Visible movements of the orofacial area: evidence for gestural or multimodal theories of language evolution?

Sławomir Wacewicz

Center for Language Evolution Studies, Nicolaus Copernicus University

wacewicz@umk.pl; Bojarskiego 1 (Room C.3.32), 87-100 Toruń, Poland

Przemysław Żywiczyński

Center for Language Evolution Studies, Nicolaus Copernicus University

przemek@umk.pl; Bojarskiego 1 (Room C.3.32), 87-100 Toruń, Poland

Sylwester Orzechowski

Institute of Psychology, Maria Curie-Sklodowska University

sylwesto@umcs.pl; Pl. Litewski 5 (Room 71), 20-080 Lublin, Poland

The age-old debate between the proponents of the gesture-first and speech-first positions has returned to occupy a central place in current language evolution theorizing. The gestural scenarios, suffering from the problem known as "modality transition" (why a gestural system would have changed into a predominantly spoken system), frequently appeal to the gestures of the orofacial area as a platform for this putative transition. Here, we review currently available evidence on the significance of the orofacial area in language evolution. While our review offers some support for orofacial movements as an evolutionary "bridge" between manual gesture and speech,

This is a post-print version. This article has been published in Gesture, 15(2), 250–282. Published 8 July 2016 [DOI: 10.1075/gest.15.2.05wac], copyright John Benjamins Publishing Company. The publisher should be contacted for permission to re-use or reprint the material in any form.

we see the evidence as far more consistent with a multimodal approach. We also suggest that, more generally, the "gestural versus spoken" formulation is limiting and would be better expressed in terms of the relative input and interplay of the visual and vocal-auditory sensory modalities.

Keywords: orofacial gestures, orofacial movements, tongue gestures, mouth gestures, evolution of language, Gestural Primacy Hypothesis, visual-vocal redundancy, hand-to-mouth links, gesture-first, speech-first, multimodality

Sławomir Wacewicz is an Assistant Professor of Linguistics at the Department of English, Nicolaus Copernicus University, Toruń, Poland. He is a co-founder and Deputy Head of Center for Language Evolution Studies. His research interests focus on the evolution of language, evolution of cognition, and the philosophy of language and mind.

Przemysław Żywiczyński is an Associate Professor of Linguistics at the Department of English, Nicolaus Copernicus University, Toruń, Poland. He is a co-founder and Head of Center for Language Evolution Studies as well as Vice-President of the Polish Society for Human and Evolutionary Studies. He publishes on language evolution, pragmatics, conversation analysis, and gesture studies.

Sylwester Orzechowski is an Assistant Professor at the Institute of Psychology, Maria Curie-Sklodowska University, Lublin, Poland. His research interests focus primarily on the role of gestures in language evolution. He is currently investigating the low-level coordination processes in human interactions.

1. Introduction

Was man's first language gestural or spoken? This age-old debate between the proponents of the gesture-first and speech-first positions has returned to central prominence in the most recent language evolution theorizing. However, this historically motivated framing of the debate about language evolution scenarios presents a false dichotomy. The traditional "gestural versus spoken" opposition would be better expressed in terms of the contrast and interplay between the sensory modalities of vision versus audition, in both production and reception. Here, we advocate such a restructuring of the debate: from the "gestural versus spoken" dichotomy, to the relative input and interplay of the visual and vocal-auditory sensory modalities, which need not necessarily contrast, but may interact in complex ways for achieving communicative effect. Taking the sensory modality as the starting point, but in a non-exclusionary way, is a step in the direction of the multimodal approaches to language origins recently championed by primatologists and gesturologists.

We devote the bulk of our paper to illustrating this point with an interesting but relatively underexplored class of phenomena: visible actions of the orofacial area. Traditionally, this topic has mostly been discussed in the context of the gestural scenarios of language emergence, where orofacial gestures – "mouth gestures" and/or "tongue-gestures" – were summoned in the context of explaining the transition from a gestural protolanguage to fully linguistic speech. Our review of the phenomenon of orofacial signalling does not rule out the possibility of orofacial gestures forming a potential "bridge" between manual gesture and speech. However, we see the evidence reviewed here as lending itself to a far better interpretation within the framework of a multimodal conception.

2. Gestural primacy hypotheses and the 'modality transition' problem

Serious discussion of accounting for the origin of language in naturalistic terms (rather than as a result of Divine intervention) developed extensively in the eighteenth century. The idea that language might have originated first from forms of visible bodily action or gesture was prominent in these discussions (see Hewes 1977a, 1996 for a historical survey). From the middle of the nineteenth century discussions of language origins fell out of fashion as being too speculative (see Stam 1976). However, from the mid-1950s advances in primatology, neuroscience, genetics, paleoanthropology and archaeology, combined with the cognitive revolution instigated by Chomsky and with the growing popularity of evolutionary explanations, made it possible to approach the problem of language origins in genuinely scientific terms. Although initial attempts heavily relied on speculation, so characteristic of traditional glossogenetic philosophizing (see e.g. Swadesh 1971; Diamond 1959), soon enough much more carefully designed proposals were formulated. A harbinger of this qualitative change was Charles Hockett, who in a series of publications launched a programme of systematic comparison between language and other, mainly non-human, communication systems (1958, 1959, 1960a, 1960b, 1966; Hockett & Altmann 1968). His design-features approach played a key role in alleviating the glossogenetic taboo in linguistics, and is still regularly appealed to in comparative contexts (see e.g. Hauser 1997; Fitch 2010 but also Liebal, Waller, Burrows & Slocombe 2014). Notwithstanding its historical importance, Hockett's framing of language as a disembodied, user-external code is making it obsolete in contemporary language evolution, which has more interest in the user-internal, evolved cognitive and socio-cognitive capacities underlying language use (for a discussion see Wacewicz & Żywiczyński 2015a).

The father of modern gestural theories of language origin was Gordon Hewes (1973, 1975, 1976, 1977a, 1977b, 1996), whose consideration for empirical detail became paradigmatic of the whole newly forming research area of language evolution. He was careful to reserve the term Gestural Primacy Hypothesis (GPH) for modern-day arguments in favour of gestural accounts of the emergence of language and managed to indicate lines of evidence that have secured GPH a lasting presence in the language evolution research. His inspiration came from Gardners' work with the chimpanzee Washoe (Gardner & Gardner 1969), which led to him to the – now firmly established – view that acquiring a (limited) vocabulary of manual as opposed to vocal signs is well within the reach of our nearest ape cousins, and thus presumably the last Common Ancestors (LCA) that humans share with chimpanzees (Arbib 2005; Tomasello 2008). Another important insight was related to the expressive, symbolic and combinatorial potential of gestures (Hewes 1973; Steklis & Harnard 1976; Corballis, 2002; Armstrong & Wilcox 2007). Hewes also singled out the problem of modality transition, i.e. transfer from the original visual modality to the primarily vocal modality of language, as the biggest challenge to gestural scenarios and argued that the prospects of GPH would depend on empirically well-founded and non-trivial explanation of the modality switch.

Over the past two decades, the gestural view of language origin has been strengthening its status as a major contender in the field of language evolution. Naturally, it has not been without its critics. For example, gesture scientists point to a very deep integration of speech with gesture in the prototypical use of language, i.e.

face-to-face conversation (Kendon 2004), taking it as evidence of a unified system with a common evolutionary history (Kendon 2011; McNeill et al. 2008; McNeill 2012). It has also been noted that in the face of the predominantly *spoken* nature of the large majority of today's natural languages, the gestural position is less explanatorily economical, as it has to postulate an extra step of "modality transition" (e.g. Bosman et al. 2005). This problem is particularly severe given the incontrovertible status of signed languages as fully functional human languages, which leaves the apparent "modality transition" or "modality switch" without any obvious motivation: a gestural protolanguage would seem to predict natural development into signed rather than spoken languages. This difficulty was already recognized by Hewes and has since then been posed as a central theoretical challenge to the gestural view by its proponents and critics alike (Burling 2005; Corballis 2002; Fitch 2010; Hewes 1973; Kendon 1991, 2008; MacNeilage 2008; Tallerman 2012).

3. Orofacial gestures and language origins

Recently, many authors have pointed to the phenomenon of the gestures of the orofacial area as a possible platform for accomplishing the required transition (e.g. Arbib 2012; Corballis 2003; Orzechowski et al. 2014). The idea of orofacial gestures as an important component in language origins itself has an interesting history. Its most recent proponents (Woll 2014; Meguerditchian et al. 2014) go back to the observation of Charles Darwin (1872), who noted imitative involvement of the orofacial area accompanying bodily routines, such as moving the jaw when using scissors, or moving the tongue in children learning to write. Although Woll (2014) credits the phonologist Henry Sweet (1888) with the first explicit statement of the possible involvement of this phenomenon in language emergence, it appears that Alfred Russel Wallace, the co-author of the natural selection theory, made a pioneering attempt to explore the glossogenetic potential of mouth and tongue gestures. In his review of Tylor's *Anthropology*, Wallace noted that the forms of the English words "come" and "go" are non-arbitrary (contrary to what Tylor claims), as the first of these is pronounced with a closure and contraction of the lips, while the latter, with a lip protrusion. Summoning the phenomenon of lip-pointing found among "many savages" as well as linguistic examples from French, German and Italian, he argued that the lip-protrusion in the word "go" can be understood as a mouth gesture for giving directions (Wallace 1881, pp. 244-245). Later, in "The Expressiveness of Speech", an essay in *The Fortnightly Review* (1895), he tried to combine this argument with both sound-symbolic and gestural accounts to propose a multi-faceted, and multimodal, scenario of language emergence.

The mouth gesture theory was revived by R. A. S. Paget (1930, 1944), who pointed to the fact that linguistic vocalizations are a product of movements performed by articulatory and phonatory organs. Drawing on Darwin's observation, Paget observed that the actions of the mouth and other articulators often echo hand movements, which led him to put forward a bold if controversial thesis: speech arose in the process of the mouth, tongue and lips involuntarily imitating body movements, out of which he considered hand gesticulations as most significant (Hawhee 2006). Since, as he argues, the hands became occupied with tool use, mouth and tongue gestures assumed the dominant role in this pantomimic mode of expression. Importantly, Paget was not concerned with the visual signalling of the orofacial area but with the acoustic consequences of the mouth and tongue movements, which– when heard – could be understood as imitative of the articulators' actions: "the significant elements in human speech are the postures and gestures [of the organs of articulation], rather than the sounds. The sounds only serve to indicate the postures and gestures which produced them. W lip-read by ear" (1930, p. 174). To support his claims, Paget amassed linguistic material from many unrelated languages (including Chinese, Sumerian and Arawak) with a view to demonstrating that phonetic and semantic resemblances between them could be best explained by the mouth-gesture theory. The Icelandic linguist Jóhanneson, who was independently engaged in a very similar project, summoned bulky evidence from Indo-European and Semitic languages to claim that as many as 85% of lexemes were derived from mouth gestures understood as the movements of lips and tongue (1949; cf. Hewes 1977b).

More recently, orofacial gestures have been proposed as a candidate "link" between the hypothesized gestural stage of protolinguistic communication and spoken language – based on a variety of arguments coming from anthropology (Hewes, e.g. 1996), linguistics (Studdert-Kennedy 2002), and primatology (e.g. Meguerditchian et al. 2011). The most comprehensive account of the role of orofacial gestures in the putative gesture-speech transition comes from Michael Corballis (2002, 2003, 2012), who provides a more-up-to-date version of Hewes's argument about the neural control of primate gestural communication. Corballis highlights the fact that monkeys and non-human apes possess voluntary control over manual as well orofacial gestural actions afforded by neocortical connections, which does not likewise extend to the control of vocalization. For Corballis, this constitutes a platform upon which language could be built. The evolutionary transition from hand and mouth gestures to speech would have been late and gradual, and even today is not complete, as illustrated by the phenomenon of co-speech gesturing.

Corballis also notes that speech itself is fundamentally a system of movements of the speech organs, and may thus be considered a system of "gestures". This echoes the idea originally developed by Armstrong, Stokoe and Wilcox (1994, 1995; also Armstrong and Wilcox 2007) of both speech and gesture being "planned sequences of musculo-skeletal actions" (although, as Kendon [2008: 13] notes, the similarity is rather superficial and obfuscates the difference in the perceptual processes for speech and gesture comprehension). Following this line of thinking, there is continuity from manual gestures, through visible orofacial gestures to the invisible gestures of the vocal tract. Vocalization is the addition that makes these movements accessible to the receiver in the auditory modality. In that sense speech is simply "swallowed" orofacial gesturing. With the emergence of complex combinatorial meanings and grammatical structure, a repertoire of these gestures would have had to grow larger and finally gain autonomy from the visually perceptible base. On this scenario, the relatively flexible orofacial area plays a leading role in the gradual evolutionary extension of flexible voluntary control to the more internal parts of the vocal tract.

Orofacial gestures form a pivotal component of the gestural-pantomimic scenario developed by Michael Arbib (2002, 2005, 2006, 2012). According to his influential Mirror System Hypothesis (MSH), the original gestural mode of communication, which involved the use of both manual and orofacial gestures, "recruited" vocalization, thus giving rise to a more open referential system. Explaining the logistics of this transition, Arbib emphasizes the role of the orofacial mirror neuron system discovered in monkeys' homologue of Broca's area (the F5 area) and argues that a comparable structure in the hominin brain provided the neural infrastructure for volitionally controlled vocalizations. In a similar vein, primatologists such as Leavens, Taglialatela and Hopkins (2014) or Meguerditchian and Vauclair (2014), appealing to comparative data, argue that “the oro-facial system might constitute a relevant mediator between the gestural communicatory system and speech in the evolution of language” (Meguerditchian & Vauclair 2014, p. 148). A closer discussion follows in sections 5.2 and 8.

A different view on the evolutionary role of the orofacial area is offered by the psycholinguist Peter MacNeilage. Rather than with orofacial gestures *per se*, his *frame-content theory* begins with the cyclical movements of the jaw during food ingestion. To McNeilage, the rhythmical mandibular oscillations (opening and closing of the jaw) that occur when masticating food did not provide a bridge between gestures and speech; rather, they were a starting point for language, which – as he insists – has existed in the vocal-auditory channel since its earliest evolutionary beginnings (MacNeilage & Davies 2005; MacNeilage, 2008). Ferrari, Gallese, Rizzolatti and Fogassi (2003) furnish supporting evidence for some link between masticatory actions and the communicative function, showing that in monkeys, observing ingestive actions – such as grasping, sucking or breaking food – causes response of mirror neurons in the F5 area, which is considered as a homologue of Broca’s region in humans. This extends the scope of the Mirror Neuron System from hand actions to mouth actions, with implications for a communicative potential of the movements involved in mastication.

4. From “gesture versus speech” to sensory modality

Does research into visible movements of the orofacial area provide evidence for a gestural theory of language origins? We wish to begin by noting that this question is very likely ill-posed. Although motivated historically, framing the debate in the binary terms of gesture versus speech is misguided and too restrictive, for several reasons. Firstly, from the vantage point of communicative dynamics, the term “gesturing” favours the producer’s role relative to that of the recipient. Referring instead to the visual and vocal modalities – as we prefer – provides a more balanced view of interaction, attending both to the producer’s communicative actions (not only manual) and to the receiver’s states of visual attention in processing those actions. Furthermore, such a view is more inclusive: instead of the narrow focus on the hand and arm, it recognizes the communicative contributions of other types of visual information. In fact, a variety of language origins scenarios classified as “gestural”, from Mandeville and Condillac to Arbib (2012) or Tomasello (2008), rely on full-body communicative action rather than manual gesture alone.

Finally, and most importantly, restructuring the debate in terms of the relative input and interplay of the *visual and vocal-auditory sensory modalities* helps to see those two modalities not as mutually exclusive alternatives, but rather as providing a range of semiotic resources which stand in complex and diverse relations to each other.

From an evolutionary point of view, this takes the focus off the question of “primacy”, and foregrounds the relative input or division of labour between the modalities on different stages of the development of the language faculty (Collins 2013). Such observations have led to the rise of the *multimodal theories,* which see a nontrivial role of each of the two major modalities – as well as their closer connection and interaction – throughout the entire process of language evolution (e.g. Kendon 2011; McNeill 2012; Sandler 2013). Although relatively recent, the multimodal accounts are quickly gaining ground, cf. the symptomatic statement by Gillespie-Lynch, Greenfield, Lyn, and Savage-Rumbaugh (2014, p. 6) “Although the study that is the focus of this review was designed to evaluate the gestural theory of language evolution, our findings revealed unexpected support for the multimodal theory of language evolution. In retrospect, a multimodal theory of language evolution is more logical than a purely gestural theory...".

It is important to observe that such approaches usually work from a better understanding of animal communication in general, and human communication in particular, which improves our conceptions of both the "starting point" and "end state" for language evolution (cf. McNeill in press). As for the "original substrate", our best inference about the putative capacities of the *Pan-Homo* LCA comes from extant communication systems in monkeys and nonhuman apes; recently available comparative data show non-human primate communication to be multimodal to a much broader extent than previously acknowledged. To some degree this is unavoidable, because many communicative actions combine visual and vocal information by their very nature (e.g. shaking a tree branch). However, apes can and do change between the communication channels flexibly and strategically. For example, Leavens, Russell and Hopkins (2010, p. 39) note that "the ability to exercise choice over modality of communication and to tactically vary the display of signals within a context-appropriate modality emerges in captive populations of chimpanzees in the complete absence of any explicit training to do so". Wild chimpanzees fall back upon the visual mode when more secrecy is needed (e.g. Hobaiter & Byrne 2012), but use the vocal-auditory mode for attracting attention, by making calls or auditory gestures, or using objects (e.g. leaf clipping to produce loud noises, Matsumoto-Oda & Tomonaga 2005). They also combine visual and vocal signalling depending on the communicative context (Taglialatela, Russell, Pope, Morton, Bogart, Reamer, Schapiro & Hopkins 2015).

Liebal, Waller, Burrows and Slocombe (2014, esp. Chapters 5 and 10) summarize impressive evidence documenting the role of both auditory and visual (manual, facial, and whole-body) signals, as well as some input from other modalities.

As for the "end state", the notion of language is now undergoing a major theoretical change. Starting with de Saussure, linguists were committed to the view that language constitutes a system independent of other, that is non-linguistic, semiotic resources. Such an attitude was accompanied by the tendency to displace language from the context of its prototypical use – face-to-face interaction – and treat it as a system of abstract rules. At this juncture, it should be remembered that linguistics as a scientific enterprise grew out the philological tradition preoccupied with the study of written texts, primarily Latin and Greek (Harpham 2009). With the rise of historical and comparative linguistics in the first half of 19th century, the philological way of thinking about language with its heavy emphasis on ruled-based morphosyntax (Ashcroft 2001) and the written language bias (Linell 2005) was solidified into a model of linguistic description. The culmination of this "longstanding obsession of linguistics" (to borrow a phrase from Hewes 1973, p. 11) was the generative programme defined by the claim that the essential property of language is combinatorial syntax (e.g. Pinker 1994, p. 124). Syntactocentrism has been robust in the discussions of evolution of language by Chomsky and his various collaborators stressing the radical disparity between non-human communication and language understood as a computational processor (Hauser et al. 2002; Fitch et al. 2005; Bolhuis et al. 2014).

Currently, these views that regard language as an autonomous module are slowly being abandoned in favour looking at language in the context of its presumed prototypical use – face-to-face conversation. This new way of thinking about language recognizes its multifaceted nature, including among the components traditionally thought of as belonging to language also the cognitive infrastructure related to intentionality (Grice 1975; Sperber & Wilson 1986; Tomasello 2008), mental imagery and meaning embodiment (Langacker 1987; Lakoff & Johnson 1999), or socio-normative elements (e.g. Watts 2003). More importantly for the context at hand, there is a growing tendency to view linguistic communication as relying not just on speech but a variety of semiotic resources, such as gesture, posture or prosody (Vigliocco et al. 2014, p. 1), which leads to the idea of language as "multi-modalic orchestration", to use Kendon's wording (Kendon 2011, p. 267). Of course, this trend for communicative holism has not appeared *de novo* in recent years; it has strong roots in functional and cognitive linguistics (Halliday 1973; Lakoff & Johnson 1980), Goffman's interactionism (1959, 1969), Vygotsky's psychology (1978, 1987), and the tradition of distributed cognition, or more specifically distributed language with its focus on languaging – the use of linguistic practices in real-life cognitive and communicative dialogic activities (Cowley 2011).

However, the decisive effort to give the multimodal conception a firm empirical footing was taken by gesturologists. The idea that linguistic communication depends on the organisation and synchronisation of body movements and speech has been a recurrent motif in Kendon's works (e.g. 1972, 2004, 2011, 2014b). McNeill has sought to elaborate a theoretical model for speech, hand movement and thought integration based on the notion of growth point – a minimal psychological unit of language expressed by speech and gesture (1992, 2000, 2005). A lot of empirical findings about the relation between gesture and conceptual organisation of utterances as well as the significance of gestures for language acquisition comes from Goldin-Meadow's research (2003, 2008).

The gesturologists were also the first to transpose arguments about the multimodal nature of language, or rather languaging, onto the evolutionary plane, claiming that from its beginnings, it "involved vocal signaling as well as signaling through visible bodily action" (Kendon 2011, p. 254). Consistent with an increasing awareness of the complex and multimodal nature of human communication is the 2014 issue the *Philosophical Transactions of the Royal Society B,* "Language as a multimodal phenomenon: implications for language learning, processing and evolution" (edited by Vigliocco, Perniss, & Vinson), documenting the extent to which multimodal conceptions has penetrated ways of thinking about language – e.g. its acquisition (Goldin-Meadow 2014; Liszkowski 2014), processing (Özyürek 2014; Skipper 2014) and, of course, evolutionary beginnings (Levinson & Holler 2014; Sereno 2014). In a comparative context, another 2014 publication (Liebal et al. 2014) illustrates exciting prospects for the multimodal approach to the study of primate communication. Certainly, multimodal approaches to language origins reflect a recent trend towards viewing language evolution from a very broad, holistic and multidisciplinary perspective; most importantly, appreciation of its embedding in a social context (Dor et al. 2014; Wacewicz & Żywiczyński 2015b). Of particular note are accounts that emphasize a more holistic use of not just modalities, but a wide range of different semiotic resources, such as song, dance, vocal imitation or pantomime (see e.g. Kendon 2014a; Levinson & Holler 2014; Lewis 2014; Zlatev 2014).

In sum, we agree with Kendon (2011) in observing that the "gestural versus spoken" opposition, and the resulting "modality transition" problem, is superficial.

Language evolution theorising frequently oversimplifies both the starting point as gestural and the end-state as speech (e.g. Mühlenbernd et al. 2014). When taken seriously, available primatological as well as anthropological evidence shows unambiguously that the communication of extant nonhuman apes (and thus the hypothetical communicative system of the *Pan-Homo* LCA) as well as present-day naturalistic communication of humans both have a very profoundly multimodal character. In what follows, we review available evidence of the communicative role of the orofacial area and then evaluate this information in the context of language origins.

5. Visible movements of the orofacial area

In language origins research, and linguistically informed research in general, the rich communicative potential of the orofacial area in producing visual signals has not been the center of interest, at least when compared to its role in sound production. However, in primates, and especially the almost exclusively diurnal monkeys and apes, the face provides some of the most vital social information. This anatomical area is fundamental to individual recognition (even kinship recognition) and to assessing the age, sex, health status, reproductive value as well as current affective state and visual attention of the individual, and it also aids in evaluating its physical strength and social status (e.g. Parr & de Waal 1999; Parr et al. 2010; Little et al. 2008; Little et al. 2011). Consequently, the facial area serves as the natural focus of visual attention in primate social interactions.

Consistently with our focus on the modality, below we take into consideration all *visible movements of the orofacial area*, i.e. those movements of the muscles of the front of the head (including the ocular, masticatory, facial and lingual muscles) that are visually accessible to other individuals. This broad class of behaviours with distinct neural or psychological mechanisms fulfil a range of primary functions, out of which we foreground the communicative function *sensu* Ekman and Friesen (1969). Consequently, central to our discussion are *orofacial gestures*, which we understand as movements characterized by a high degree of voluntary control, whose production is volitional, flexible, and originates from a communicative intention: they are intended to convey specific information to other individuals. To use Goffman's (1963, pp.13-14) famous distinction between information one "gives" and "emits, exudes, or gives off", the interactional status of orofacial gestures, as we define them, is that of "expressions given" rather than "given off". Prototypically, orofacial gestures are such movements that can be readily supplemented with the production of sound, and whose elements can be controlled individually rather than as complexes. We single out *articulatory gestures* as a separate class; although they are volitional and visually perceptible to a certain extent, their communicative function is realized solely through the production of speech sounds (Browman & Goldstein 1989; Eccardt 2006).

A related category are *mouth and tongue gestures* understood as kinetic echoes of limb, and more narrowly hand, movements. In contrast to orofacial gestures they are performed with little voluntary control and communicative intention but can be easily co-opted for sound production. Historically, authors such as Darwin, Wallace, Paget and Jóhanneson were concerned with this type of mouth and tongue actions; their speculations are often referred to as the mouth gesture theory of language origin, the term coined by Wallace (1895), although "the mouth and tongue gesture theory" would perhaps be an apter description. A crucial aspect of this way of thinking about the emergence of language is the versatile potential of mouth and tongue gestures for sound production, providing a “link” between the visual and vocal-auditory modalities.

Next, there are complex actions of the face related to exhibiting emotions. Following some psychological literature, we distinguish *facial displays*, which intentionally convey information pertaining to actions that an organism is ready to pursue (Fridlund 1997). Thus, “facial displays are expressive *to* another person rather than expressive *of* an underlying state” (Chovil 1997, p. 321), including discourse-oriented facial displays such as brow movements or upper lip raising as “linguistic elements of a message” (Chovil 1991/1992, p. 166). This differs from *facial expression*, which is frequently taken as involuntary expression of emotion or other internal states, with its types tightly linked to specific classes of releaser stimuli. However, the degree of voluntary control and intentionality may vary (see 4.2.), and the distinction between facial expression and facial displays in the sense above is not always sharp. Note that from the point of view of the signalling theory, facial expressions are signals rather than cues. *Signals* are behaviours or features that evolved “for” signalling, and their design is specifically shaped by selection pressures to influence the actions of the receivers, whereas *cues* are behaviours or features that supply information but did not evolve “for” the signalling function (Maynard Smith & Harper 2003, pp. 3-8).

Other visible movements of the face and mouth are not directly motivated by the social interaction context in either the functional or phylogenetic sense: they are above all cues, not signals. Still, on Ekman and Friesen’s (1969) classification, they can be *informative*, i.e. decoded by observers as “meaningful” even if not produced by the sender with an intention to convey information. This includes *ingestive movements*, performed during chewing and swallowing food, as well as *manipulative movements* for operating objects (indeed, dental wear evidence indicates that pre-sapiens hominins relied on the mouth and the teeth as a third hand to a greater extent than present day humans, (see e.g. Le Cabec et al. 2013). Interestingly, both ingestion and oral grasping were found to activate the mirror neuron system in the macaque (Ferrari et al. 2003).

5.1. Facial expression

Interest in facial expressions in primates has a long history (e.g. Darwin 1872; Hinde & Rowell 1962; van Hooff 1967). Facial expressions are essential in the social lives of apes as "critically important for coordinating social interaction, facilitating group cohesion and maintaining individual social relationships" (Parr & Waller 2006, p. 221). Although facial expression is perceived visually, recently there has been growing appreciation of its essentially *multimodal* nature, resulting in appeals for more integrative research (e.g. Slocombe et al. 2011). The production of many vocalizations is naturally coupled with a "dedicated" facial expression (see e.g. Goodall 1986), but more interestingly this is also reflected in reception. For example, Ghazanfar and Logothetis (2003) showed that rhesus monkeys perceive vocalization and the accompanying facial expression cross-modally, as a unified signal. Similarly, Parr (2004) found that captive chimpanzees can recognize a facial expression only by the corresponding vocalization – although, characteristically, composite visual-vocal stimulus in the multimodal condition led to improved performance relative to the unimodal condition.

5.2. Voluntary control and plasticity

Both facial expression (as opposed to facial displays) and vocalization in non-human primates have traditionally been seen as largely innate, emotionally driven and involuntary, and thus dissociated from language-like communication or its precursors. However, the discontinuity between emotional, involuntary facial expression and voluntary orofacial praxis is in fact far from obvious. In humans, almost every single muscle of the face may be voluntarily moved, even if for some of them this requires training or prior electrical stimulation (Ekman et al. 2002). The performance of professional actors and mimes illustrates the human potential to bring under voluntary control and deploy intentionally even seemingly inflexible facial configurations. Thus, as we noted above, the division into facial expressions and facial displays is not always clear-cut.

The proposed rigidity of vocal signals served as one of the keystones of the argumentation in favour of the gestural hypotheses (e.g. Hewes 1976, 1977a, 1996; Corballis 2002; Tomasello 2008). This line of argumentation stresses the discontinuity between subcortically controlled, species-specific vocalisations of monkeys and apes, which are primarily related to emotional expression, and the voluntary execution of speech in humans. As stressed by Hewes (e.g. 1973, 1977a) and Corballis (2002), since apes are able to engage in volitional gestural communication, it constitutes a likely platform for the origin of protolanguage.

Now a wealth of recent studies have converged to qualify if not question this received view in several ways. Firstly, even the relatively rigid vocalizations manifest a degree of plasticity dependent on social and cognitive factors. For example, Clay and Zuberbühler (2014) review evidence for chimpanzee and bonobo calls being combined in sequences, and modified based on the composition of the audience ("audience effects") or even on their state of knowledge, since calls are apparently used to inform naive individuals. Watson et al. (2015) report a case of captive chimpanzees who, after their transfer into a new social group, have changed the acoustic structure of a food call to match more closely that characteristic of the target group; this demonstrates some potential for ontogenetic plasticity. Leavens, Taglialatela and Hopkins (2014) discuss flexible use of attention-getting calls in chimpanzees, including at least one voiced call (extended grunt), made with the use of the vocal folds.

Secondly, a distinction should be made between articulation and phonation. While it is true that monkeys and apes have limited phonatory control (see e.g. Sutton 1979; Eberl 2010; Ackermann & Ziegler 2010), this does not similarly extend to their control of the supralaryngeal vocal tract. Unsurprisingly, monkeys and apes possess considerable voluntary control of the tongue and the mouth, which means that the motor and neural substrates are in place for relatively refined articulation of learned sounds. Zimmermann et al. (2013, p. 126) speak of "an unexpected degree of freedom in orofacial gestures of learned sounds … in several nonhuman mammalian species", which include chimpanzees and orangutans. In addition to the extended grunts mentioned above, Leavens et al. (2014) list kisses, lip smacks, pants, raspberries, and teeth chomps as examples of attention-getting calls used voluntarily by both wild and captive apes; they also cite examples of voluntary skilled orofacial action in captive apes, such as blowing up a balloon or smoking. Clark and Perlman (2014) report a range (admittedly small) of acquired sounds in the enculturated female gorilla Koko, who uses them in coordination with manual behaviours and gestures. Furthermore, lip smacking in macaque monkeys was found to display structural, especially rhythmical, similarities to human speech (Ghazanfar et al. 2012). Here, a particularly impressive illustration is the rhythmical lip smacking in geladas accompanied by a derived vocalization component (Bergman 2013), which possesses the periodicity so closely resembling human speech (6-9 Hz) that reports exist of geladas being mistaken for talking humans.

We want to emphasize that where the sound signalling in apes is communicatively flexible and ontogenetically plastic, it appears to involve mostly the front of the mouth and the orofacial area, that is the areas best *visually* accessible to the observer. Notably, such behaviours are most common in affiliative, relaxed social interactions, which could be termed "proto-conversational" from the human perspective. For example, wild chimpanzees apparently use lip-smacking, "a distinct multimodal oral gesture produced during grooming", to coordinate bouts of grooming in a cooperative way (Fedurek et al. 2015). Leavens et al. (2014) take similar observations to support Corballis's "hand-to-mouth" idea of gesture-to-speech evolutionary transition – especially in conjunction with Dunbar's (1996) vocal grooming hypothesis, which stresses the social context and social purpose of communication. We find merit in this proposal, but would like to suggest that the above data lend themselves to a broader interpretation, viewing both the vocal and the visual to be part of a composite and partly redundantly structured message (see below).

Finally, non-human primates possess the neural substrates for the multimodal mapping between the visual input and motor output of orofacial movement – a necessary (if not sufficient) precondition for social learning of such movements (Arbib 2012, pp. 73-74). Although apes and monkeys are incapable of imitation – at least in the sense of spontaneous, high-fidelity copying of instrumental actions (e.g. Call & Carpenter 2003) – they apparently share with humans the ability for neonatal imitation of facial movements (e.g. Bard 2007). Particularly informative are the monkey data on mirror neurons. There are bimodal – that is audio-visual – mirror neurons which code actions independently of whether they are performed, seen or heard (Kohler et al. 2002; Gazzola et al. 2006); however, no such mirror neurons seem to exist for monkey conspecific vocalization (e.g. Coudé et al. 2011). But there are mirror systems for orofacial movement, including both ingestive action such as chewing or lip smacking and communicative gestures such as silent human speech (see Ferrari et al. 2003; Buccino et al. 2004); likewise, auditory mirror systems exist for listening to and execution of instrumental mouth actions (Gazzola et al. 2006).

6. Orofacial action and language

Visible movements of the area of the face and mouth – mainly facial displays as defined in the section 5 – fulfil a variety of roles in face-to-face use of language. There is now a sizeable body of research showing them as an integral component of conversational meaning-making, both on the levels of conveying content and regulating interaction. Birdwhistell (1970) observes that facial displays serve such linguistic functions as marking emphasis, supplementing speech and backchanneling; this last point is developed by Brunner (1979), who demonstrates that smiles constitute an important type of backchanneling signal. Content-wise, facial displays bear on messages communicated by speech in both redundant and non-redundant ways, and their context-dependent character is often emphasized (Chovil 1991/1992, 1997). Facial displays also aid a variety of syntactic and prosodic purposes, e.g. help formulate questions, mark punctuation, indicate speech emphasis or syntactic functions (Ekman 1979). This is related to the vital role of facial displays for discourse organisation, for example topic initiation (Chovil 1997). Clark and Gerrig (1990) enumerate three principal means by which this type of visual information conveys meaning: it adds on the information communicated linguistically, draws interactants' attention to objects or events, and helps them act out elements of verbal messages. In sum, "... facial displays of conversants are *active, symbolic* components of *integrated* messages" (Bavelas & Chovil 1997, p. 334) as well as play an important regulatory role in conversational dynamics.

6.1. Orofacial gestures in sign language

Signed languages of the deaf define a special context in which orofacial gestures are employed strictly linguistically. Orofacial gestures may function grammatically in a sign language and can even show patterns of L1 to L2 transfer in sign-speech bilinguals (e.g. eyebrow raise, which marks conditionals in the American Sign Language; Pyers & Emmorey 2008). Orofacial signs are either *mouthings* – oral patterns tracing the pronunciation of the corresponding word in a spoken language – or *mouth gestures*, which are not derived from spoken languages (Woll 2001; Sutton-Spence 2005). Note that used in this sense, *mouth gestures* differ from our definition of the term: although they do accompany hand actions, mouth gestures related to sign language are executed volitionally and with a communicative intention. Mouth gestures of this sort frequently involve sound production: "[s]ome LSN signers mark certain adverbial and aspectual information with mouth gestures often accompanied by distinct vocalizations that are visually detectable …" (Kegl et al. 1999, p. 183). Mouth gestures are often iconic (Sutton-Spence 2005, p. 11), and this iconicity appears to be "overdetermined" visually and vocally, i.e. the modalities work together to express the same motivated information (see section 7).

Woll (2014, p. 4) highlights a special class of mouth gestures, Echo Phonology, where "the mouth action is a visual and motoric <echo> of the hand action in a number of respects: onset and offset, dynamic characteristics (speed and acceleration) and type of movement (e.g., opening or closing of the hand, wiggling of the fingers)". In a way strongly reminiscent of Paget and Jóhanneson, Woll sees echo phonology as a candidate mechanism for the evolutionary transition from manual gesture to speech, and especially from motivated manual signals to abstract vocal ones. She proposes "a possible leap from echo phonology in signs to a situation where voicing accompanies these mouth gestures so that they begin to have independent existence as lexical items" (2014, pp. 5-6); this is reinforced by neuroimaging data showing that the brain activation patterns of echo phonological mouth gestures resemble those of manual-only signs rather than speech (Capek et al. 2008 in Woll 2014). To this, we would like to add that echo phonology is another case where the message is "overdetermined", as the mouth and the hand work together to convey the message with some degree of redundancy. Other examples of partly redundant and partly complementary hand and mouth communicative action involve "symbiotic" orofacial co-sign gesturing, in which "iconic mouth gestures are co-temporal with the manual verbal string, and they complement or embellish it" (Sandler 2009, p. 255) as well as manual emblems that have an obligatory orofacial component (Ricci Bitti & Poggi 1991).

6.2. The orofacial area in speech perception

Visual information from the orofacial area has an important and well documented role in speech recognition: it is processed jointly with the vocal information and integrated with it on early processing stages. An often cited example is the so-called McGurk effect, in which a visual image of the speaker's mouth producing the syllable /ga/ shown with the sound of the syllable /ba/ results in the perception of the syllable /da/, /d/'s articulatory position being intermediate between /b/ and /g/ (McGurk & MacDonald 1976). It is particularly noteworthy that phenomenologically this artificial composite stimulus "feels" completely natural to observers, which shows that the integration of both modalities works on a very fundamental level in the cognitive unconscious. Under normal circumstances, visual information "overdetermines" the acoustic message, being used to improve recognition accuracy. As is well known, it is possible to decode speech visually, which is perhaps best evidenced by the high success rates of deaf persons in lip and face reading (Summerfield 1992). Seeing the face of the speaker also aids speech comprehension in hearing individuals, including comprehension in L2 learners (Sueyoshi & Hardison 2005; note that the largest improvement was found in advanced learners). The gains from visually transmitted information in decoding speech are made clear by the development of *bimodal*, that is auditory plus visual, automatic speech recognition systems (e.g. Chibelushi et al. 2002). The bimodal integration of speech signals is convincingly explained by the fact that vocal tract movements result in predictable orofacial configurations, as empirically demonstrated by Jiang et al. (2002) and Yehia et al. (1998) and Yehia et al. (2002). In relation to this point, Ghazanfar and Lewkowicz (2008) show that the behaviour of the vocal tract determines the spatiotemporal behaviour of the orofacial area. The bimodal account gains further support from studies showing increased excitability of the orofacial muscles during speech perception tasks (e.g. Watkins & Paus 2004; cf. Ravizza 2005). It is worth observing that all the data summoned in this section are consistent with the motor theory of speech perception, which views the core representation of the phonological system not as sound values but rather as movements – gestures – executed by the articulatory apparatus (Liberman et al. 1967; Liberman & Mattingly 1985; Liberman & Whalen 2000; see also Galantucci et al. 2006).

6.3 Orofacial praxis and language

A possibly important piece of the puzzle is the role of the genetic substrates for orofacial *praxis* – i.e. generalized fine motor control – on the one hand, and a broad range of linguistic skills on the other. An interesting case of disruptions to orofacial praxis is exemplified by developmental verbal dyspraxia (DVD) caused by a mutation to the *FOXP2* gene (i.e. the KE phenotype; Vargha-Khadem et al. 1998). This condition represents a link on a genetic level between orofacial movement and language. It involves impairment of orofacial praxis: the control of the muscles of that area is impaired, disrupting the performance of complex volitional orofacial movements, *not necessarily related to speech*. However, as is well known, the affected individuals also display a range of *language-related* deficits, including impaired speech production and comprehension of syntactically complex sentences (see e.g. Marcus & Fisher 2003; Enard et al. 2002). The gene *SRPX2* – itself a target of *FOXP2* – likewise seems to affect both orofacial praxis (also finger praxis) and verbal ability more generally (Roll et al. 2006; see also Sia et al. 2013).

7. Visual-vocal redundancy

As we have seen repeatedly, sound production by the vocal apparatus is a motor process that to a significant extent manifests itself externally in the form of predictable orofacial configurations. As a result, the specific anatomic configurations leading to the production of particular sounds are visually observable, and frequently also visually identifiable. In many cases we can speak of obligatory pairings of visual and vocal signals – e.g. an open mouth and roaring vocalization (Chevalier-Skolnikoff 1973) – where the two modalities convey information that is "redundant" from the receiver's perspective.

One well-studied aspect of such orofacial-vocal redundancy is related to the observation made originally by Darwin (1872; later corroborated by ethological research: e.g. Andrew 1963; van Hooff 1962, 1967, 1972) that humans and non-human primates exhibit similar facial gestures during emotional expressions: specifically, the smile (or the lip-corner retraction gesture) is linked to a variety of submissive displays, whereas the so-called "o-face", performed with the lips rounded and protruded, is involved in expressing aggression and disapproval. Based on this idea, Morton (1977) and Ohala (1983, 1994) postulated the existence of an orofacial-frequency code in which the two gestural complexes are correlated with vocalizations of distinct frequencies. In smiling, the face is shaped so as to produce a higher-pitch (higher F_0) signal, indicative of a smaller body size and typical of submissive screams. The "o-face", in contrast, is a shape associated with the production of lower-F_0 signals, which exaggerate the size of the animal, leading to the perception of that expression as a threat signal. Empirical studies, targeting mainly monkeys and non-human apes, confirmed this assumption (August & Anderson 1987; Bauer 1987; Hauser 1993, 1997).

This affords an interesting perspective on *sound symbolism*, which has been recently acknowledged as a robust phenomenon and proposed to help bootstrap language acquisition and potentially language evolution (Imai & Kita 2014). Note that nonarbitrary pairings between sounds and meanings are accompanied by predictable orofacial configurations that are necessary to produce those sounds. Consequently, such configurations may also become associated with the corresponding meanings visually, and sound symbolism may come to have a visual component (cf. Gentilucci & Corballis 2006; see also section 6). This type of overdetermination can lead to interactive and cognitive benefits, e.g. intersensory redundancy was found to aid aspects of early language acquisition (Gogate & Bahrick 1998; see also Gillespie-Lynch et al. 2014). It can also be seen in mouth gestures in signed languages (see 4.1), and is particularly evident with motivated meanings, which seem to display convergent vocal and visual iconicity, cf. Sutton-Spence's (2005, p. 11) examples: "pursed lips for something very small, air escaping through vibrating lips for a steady movement".

8. Hand-to-mouth: neural and behavioural links

Recently, an increasing number of studies have emphasized neuro-motor and behavioural links in human and nonhuman primates between the orofacial area and the upper limb, and in particular between the lips and the fingers. In monkeys, electrical stimulation of the Brodmann area 44, Broca's homologue, was found to produce hand and lip movements (Petrides et al. 2005; see also Petrides & Pandya 2009). More fine-grained behavioural links have been observed in nonhuman apes. For example, captive chimpanzees move the lips when precision-gripping and use the mouth as a third hand (cf. Leavens et al. 2014). They also perform more "sympathetic" mouth movements in the context of concurrent fine object manipulation than that of concurrent gross object manipulation (Waters & Fouts 2002). Similarly, Meguerditchian et al. (2014) report "involuntary" or automatic orofacial movements associated with finger movements in wild chimpanzees and note that precision gripping tends to be preferentially accompanied by lip smacks, while gestures such as "hair sweeping" – by a greater proportion of "chewing" or "kisses". This leads them to postulate "strong neuro-motor links" in our ape cousins.

In humans, there is ample evidence for simultaneous, spontaneous, concerted hand and lip movement, starting with Darwin's observations (see 3). Movements of the facial area and finger praxis are often linked in fine skilled action, as e.g. knitting or playing a musical instrument. This is corroborated by recent neurological and psychological evidence. Grasping a small versus large object while simultaneously saying a syllable leads to differential lip aperture sizes (greater when grasping a large object) and distinct voice spectra; this was true even when the subjects observed the grasping action rather than performing it themselves (Gentilucci & Corballis 2006). Gentilucci and Dalla Volta (2008) complement similar behavioural data with neuroimaging and rTMS findings to argue for a systemic relationship between arm motor control and linguistic vocalizations, claiming that orofacial behaviours are manifestations of this system. Higginbotham et al. (2008) found that precision grips and finger points or curls cause concurrent activity in the muscles responsible for the articulation of bilabial stops. Vainio et al. (2014) report faster grasp reaction times when pronouncing a "congruent" syllable: 'ka', articulated in the back of the oral cavity, for power grip, and 'ti', which has a front articulation, for precision grip. They conclude that precision grip shares some of the same mechanism with tip-of-the-tongue articulations (e.g. /t/), and with the vocal tract relatively closed (e.g. /i/). Finally, Forrester and Rodriguez (2015) report that in four year olds, performing a concurrent manual task leads to an increased number of tongue protrusions, and that specifically *fine motor control* tasks produce a right-side bias in those protrusions.

9. Summary and general discussion

From our review there emerges a complex picture of the visible movements of the orofacial area. In this light, the gestural scenario of language origins certainly remains a contender. We underscored the rich communicative potential of the orofacial area, and emphasized its flexible voluntary control in nonhuman primates as well as humans, in a way accessible for intentional communication – all of which is at least consistent with a "manual gesture to orofacial gesture to speech" scenario, showing that hominins would have had the substrates for developing an ontogenetically plastic system of visual signalling with the face. We also pointed to the new evidence concerning the connection of orofacial movements to manual action, which lends some plausibility to the speculation about the manual signalling transforming into orofacial signalling. Indeed, several of the findings we cite have been reported specifically in this context, i.e. as supporting the gestural scenario (e.g. Woll 2014).

Nevertheless, the totality of the evidence presented above is difficult to interpret generically as favouring the "orofacial transition" scenario of language origins. Many of the observations are outside of the range of predictions that would seem to follow from this approach. In particular, the most recent studies on the control of the vocal apparatus in non-human primates have dented the conviction that has long been held (cf. Tomasello 2008) that vocal communication in the LCA was likely involuntary and inflexible. In the light of those data, the discontinuity in the production of sound signals between humans and non-human primates might be smaller than previously estimated, which correspondingly reduces the need for postulating an intervening gestural stage in language origins. This is compatible with other recent results that have been taken to support the multimodal position, e.g. a study of four captive chimpanzees showed that manual gesturing caused selective activation in the Broca's area only if the gestures were accompanied by attention getting calls (Taglialatela et al. 2011).

What is uncontroversial is the vital role of visual information from the orofacial area in human communication: nonverbal communication (stand-alone and accompanying language), sign language and speech perception – a point repeatedly made in gestural studies. However, in most cases the visual component can only be singled out analytically, because in the actual communicative acts the vocal and visual information works together and is holistically processed by the receiver as composite audio-visual stimulus. We have stressed the tight coupling of the visual and vocal-auditory modalities in primate communication, including human verbal communication, and the tendency of many orofacial communicative signals to co-occur in both the auditory and visual modalities, leading to a message that is at least partly redundant and therefore amplified and more robust in comprehension. Although this does not necessarily directly favour a multimodal over a gestural account of language origins, it is conducive to a different way of thinking about language, beyond the opposition of gestural versus spoken.

This is not to say that we are in a position to falsify certain predictions of the gestural scenario of language origins; rather, we claim that the multimodal approach is able to better accommodate the evidence coming from the research on the visible movements of the orofacial area. In our view, the better fit results from the multimodal approach being a more realistic description of what communication is in general – a point we developed in section 4. To sum up, ethologically inspired research that stresses ecological validity shows meaning-making (in monkeys, apes and humans) to be a goal-oriented rather than form-oriented process, often improvised in a make-do fashion, with the dynamic use of any semiotic resources currently at hand. All of this promotes a “continuity” scenario, in which first protolinguistic acts (i.e. communicative acts that could be assigned a propositional interpretation) had the form of visuo-vocal ensembles, with both modalities contributing to the propositional interpretation via multiple semiotic resources.

We must not overlook the fact that from an evolutionary perspective, the broad scope of the multimodal perspective at the same time constitutes a potential flaw. The increase of explanatory power that results from marrying the two rival options comes at a price, as it correspondingly lowers falsifiability. It is therefore vitally important for the proponents of multimodal theories to formulate testable predictions, in the sense of specifying ranges of possible data which, if produced by actual research, could potentially falsify their accounts. This could be achieved by developing more refined distinction within the spectrum of multimodal theories, for example between the conception of a single, unified multimodal system (McNeill 2012) and the conception of interacting systems with an initial advantage of vision as proposed by the bodily mimesis theorists, such as Donald (1991) and Zlatev (2014).

We wish to conclude by noting that movements of the orofacial area are an intriguing class of actions, whose significance for language evolution is greater than usually acknowledged. Despite the visible increase of interest in this topic (e.g. Corballis 2003; Meguerditchian et al. 2011; Arbib 2012) it still appears to be underexplored in the language evolution literature, which is evidenced by only marginal treatment in the existing overview works: the *Oxford Handbook of Language Evolution* (ed. Tallerman & Gibson 2012) or the available textbooks (Johansson 2005; Hurford 2007; Fitch 2010). As is so often the case in language evolution, relevant findings come from many distinct disciplines, highlighting the need for interdisciplinary collaboration. The starting point is synthetic and integrative work exemplified in the present paper, which should enable more finely calibrated studies, both experimental and observational. Of course, no single *experimentum crucis* is possible, but the strength of cumulative evidence may ultimately make it possible to adjudicate between the competing models.

Acknowledgements

This research was supported by grant UMO-2012/07/E/HS2/00671 from the Polish National Science Centre.

We wish to thank Michael Corballis, Adrien Meguerditchian and Ines Adornetti for helpful remarks. We are most grateful to Adam Kendon, whose penetrating comments on an earlier draft of this paper were of particular value.

References

Ackermann, Hermann & Wolfram Ziegler (2010). Brain Mechanisms Underlying Speech Motor Control. 2nd edition. In William Hardcastle, John Laver, & Fiona Gibbon, (Eds.), *The handbook of phonetic sciences. 2nd edition* (pp. 202-250). Malden, MA: Wiley-Blackwell.

Andrew, Richard (1963). Evolution of facial expression. *Science*, *142* (3595), 1034-1041.

Arbib, Michael (2002). The mirror system, imitation, and the evolution of language. In Kerstin Dautenhahn & Chrystopher Nehaniv (Eds.), *Imitation in animals and artifacts. Complex adaptive systems* (pp. 229-280). Cambridge, MA: MIT Press.

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

Arbib, Michael (2006). The Mirror System Hypothesis on the linkage of action and Languages. In Michael Arbib (Ed.), *Action to Language via Mirror Neuron System* (pp. 3-47). Cambridge: Cambridge University Press.

Arbib, Michael (2012). *How the brain got language*. Oxford: Oxford University Press.

Armstrong, David, William C. Stokoe & Sherman E. Wilcox (1994). Signs of the origin of syntax. *Current Anthropology, 35(4)*, 349-368.

Armstrong, David, William C. Stokoe & Sherman E. Wilcox (1995). *Gesture and the Nature of Language.* Cambridge: Cambridge University Press.

Armstrong, David & Sherman Wilcox (2007). *The gestural origin of language*. Oxford: Oxford University Press.

Ashcroft, Bill (2001). Language and Race. *Social Identities*, *7*(3), 311–328.

August, Peter & John Anderson (1987). Mammal sounds and motivation-structural rules: A test of the hypothesis. *Journal of Mammology, 68*, 1-9.

Bard, Kim A. (2007). Neonatal imitation in chimpanzees (*Pan troglodytes*) tested with two paradigms. *Animal Cognition*, *102*, 233-242.

Bauer, Harold R. (1987). Frequency code: Orofacial correlates of fundamental frequency. *Phonetica*, *44*, 173-191.

Bavelas, Janet B. & Nichole Chovil (1997). Faces in dialogue. In James A. Russell, & José Miguel Fernández-Dols (Eds.), *The Psychology of Facial Expression* (pp. 334-346). Cambridge: Cambridge University Press.

Bergman, Thore J. (2013). Speech-like vocalized lip-smacking in geladas. *Current Biology*, *23*, R268-R269.

Birdwhistell, Raymond L. (1970). *Kinesics and Context: Essays on Body Motion Communication*. Philadelphia: University of Pennsylvania Press,

Bolhuis, Johan J., Ian Tattersall, Noam Chomsky & Robert C. Berwick (2014). How could language have evolved? *PLOS Biology*, *12*, e1001934.

Bosman, Conrado, Vladimir Lopes & Francisco Aboitz (2005). Sharpening Occam's razor: Is there need for a hand-signing stage prior to vocal communication? [Peer commentary on "From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics" by Michael A. Arbib]. *Behavioral and Brain Sciences, 28*(2), 105–167.

Browman, Catherine P. & Louis Goldstein (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201-251.

Brunner, Lawrence J. (1979). Smiles Can Be Back Channels. *Journal of Personality and Social Psychology*, *37*(5), 728–734.

Buccino, Giovanni, Ferdinand Binkofski & Lucia Riggio (2004). The mirror neuron system and action recognition. *Brain and Language*, *89*, 370-376.

Burling, Robbins (2005). *The talking ape: How language evolved*. Oxford: Oxford University Press.

Call, Josep & Malinda Carpenter (2003). On imitation in apes and children. *Infancia y aprendizaje, 26,* 325-349.

Capek, Cheryl M., Dafydd Waters, Bencie Woll, Mairéad MacSweeney, Michael J. Brammer, Philip K. McGuire, Anthony S. David & Ruth Campbell (2008). Hand and mouth: Cortical correlates of lexical processing in British Sign Language and speechreading English. *Journal of Cognitive Neuroscience*, *20*, 1220-1234.

Chevalier-Skolnikoff, Suzanne (1973). Facial expression of emotion of nonhuman primates. In Paul Ekman (Ed.), *Darwin and facial expression: A century of research in review* (pp. 11–89). New York: Academic Press.

Chibelushi, Claude C., Farzin Deravi & John S. D. Mason (2002). A review of speech-based bimodal recognition. *IEEE Transactions on Multimedia*, *4*, 23-37.

Chovil, Nicole (1991/1992.). Discourse-Oriented Facial Displays in Conversation. *Research on Language and Social Interaction, 25*, 163–194.

Chovil, Nicole (1997). Facing others: A social communicative perspective on facial displays. In James A. Russell & José Miguel Fernández-Dols (Eds.), *The Psychology of Facial Expression* (pp. 321-333). Cambridge: Cambridge University Press.

Clark, Herbert H. & Richard J. Gerrig (1990). Quotations as demonstrations. *Language, 66*, 764-805.

Clark, Nathaniel & Marcus Perlman (2014). Breath, vocal and supralaryngeal flexibility in a human-reared gorilla. In Bart de Boer & Tessa Verhoef (Eds.), *Proceedings of the Workshop on Evolution of Signals, Speech, and Signs. Evolang Press* (pp. 5-9). Vienna.

Clay, Zanna & Klaus Zuberbühler (2014). Vocal communication and social awareness in chimpanzees and bonobos: insights from studies of vocal communication. In Daniel Dor, Chris Knight, & Jerome Lewis (Eds.), *The Social Origins of Language* (pp. 141-156). Oxford: Oxford University Press.

Collins, Christopher (2013). *Paleopoetics. The evolution of the literary imagination*. New York: Columbia University Press.

Corballis, Michael C. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.

Corballis, Michael C. (2003). From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences, 26*, 199-208.

Corballis, Michael C. (2012). How language evolved from manual gestures. *Gesture*, *12*, 200-226.

Coudé, Gino, Pier Francesco Ferrari, Francesco Rodà, Monica Maranesi, Eleonora Borelli, Vania Veroni, Fabio Monti, Stefano Rozzi & Leonardo Fogassi (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS One*, *6*, e26822.

Cowley, Stephen J. (2011). Distributed language. In Stephen J. Cowley (Ed.), *Distributed language* (pp. 1-14). Amsterdam: John Benjamins.

Darwin, Charles (1872). *The expression of the emotions in man and animals*. London: Murray.

Donald, Merlin (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Cambridge: Harvard University Press.

Dor, Daniel, Chris Knight & Jerome Lewis (2014). Introduction: a social perspective on how language began. In Daniel Dor, Chris Knight & Jerome Lewis (Eds.), *The Social Origins of Language* (pp. 1-14). Oxford: Oxford University Press.

Diamond, Arthur (1959). *The History and Origin of Language*. London: Methuen.

Dunbar, Robin (1996). *Grooming, gossip and the evolution of language*. London: Faber & Faber.

Eberl, Eva (2010). Control of gestures and vocalizations in primates. In Stefan Brudzynski (Ed.), *Handbook of Mammalian Vocalization* (pp. 469-476). Amsterdam: Academic Press.

Eccardt, Thomas (2006). The case for articulatory gestures-not sounds-as the physical embodiment of speech signs. *Studies in Functional and Structural Linguistics, 57*, 283-308.

Ekman, Paul (1979). About brows: Emotional and conversational signals. In Mario von Cranach, Klaus Foppa, Wolfgang Lepenies & Detlev Ploog (Eds.), *Human ethology* (pp. 169-202). Cambridge: Cambridge University Press.

Ekman, Paul & Wallace V. Friesen (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*, 49-98.

Ekman, Paul, Wallace V. Friesen & Joseph C. Hager (2002). *Facial Action Coding System. The Manual*. Salt Lake City: Research Nexus.

Enard, Wolfgang, Molly Przeworski, Simon E. Fisher, Cecilia S. L. Lai, Victor Wiebe, Takashi Kitano, Anthony P. Monaco & Svante Pääbo (2002). Molecular evolution of *FOXP*2, a gene involved in speech and language. *Nature, 418,* 869-872.

Fedurek, Pawel, Katie E. Slocombe, Jessica A. Hartel & Klaus Zuberbühler (2015). Chimpanzee lip-smacking facilitates cooperative behaviour. *Scientific Reports, 5*, 13460.

Ferrari, Pier F., Vittorio Gallese, Giacomo Rizzolatti & Leonardo Fogassi (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17*, 1703-1714.

Fitch, W. Tecumseh (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Fitch, W. Tecumseh, Mark D. Hauser & Noam Chomsky (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, *97*, 179–210.

Forrester, Gillian & Alina Rodriguez (2015). Slip of the tongue: Implications for evolution and language development. *Cognition*, *141*, 103-111.

Fridlund, Alan J. (1997). The new ethology of facial expressions. In James A. Russell & José Miguel Fernández-Dols (Eds.), *The Psychology of Facial Expression* (pp. 103-129). Cambridge: Cambridge University Press.

Galantucci, Bruno, Carol A. Fowler & M. T. Turvey (2006). The motor theory of speech perception reviewd. *Psychonomic Bulletin and Review, 13*(3), 361–377.

Gardner, R. Allen & Beatrice T. Gardner (1969). Teaching Sign Language to a Chimpanzee. *Science, 165*(3894), 664–672.

Gazzola, Valeria, Lisa Aziz-Zadeh & Christian Keysers (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, *16*, 1824-1829.

Gentilucci, Maurizio & Michael C. Corballis (2006). From manual gesture to speech: A gradual transition. *Neuroscience and Biobehavioral Reviews*, *30*, 949-960.

Gentilucci, Maurizio & Riccardo Dalla Volta (2008). Spoken language and arm gestures are controlled by the same motor control system. *The Quarterly Journal of Experimental Psychology*, *61*, 944-957.

Ghazanfar, Asif A. & Nikos K. Logothetis (2003). Neuroperception: Facial expressions linked to monkey calls. *Nature*, *423*, 937-938.

Ghazanfar, Asif & David Lewkowicz (2008). The Ontogeny and Phylogeny of Bimodal Primate Vocal Communication. In Nobuo Masataka (Ed.), *The Origins of Language* (pp. 85-110). Tokyo: Springer.

Ghazanfar, Asif A., Daniel Y. Takahashi, Neil Mathur, & Tecumseh Fitch (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current Biology*, *22*, 1176-1182.

Gillespie-Lynch, Kristen, Patricia M. Greenfield, Heidi Lyn & Sue Savage-Rumbaugh (2014). Gestural and symbolic development among apes and humans: support for a multimodal theory of language evolution. *Frontiers in Psychology*, *5*, 1228.

Goffman, Erving (1959). *The Presentation of Self in Everyday Life.* Garden City: Doubleday Anchor Books.

Goffman, Erving (1963). *Behavior in Public Places: Notes on the Social Organization of Gatherings.* New York and London: The Free Press of Glencoe.

Goffman, Erving (1969). *Strategic Interaction.* Philadelphia: University of Pennsylvania.

Gogate, Lakshmi J. & Lorraine E. Bahrick (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*, 133-149.

Goldin-Meadow, Susan & Melissa A. Singer (2003). From children's hands to adults' ears: Gesture's role in teaching and learning. *Developmental Psychology, 39*(3), 509-520

Goldin-Meadow, Susan, Wing Chee So, Asli Özyürek & Carolyn Mylander (2008). The natural order of events: How speakers of different languages represent events nonverbally. *PNAS, 105*(27), 9163-9168.

Goldin-Meadow, Susan (2014). Widening the lens: what the manual modality reveals about language, learning and cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369,* 20130295.

Goodall, Jane (1986). *The chimpanzees of Gombe*. Boston: Houghton Mifflin Publishing.

Grice, H. Paul (1975). Logic and conversation. In Peter Cole & Jerry L. Morgan (Eds.) *Syntax and semantics, Vol. 3: Speech acts* (pp. 43–58). New York: Academic Press.

Halliday, Michael A. K. (1973). *Explorations in the Functions of Language*. London: Edward Arnold.

Harpham, Geoffrey G. (2009). Roots, Races, and the Return to Philology. *Representations,* *106*(1), 34–62.

Hauser, Marc D. (1993). The evolution of nonhuman primate vocalizations: Effects of phylogeny, body weight and motivational state. *American Naturalist*, *142*, 528-542.

Hauser, Marc D. (1997). *The Evolution of Communication*. Cambridge, MA: MIT Press.

Hauser, Mark. D., Noam Chomsky & W. Tecumseh Fitch (2002). The Faculty of Language: What Is It , Who Has It , and How Did It Evolve? *Science*, *298*, 1569–1579.

Hawhee, Debra (2006). Language as Sensuous Action: Sir Richard Paget, Kenneth Burke, and Gesture-Speech Theory. *Quarterly Journal of Speech*, *92*, 331-354.

Hewes, Gordon W. (1973). Primate communication and the gestural origin of language. *Current Anthropology*, *14*, 5-24.

Hewes, Gordon W. (1975). *Language origins: a bibliography*. The Hague: Mouton.

Hewes, Gordon W. (1976). The current status of the gestural theory of language origin. *Annals of the New York Academy of Sciences*, *280*, 482-504.

Hewes, Gordon W. (1977a). A Model for Language Evolution. *Sign Language Studies, 15*, 97-168.

Hewes, Gordon W. (1977b). Language Origin Theories. In Duane Rumbaugh (Ed.), *Language learning by a chimpanzee: The Lana Project* (pp. 3-53). New York: Academic Press.

Hewes, Gordon W. (1996). A history of the study of language origins and the gestural primacy hypothesis. In Andy Lock & Charles Peters (Eds.), *Handbook of human symbolic evolution* (pp. 263-269). Oxford: Oxford University Press.

Higginbotham, David R., Matthew I. Isaak & James N. Domingue (2008). The exaptation of manual dexterity for articulate speech: an electromyogram investigation. *Experimental Brain Research*, *186*, 603-609.

Hinde, Robert & Thelma Rowell (1962). Communication by postures and facial expressions in the rhesus monkey (*Macaca mulatta*). *Proceedings of the Zoological Society of London*, *138* (1), 1-21.

Hobaiter, Catherine & Richard, W. Byrne (2012). Gesture use in Consortship: wild chimpanzees' use of gesture for an 'evolutionarily urgent' purpose. In Simone Pika & Katja Liebal (Eds.), *Developments in Primate Gesture Research* (pp.127-144). Amsterdam: John Benjamins Press.

Hockett, Charles. F. (1958). *A Course in Modern Linguistics*. New York: Macmillan.

Hockett, Charles. F. (1959). Animal "languages" and human language. *Human Biology 31*, 32–39.

Hockett, Charles F. (1960a). The origin of speech. *Scientific American*, *203,* 88–111.

Hockett, Charles F. (1960b). Logical considerations in the study of animal communication. In Wesley Lanyon &William Tavolga (Eds.), *Animal sounds and communication* (pp. 392–430). Washington: American Institute of Biological Sciences.

Hockett, Charles. F. (1966). The problem of universals in language. In Jospeh Greenberg (Ed.) *Universals of Language*, (pp. 1-29). Cambridge, MA: The MIT Press.

Hockett, Charles F. & Stuart A. Altmann (1968). A note on design features. In Thomas A. Sebeok (Ed.), *Animal Communication: techniques of study and results of research*, (pp. 61-72). Bloomington: Indiana University Press.

Hooff, van, Jan (1962). Facial expressions in higher primates. *Symposia of the Zoological Society of London*, *8*, 97-125.

Hooff, van, Jan (1967). The facial displays of the catarrhine monkeys and apes. In Desmond Morris (Ed.), *Primate Ethology* (pp. 7-68). Chicago: Aldine.

Hooff, van, Jan (1972). A comparative approach to the phylogeny of laughter and smiling. In: Robert Hinde (Ed.), *Non-Verbal Communication* (pp. 209-238). Cambridge: Cambridge University Press.

Hurford, James R. (2007). *The Origins of Meaning*. Oxford: Oxford University Press.

Imai, Mitsumi & Sotaro Kita (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130298.

Jiang, Jintao, Abeer Alwan, Patricia A. Keating, Edward T. Auer Jr. & Lynne E. Bernstein (2002). On the relationship between face movements, tongue movements, and speech acoustics. *EURASIP Journal on Applied Signal Processing*, *11*, 1174-1188.

Jóhanneson, Alexander (1949). *Origin of Language. Four Essays*. Reykjavik: Leiftur.

Johansson, Sverker (2005). *Origins of Language. Constraints and Hypothesis*. Amsterdam/Philadeplphia: John Benjamins Publishing Company.

Kegl, Judy, Ann Senghas & Marie Coppola (1999). Creation through contact: Sign language emergence and sign language change in Nicaragua. In Michel DeGraff (Ed.), *Language Creation and Language Change: Creolization, Diachrony, and Development* (pp. 179-237). Cambridge, MA: MIT Press.

Kendon, Adam (1972). Some relationships between body motion and speech. An analysis of an example. In Aron Wolfe Siegman & Benjamin Pope (Eds.), *Studies Dyadic Communication* (pp. 177-210). New York: Pergamon Press.

Kendon, Adam (1991). Some considerations for a theory of language origins. *Man*, *26*, 199-221.

Kendon, Adam (2004). *Gesture. Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kendon, Adam (2008). Signs for Language Origins? *The Public Journal of Semiotics*, *2*, 2-29.

Kendon, Adam (2011). 'Gesture First' and 'Speech First' in Theories of Language Origins. In Gaurav Mathur and Donna Jo Napoli (Eds.), *Deaf Around the World* (pp. 251-267). New York: Oxford University Press.

Kendon, Adam (2014a). The 'poly-modalic' nature of utterances and its relevance for inquiring into language origins. In Daniel Dor Chris Knight, & Jerome Lewis (Eds.), *The Social Origins of Language* (pp. 67-76). Oxford: Oxford University Press.

Kendon (2014b) Semiotic diversity in utterance production and the concept of 'language'. Phil. Trans. Roy. Soc., series B, **369**: 20130293

Kohler, Evelyne, Christian Keysers, M. Alessandra Umilta, Leonrado Fogassi, Vittorio Gallese & Giacomo Rizzolatti (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, *297*, 846-848.

Lakoff, George & Mark Johnson (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, George & Mark Johnson (1999). *Philosophy in the Flesh*. New York: Basic Books.

Langacker, Ronald (1987). *Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites*. Stanford: Stanford University Press.

Leavens, David A., Jamie L. Russell & William D. Hopkins (2010). Multimodal communication by captive chimpanzees *(Pan troglodytes). Animal Cognition*, *13*, 33–40.

Leavens, David, Jared Taglialatela &William Hopkins (2014). From grasping to grooming to gossip: innovative use of chimpanzee signals in novel environments supports both vocal and gestural theories of language origins. In Marco Pina & Nathalie Gontier (Eds.), *The evolution of social communication in primates: A multidisciplinary approach* (pp. 179-194). New York: Springer.

Le Cabec, Adeline, Philipp Gunz, Kornelius Kupczik, José Braga & Jean-Jacques Hublin (2013). Anterior tooth root morphology and size in Neanderthals: Taxonomic and functional implications. *Journal of Human Evolution*, *64*, 169-193.

Levinson, Steven C. & Judith Holler (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 201303302.

Lewis, Jerome (2014). BaYaka Pygmy multi-modal and mimetic communication traditions. In Daniel Dor, Chris Knight & Jerome Lewis (Eds.), *The Social Origins of Language* (pp. 77-91). Oxford: Oxford University Press.

Liberman, Alvin, Franklin Cooper, Donald Shankweiler & Michael Studdert-Kennedy, (1967). Perception of the speech code. *Psychological Review*, *74*, 431-461.

Liberman, Alvin M. & Ignatius G. Mattingly (1985). The motor theory of speech perception revisited. *Cognition*, *21*, 1-36.

Liberman, Alvin M. & Doug H. Whalen (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196.

Liebal, Katja , Bridget M. Waller, Anne M. Burrows & Katie E. Slocombe (2014). *Primate Communication: A Multimodal Approach*. Cambridge and New York: Cambridge University Press.

Linell, Per (2005). *The Written Language Bias in Linguistics: Its Nature, Origins, and Transformations*. London: Routledge.

Liszkowski, Ulf (2014). Two sources of meaning in infant communication: preceding action contexts and act-accompanying characteristics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369*, 20130294.

Little, Anthony C., Benedict C. Jones, Corri Waitt, Bernard P. Tiddeman, David R. Feinberg, David I. Perrett, Coren L. Apicella & Frank W. Marlowe (2008). Symmetry is related to sexual dimorphism in faces: data across culture and species. *PLoS ONE*, *3*, e2106.

Little, Anthony C., Benedict C. Jones, & Lisa M. DeBruine (2011). Facial attractiveness: evolutionary based research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1638-1659.

MacNeilage, Peter F. (2008). *The Origin of Speech*. Oxford: Oxford University Press.

MacNeilage, Peter F. & Barabara L. Davies (2005). Evolutionary sleight of hand: Then, they saw it; now we don’t. *Behavioral and Brain Sciences*, *28*, 153-163.

Matsumoto-Oda, Akiko & Masaki Tomonaga (2005). "Intentional" control of sound production found in leaf-clipping display of Mahale chimpanzees. *Journal of Ethology*, *23*, 109-112.

McNeill, David (1992). *Hand and Mind*. Chicago: University of Chicago Press.

McNeill, David (2000). Growth points in thinking-for-speaking. In David McNeill (Ed.) *Language and Gesture* (pp. 141-161). Cambridge: Cambridge University Press.

McNeill, David (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

McNeill, David (2012). *How Language Began. Gesture and Speech in Human Evolution*. Cambridge: Cambridge University Press.

McNeill, David (in press). *Why We Gesture: The surprising role of the hands in communication*. Cambridge: Cambridge University Press.

McNeill, David, Susan D. Duncan, Jonathan Cole, Shaun Gallagher & Bennett Bertenthal (2008). Growth points from the very beginning. *Interaction Studies*, *9*, 117–132.

Marcus, Garry F. & Simon E. Fisher (2003). *FOXP2* in focus: what can genes tell us about speech and language? *Trends in Cognitive Sciences*, *7*, 257-262.

Maynard Smith, John & David Harper (2003). *Animal Signals*. Oxford: Oxford University Press.

McGurk, Harry & John MacDonald (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

Meguerditchian, Adrien, Hélène Cochet & Jacques Vauclair (2011). From gesture to language: ontogenetic and phylogenetic perspectives on gestural communication and its cerebral lateralization. In Anne Vilain, Jean-Luc Schwartz, Christian Abry & Jacques Vauclair (Eds.), *Primate communication and human language: vocalization, gestures, imitation and deixis in humans and non-humans* (pp. 89-118). Amsterdam: John Benjamins.

Meguerditchian, Adrien, Marie Plouvier, Jill Pruetz & William Hopkins (2014). From hand to mouth: fine precision grip during mutual grooming elicited wide lip movements in wild fongoli chimpanzees. In Erica Cartmill, Sean Roberts, Heidi Lyn & Hannah Cornish (Eds.), The Evolution of Language. *Proceedings of the 10th International Conference (EVOLANG 10)* (pp. 487-488). Singapore: World Scientific.

Meguerditchian, Adrien & Jacques Vauclair (2014). Communicative Signaling, Lateralization and Brain Substrate in Nonhuman Primates: Toward a Gestural or a Multimodal Origin of Language? *Humana Mente Journal of Philosophical Studies, 27,* 135-160.

Morton, Eugene S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist, 111,* 855-869.

Mühlenbernd, Roland, Dankmar Enke, Matthias Villing, Natalja Gavrilov & Jonas David Nick (2014). Modality switch in human language evolution. In Erica A. Cartmill, Seán Roberts, Heidi Lyn & Hannah Cornish (Eds.), *Evolution of Language. Proceedings of the Tenth International Conference EVOLANG* (pp. 161-168). Singapore: World Scientific.

Ohala, John (1983). Cross-language use of pitch: an ethological view. *Phonetica, 40* (1), 1-18.

Ohala, John (1994). The frequency code underlies the sound-symbolic use of voice-pitch. In Leanne Hinton, Johann Nichols & John Ohala (Eds.), *Sound Symbolism* (pp. 325-347). Cambridge: Cambridge University Press.

Orzechowski, Sylwester, Sławomir Wacewicz, & Przemysław Żywiczyński (2014). Orofacial gestures in language evolution. The auditory feedback hypothesis. In Erica Cartmill, Sean Roberts, Heidi Lyn & Hannah Cornish (Eds.), *The*

Evolution of Language. Proceedings of the 10th International Conference (EVOLANG 10) (pp. 221-227). Singapore: World Scientific.

Özyürek, Asli. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130296.

Paget, Richard A. S. (1930). *Human Speech: Some Observations, Experiments, and Conclusions as to the Nature, Origin, Purpose and Possible Improvement of Human Speech*. London: Kegan Paul & Trench Trübner.

Paget, Richard A. S. (1944). The origin of language, *Science*, *99*, 14-15.

Parr, Lisa A. (2004). Perceptual biases for multimodal cues in chimpanzee *(Pan troglodytes)* affect recognition. *Animal Cognition*, *7*, 171-178.

Parr, Lisa A. & Frans B. M. de Waal (1999). Visual kin recognition in chimpanzees. *Nature*, *399*, 647–648.

Parr, Lisa A. & Bridget M. Waller (2006). Understanding chimpanzee facial expression: insights into the evolution of communication. *Social Cognitive and Affective Neuroscience*, *1*, 221-228.

Parr, Lisa A., Matthew Heintz, Elizabeth Lonsdorf & Emily Wroblewski (2010). Visual kin recognition in nonhuman primates: (*Pan troglodytes* and *Macaca mulatta*): inbreeding avoidance or male distinctiveness? *Journal of Comparative Psychology*, *124*, 343-350.

Petrides, Michael, Geneviève Cadoret & Scott Mackey (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature*, 435, 1235-1238.

Petrides, Michael & Deepak N. Pandya (2009). Distinct Parietal and Temporal Pathways to the Homologues of Broca's Area in the Monkey. *PLoS Biol, 7,* e1000170.

Pinker, Steven (1994). *The Language Instinct*. New York: Harper Perennial Modern Classics.

Pyers, Jennie E. & Karen Emmorey (2008). The face of bimodal bilingualism: Grammatical markers in American Sign Language are produced when bilinguals speak to English monolinguals. *Psychological Science, 19*, 531–535.

Ravizza, Susan (2005). Neural Regions Associated with Categorical Speech Perception and Production. In Henri Cohen & Claire Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science* (pp. 601-615). Amsterdam: Elsevier.

Ricci Bitti, Pio & Isabella Poggi, (1991). Symbolic nonverbal behavior: Talking through gestures. In Robert Feldman & Bernard Rime (Eds.), *Fundamentals of Nonverbal Behavior* (pp. 433-457). Cambridge: Cambridge University Press.

Roll, Patrice, Gabrielle Rudolf, Sandrine Pereira, Barbara Royer, Ingrid E. Scheffer, Annick Massacrier, Maria-Paola Valenti, Nathalie Roeckel-Trevisiol, Sarah Jamali, Christophe Beclin, Caroline Seegmuller, Marie-Noëlle Metz-Lutz, Arnaud Lemainque, Marc Delepine, Christophe Caloustian, Anne de Saint Martin, Nadine Bruneau, Danièle Depétris, Marie-Geneviève Mattéi, Elisabeth Flori, Andrée Robaglia-Schlupp, Nicolas Lévy, Bernd A. Neubauer, Rivka Ravid, Christian Marescaux, Samuel F. Berkovic, Edouard Hirsch, Mark Lathrop, Pierre Cau & Pierre Szepetowski (2006). *SRPX2* mutations in disorders of language cortex and cognition. *Human Molecular Genetics, 15*, 1195-1207.

Sandler, Wendy (2009). Symbiotic symbolization by hand and mouth in sign language. *Semiotica*, *2009* (174), 241-275.

Sandler, Wendy (2013). Vive la différence: Sign language and spoken language in language evolution. *Language and Cognition*, *5*, 189–204.

Sereno, Martin (2014). Origin of symbol-using systems: speech, but not sign, without the semantic urge. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 20130303.

Sia, Gek-Ming, Roger Clem & Richard Huganir (2013). The Human Language-Associated Gene *SRPX2* Regulates Synapse Formation and Vocalization in Mice. *Science, 342* (6161), 987-991.

Skipper, Jeremy (2014). Echoes of the spoken past: how auditory cortex hears context during speech perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369*, 20130297.

Slocombe, Katie E., Bridget M. Waller & Katja Liebal (2011). The language void: the need for multimodality in primate communication research. *Animal Behaviour, 81*, 919-924.

Sperber, Dan & Deirdre Wilson, D. (1986) *Relevance: Communication and cognition*. Cambridge: Harvard University Press/Blackwell.

Stam, James H. (1976). *Inquiries into the Origin of Language: the Fate of a Question.* New York: Harper and Row, Publishers.

Steklis, Horst & Stevan R. Harnad (1976). From hand to mouth: Some critical stages in the evolution of language. *Annals of the New York Academy of Sciences, 280,* 445-455.

Studdert-Kennedy, Michael (2002). Mirror neurons, vocal imitation and the evolution of particulate principle. In Maxim Stamenov & Vittorio Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 207-227). Amsterdam: John Benjamins,.

Sueyoshi, Ayano & Debra M. Hardison (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55, 661–699.

Summerfield, Quentin (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*, 71-78.

Sutton, Dwight (1979). Mechanisms underlying vocal control in nonhuman primates. In Horst Steklis & Michael Raleigh (Eds.), *Neurobiology of social communication in primates: An evolutionary perspective* (pp. 45-68). New York: Academic Press.

Sutton-Spence, Rachel (2005). *Analysing Sign Language Poetry*. New York: Palgrave Macmillan.

Swadesh, Morris (1971). *The Origin and Diversification of Language*. Chicago: Alidne & Atherton.

Sweet, Henry (1888*). A history of English sounds from the earliest period, with full word-lists*. Oxford: Clarendon (retrieved from https://archive.org/details/ahistoryenglish05sweegoog).

Taglialatela, Jared P., Jamie L. Russell, Jennifer A. Schaeffer & William D. Hopkins (2011) Chimpanzee Vocal Signaling Points to a Multimodal Origin of Human Language. *PLoS ONE* 6(4): e18852.

Taglialatela, Jared P., Jamie L. Russell, Sarah M., Pope, Tamara Morton, Stephanie Bogart, Lisa A. Reamer, Steven J. Schapiro & William D. Hopkins (2015). Multimodal communication in chimpanzees. *American Journal of Primatology*, 1098-2345.

Tallerman, Maggie (2012). Protolanguage. In Maggie Tallerman & Kathleen Gibson (Eds.), *The Oxford Handbook of Language Evolution* (pp. 479-491). Oxford: Oxford University Press.

Tallerman, Maggie, Kathleen Gibson (Eds.), (2012). *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press.

Tomasello, Michael. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Vainio, Lari, Mikko Tiainen, Kaisa Tiippana & Martti Vainio, (2014). Shared processing of planning articulatory gestures and grasping. *Experimental Brain* Research, 232(7), 2359-2368.

Vargha-Khadem, Faraneh, Kate E. Watkins, C. J. Price, John Ashburner, Katie J. Alcock, Alan Connelly, Richard S. J. Frackowiak, Karl J. Friston, M. E. Pembrey, Mortimer Mishkin, David G. Gadian & Richard E. Passingham (1998). Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences, 95* (21), 12695-12700.

Vigliocco, Gabriella, Pamela Perniss, & David Vinson (2014). Introduction: Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369*, 20130292.

Vygotsky, Lev S. (1978). *Mind in Society. The Development of Higher Psychological*

Processes. Michael Cole, Vera John-Steiner, Sylvia Scribner & Ellen Souberman (Eds.). Cambridge: Harvard University Press.

Vygotsky, Lev S. (1987). *The Collected Works of L.S. Vygotsky. Volume 1. Problems of General Psychology*. Robert W. Rieber & Aaron S. Carton (Eds.). New York: Plenum Press.

Wacewicz, Sławomir & Przemysław Żywiczyński (2015a). Language Evolution: Why Hockett's Design Features are a Non-Starter. *Biosemiotics, 8*(1), 29-46.

Wacewicz, Sławomir & Przemysław Żywiczyński (2015b). From the narrow to the broad. Multiple perspectives on language evolution. *Theoria et Historia Scientiarum, 11*, 5-18.

Wallace, Alfred .R. (1881). Review of Tylor's Anthropology. *Nature*, *24*, 242-245.

Wallace, Alfred R. (1895). The Expressiveness of speech, or mouth gesture as a factor in the evolution of speech. *Fortnightly Review*, *58*, 528-543.

Waters, Gabriel S. & Roger S. Fouts (2002). Sympathetic mouth movements accompanying fine motor movements in chimpanzees *(Pan troglodytes)* with implications toward the evolution of language. *Neurological Research, 24*, 174-180.

Watkins, Kate & Tomáš Paus (2004). Modulation of motor excitability during speech perception: the role of Broca's area. *Journal of Cognitive Neuroscience*, *16*, 978-987.

Watts, Richard (2003). *Politeness*. Cambridge: Cambridge University Press.

Watson, Stuart K., Simon W. Townsend, Anne M. Schel, Claudia Wilke, Emma K. Wallace, Leveda Cheng, Victoria West & Katie Slocombe (2015). Vocal

Learning in the Functionally Referential Food Grunts of Chimpanzees. *Current Biology*, *25*, 495-499.

Woll, Bencie (2001). The sign that dares to speak its name: echo phonology in British Sign Language (BSL). In Penny Boyes-Braem & Rachel Sutton-Spence (Eds.), *The Hands are the Head of the Mouth* (pp. 87-98). Hamburg: Signum Press.

Woll, Bencie (2014). Moving from hand to mouth: echo phonology and the origins of language. *Frontiers in Psychology*, *5*, article 662.

Yehia, Hani, Philip Rubin, & Eric Vatikiotis-Bateson (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication*, *26*, 23-43.

Yehia, Hani. C., Takaaki Kuratate, & Eric Vatikiotis-Bateson (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, *30*, 555-568.

Zimmermann, Elke, Lisette Leliveld, & Simone Schehka (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices. In Eckart Altenmüller, Sabine Schmidt & Elke Zimmermann (Eds.), *Evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man* (pp. 116-132). Oxford: Oxford University Press.

Zlatev, Jordan (2014). Human Uniqueness, Bodily Mimesis and the Evolution of Language. *Humana Mente. Journal of Philosophical Studies*, *27*, 197-219.