report measures of their lexicons. *Merrill-Palmer Quarterly*. 48, 405–426.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 75-78.

Mercado, E., Mantell, J. T., & Pfordresher, P. Q. (2014). Imitating Sounds: A Cognitive Approach to Understanding Vocal Imitation. *Comparative Cognition and Behavior Reviews*. 9, 1-57. doi:10.3819/ccbr.2014.90002

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics*, 72, 1614–1625. doi:10.3758/APP.72.6.1614

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2013). Is Speech Alignment to Talkers or Tasks? *Attention, Perception and Psychophysics* , 75, 1817–1826. doi:10.3758/s13414-013-0517-y

Mitterer, H., & Ernestus, M. (2008). The link between perception and production is phonological and abstract: evidence from the shadowing task. *Cognition,* 109, 168–173. doi: 10.1016/j.cognition.2008.08.002

Mühlenbernd, R., & Quinley, J. (2013). Signaling and Simulations in Sociolinguistics. *Penn Working Papers in Linguistics*, 19(1), 129-138.

Munro, M. J., Derwing, T. M., & Flege, J. E. (1999). Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics*, 27, 385–403.

Namy L., Nygaard L. C, & Sauerteig D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21, 422–432.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790-804.

Nguyen, N., Dufour, S., & Brunellière, A. (2012). Does imitation facilitate word recognition in a non-native regional accent?. *Frontiers in psychology*, *3*, Article 480.

Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132-142.

Oudeyer, P. Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology*, *233*(3), 435-449.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382-2393.

Pardo, J. S., Cajori J., I., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics,* 72(8), 2254-2264.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190-197.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioural & Brain Sciences*, 27, 169–226.

Reiterer S., Hu X., Sumathi, T., & Singh N. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology,* 4. doi:10.3389/fpsyg.2013.00782

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. Trends in neurosciences, 21(5), 188-194.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. Nature Reviews Neuroscience, 2(9), 661-670.

Sanchez, K., Miller, R. M., & Rosenblum, L. D. (2010). Visual influences on alignment to voice onset time. *Journal of Speech, Language, and Hearing*, 53, 262–272.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421-436.

Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J. L., & Nguyen, N. (2013). Converging toward a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in psychology*, 4, Article 422.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in cognitive sciences*, *14*(9), 411-417.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66 (3), 422–429.

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass,* 3(1), 236□–264.

Studdert-Kennedy, M. (2000a) Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function. In Knight, C., Hurford, J. R., & Studdert-Kennedy, M. (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form* (pp.161-176). Cambridge: Cambridge University Press.

Studdert-Kennedy, M. (2000b). Imitation and the emergence of segments. *Phonetica*, 57, 275-283.

Studdert-Kennedy, M. (2005). How did language go discrete? Language Origins: Perspectives on Evolution, Oxford: Oxford University Press, pp. 48-67.

Tan, R., & Fay, N. (2011). Cultural transmission in the laboratory: agent interaction improves the intergenerational transfer of information. *Evolution and Human Behavior*, *32*(6), 399-406.

Trudgill, P. (2004). *New-dialect formation. The inevitability of colonial Englishes.* Edinburgh: Edinburgh University Press. 165pp.

Trudgill, P. (2008). Colonial Dialect Contact in the History of European Languages: On the Irrelevance of Identity to New-Dialect Formation. *Language in Society*, 37(2), 241-254.

Uther, M., Knoll, M., & Burnham, D. (2007). Do you speak E-N-G-L-I-S-H? A comparison of foreigner- and infant-directed speech. *Speech Communication*, 49, 2–7.

Verhoef, T., de Boer, B., Del Giudice, A., Padden, C., & Kirby, S. (2011). *Cultural evolution of combinatorial structure in ongoing artificial speech learning experiments* (Vol. 23, pp. 3-11). Center for Research and Language Technical Report, University of California, San Diego.

Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, *43*, 57-68.

Webb, J. T. (1970). Interview synchrony: An investigation of two speech rate measures in an automated standardized interview. In Siegman A. W., & Pope B. (Eds.), *Studies in dyadic communication: Proceedings of a research conference on the interview* (pp.115–133). New York: Pergamon.

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic Imitation from an Individual-Difference Perspective: Subjective Attitude, Personality and "Autistic" Traits. *PLoS ONE*, 8(9), e74746. doi:10.1371/journal.pone.0074746