

Connection between movements of mouth and hand: Perspectives on development and evolution of speech

Vainio, L.

Lari Vainio (1,2,3) – the corresponding author; Email: lari.vainio@helsinki.fi; Tel: +358-2-94129392

1 - University of Helsinki, Helsinki Collegium for Advanced Studies, P.O. Box 4 (Fabianinkatu 24),
FIN 00014 University of Helsinki, Finland.

2 - Perception, Action & Cognition Research Group, Department of Psychology and Logopedics,
Faculty of Medicine, University of Helsinki, Haartmaninkatu 8, 00014, Finland

3 - Phonetics and speech synthesis research group, Department of Digital Humanities, University of
Helsinki, Unioninkatu 40, 00014, Finland

Abstract

Mounting evidence shows interaction between manipulative hand movements and movements of tongue, lips and mouth in a vocal and non-vocal context. The current article reviews this evidence and discusses its contribution to perspectives of development and evolution of speech. In particular, the article aims to present novel insight on how processes controlling the two primary grasp components of manipulative hand movements, the precision and power grip, might be systematically connected to motor processes involved in producing certain articulatory gestures. This view assumes that due to these motor overlaps between grasping and articulation, development of these grip types in infancy can facilitate development of specific articulatory gestures. In addition, the hand-mouth connections might have even boosted the evolution of some articulatory gestures. This account also proposes that some semantic sound-symbolic pairings between a speech sound and a referent concept might be partially based on these hand-mouth interactions.

Keywords: speech, articulation, manual movements, grasping, speech development, speech evolution

1. Introduction

Traditional investigations related to interaction between speech and manual processes have been largely carried out by observing hand movements during communication between individuals revealing systematic occurrence of different kinds of communicative manual gestures during speech production (McNeill, 1992). The phenomenon related to mouth-hand interaction that is also acknowledged by researchers, but that has received less empirical attention, is that non-communicative movements of mouth, tongue and lips can accompany hand movements. As a familiar example, Darwin (1872) noted that “persons cutting anything may be seen to move their jaws simultaneously with the blades of the scissors”, and that “children learning to write often twist about their tongues as their fingers move...” (Darwin, 1872, p.34). Particularly, theories of language evolution have often emphasized theoretical significance of this involuntary tendency to produce mouth actions during, for example, manual manipulation. For instance, Paget (1944) and Hewes (1973) have proposed that this kind of mimicking of hand movements with the mouth, tongue and lips might have contributed to evolution of speech. As a continuation for this view, the present article assumes that understanding the mechanisms that connect non-communicative mouth and hand actions shed light on integrated mechanisms between speech and hand movements and also contribute to understanding evolutionary and developmental aspects of speech and language. As such, this article, firstly, reviews evidence on the connections between mouth movements and non-communicative hand movements such as grasping, and secondly it discusses the potential relevance of these findings in light of views related to language evolution and development as well as sound symbolism. In particular, the article aims to present novel perspectives on how connections between mouth motor programs controlling specific articulatory gestures and grasp motor programs controlling the precision and power grip might play a facilitatory role in development and evolution of certain articulatory gestures.

1.1. Evidence linking mouth movements to manipulative hand movements

Individuals appear to have a tendency to accompany manipulative hand movements with movements of mouth, tongue and lips. As an example, most of us are familiar with the phenomenon that while performing a manual task that requires high precision manual dexterity, such as threading a needle, tongue tip is pressed between the lips or the protruded tongue is moved in synchrony with the manipulative hand movements. Corresponding phenomena have been also observed in chimpanzees. Waters and Fouts (2002) observed that captive chimpanzees increasingly perform protrusion and compression of the lips and tongue during fine manual manipulation requiring precision grip. In addition, lip-smacking gestures of chimpanzees are more frequently observed during manual grooming actions, requiring utilization of fine precision grip, than during any other grooming-related actions such as hair sweeping or hair holding (Meguerditchian, Plouvier, Pruetz, & Hopkins, 2014).

In humans, Forrester and Rodriguez (2015) have shown in typically developing 4-year old children that “tongue thrusts” (i.e., protrusion of the tongue tip between incisors or lips while performing demanding and precise manual tasks) are elicited by motor tasks consisting of fine (e.g., requiring object manipulation using precision grip) or gross (e.g., requiring effortful spatial manipulation) manual actions. In general, these tongue thrusts tend to be commonly observable in early childhood but extinguish in later childhood (Mason & Proffitt, 1974). In addition, electromyographic (EMG) and magnetoencephalographic (MEG) evidence is in line with the view that motor processes involved in protruding lips and tongue are connected to motor processes responsible for moving hands. For example, Higginbotham, Isaak and Domingue (2008) found increased EMG responses of the orbicularis oris muscle—linked to the articulation of bilabial stops and round vowels—during the curl gesture, precision grasping, and manual pointing. In addition, Salmelin and Sams (2002) have shown MEG evidence that the hand motor cortex is significantly involved in non-verbally produced lip (lip protrusion) and tongue (touching the upper teeth with the tip of the tongue) movements. Hence, a variety of evidence shows that hand motor processes are connected to processes that program lip and

tongue-related actions in a non-communicative context. Furthermore, processes related to fine manual control are not only linked to involuntary non-vocal mouth movements, but they are also linked to motor processes responsible for moving speech organs for articulation. Indeed, numerous studies have shown the functional connection between development of articulatory movement skills and development of fine motor skill such as fastening buttons and threading beads (e.g., Zelaznik & Goffman, 2010; Owen & McKinlay, 1997). Hence, processes responsible for controlling manipulative hand movement, such as fine manual grasping, appear to be connected, at least to some extent, to vocal and non-vocal movements of mouth, tongue and lips, and according to some accounts these connections might even play some role in speech development.

1.2. Neural evidence linking grasping to mouth movements

Manual grips that form the central components of manipulative movements can be in general divided into precision and power grips on the basis of functional, phylogenetic and developmental considerations (Halverson, 1931; Napier, 1956). The precision grip is formed by pinching small objects between the tips of the index finger and thumb, whereas the power grip is formed by grasping and holding large objects using a clamp shaped by the flexed fingers and the palm. Some of the evidence discussed above suggests that not only manipulative hand movements in general but grasping in particular might be to some extent connected to motor processes controlling mouth actions (e.g., Waters & Fouts, 2002; Higginbotham et al., 2008). However, the first evidence directly recorded from single neurons, which clearly demonstrates integrated planning of grasping and mouth movements, was reported by Rizzolatti, Camarda, Fogassi, Gentilucci, Luppino, and Matelli (1988). The study showed that macaque monkeys' premotor area F5 (the ventral premotor cortical area 6) includes neurons that discharge when the animal grasps the object with the mouth or hand. Although the effect did not depend on synergisms between hand and mouth movements (i.e., the same neuron fired even when the

hand and mouth grasps were performed separately), it was proposed that the effect primarily reflects motor processes that coordinate hand-to-mouth movements for self-feeding behavior. This interpretation is supported by the finding showing that electrical micro-stimulation of this same area in monkeys triggers the hand to shape into a grip posture, the mouth to open, and move the hand to the mouth (Graziano, Taylor & Moore, 2002).

The other single-cell recoding evidence revealing integration between manual grasping and mouth movements comes from mirror neuron investigations showing that some neurons of the monkey premotor F5 area represent observed grasping actions regardless of whether the action is performed with the hand or mouth (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). Briefly, mirror neuron studies have shown that action observation, such as observing someone else grasping an object, triggers activation in the neurons of the premotor region F5 of a monkey involved in executing the same action (Di Pellegrin, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). Subsequently, for example, brain imaging (Iacoboni, Woods, Brass, Bekkering, Mazziotta, & Rizzolatti, 1999) and single-cell recording (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010) studies have shown that similar mirror neuron processes can also be observed in humans. Moreover, some mirror neurons in monkeys encode actions in a strictly congruent manner (Rizzolatti, Fogassi, & Gallese, 2001). For example, perceiving precision grasping activates the motor representation that is involved in producing the precision grasp whereas perceiving power grasping activates the motor representation that is involved in producing the power grasp. In addition, Ferrari, Gallese, Rizzolatti, and Fogassi (2003) showed that some mirror neurons are associated with lip-smacking behavior, which is the most common communicative mouth gesture in monkeys, while some mirror neurons are related to eating-related mouth movements. Additionally, in the macaque monkey, electrical stimulation of this same area has been found to trigger involuntary hand and lip movements (Petrides, Cadoret, & Mackey, 2005). Hence, it seems that this premotor region F5, in monkeys, provides a basis for representing manual grasping and certain mouth movements in a relatively integrated manner.

Some of the neurons in area F5 are connected to both the hand field of the primary motor cortex (M1) and the brainstem (Rizzolatti, Luppino, & Matelli, 1998; Rizzolatti & Luppino, 2001). It has been proposed that connections from this area to the motor representations of M1 might be primarily responsible for the area's role in motor control (Cerri, Shimazu, Maier, & Lemon, 2003; Davare, Lemon, & Olivier, 2008). However, F5 can also contribute to the control of actions, in parallel with M1, via direct projections to the brainstem (Borra, Belmalih, Gerbella, Rozzi, & Luppino, 2010). Therefore, it is possible that premotor projections to the brainstem might also have a role, for example, in the coordinated hand-to-mouth self-feeding behavior, hand movement-related "tongue thrusts" and other phenomena that show tight connection between the movements of hand and mouth. Moreover, given that many manipulative manual skills appear to be rhythmically organized similar to many other motor behaviors such as mastication and vocalization (Kunesch, Binkofski & Freund, 1989; Lund & Kolta, 2006; Jürgens, 2002), the central pattern generators (CPG), identified mainly in the brainstem (Cohen, Rossignol, & Grillner, 1988), might also contribute to these interactive hand-mouth phenomena.

Involvement of premotor region F5, in monkeys, in connecting manual grasping with mouth actions is an intriguing finding because it has been proposed that the homologue of the monkey ventral premotor area F5c is Brodmann's area 44, which is a part of Broca's area (Rizzolatti & Arbib, 1998). Research has shown that in humans, this area is involved, for example, in the production of articulatory gestures (Papoutsis, de Zwart, Jansma, Pickering, Bednar, & Horwitz, 2009), encoding articulatory gestures (Demonet, Chollet, Ramsay, Cardebat, Nespoulous, Wise, et al., 1992; Zatorre, Evans, Meyer, & Gjedde, 1992), observation of silent speech (Buccino, Lui, Canessa, Patteri, Lagravinese, Benuzzi, et al., 2004), observation of mouth movements related to eating (Buccino, Binkofski, Fink, Fadiga, Fogassi, Gallese, et al. 2001), performing manual manipulation of objects (Binkofski, Buccino, Stephan, Rizzolatti, Seitz, & Freund, 1999), observation of grasp actions (Rizzolatti, Fadiga, Matelli, Bettinardi, Paulesu, Perani, et al. 1996), and mental imagery of grasping movements (Decety, Perani, Jeannerod, Bettinardi, Tadary, Woods, et al., 1994). Given that this area is involved in speech in general

and in articulatory gestures in particular, and that the area is also involved in manual gestures in general and in grasping in particular, it has been proposed that this area is also involved in connecting representations of manual grasp actions to mouth gestures (Gentilucci & Dalla Volta, 2008). That is, it is possible that a corresponding premotor area, which in monkeys integrate grasping with mouth movements, in human, not only provides a core system for speech processes but might also provide a neural basis for processes that integrate manual movements with the movements of lips, mouth and tongue observed, for example, in an earlier mentioned automatic tendency for mimicking of hand movements with the mouth, tongue and lips.

1.3. Behavioral evidence for interaction between grasping and mouth actions related to vertical mouth movements and vowel production

Regarding behavioral evidence of interplay between mouth movements and manual grasp actions, Gentilucci and his colleagues have carried out many studies exploring how reaching-to-grasp performance influences the lip kinematics and the features of voice spectra. For instance, they have shown that when the participants are required to hold their mouths open while they perform an object-directed grasp action with the thumb and the index finger, the grasp kinematics influence the mouth opening (Gentilucci, Benuzzi, Gangitano, & Grimaldi, 2001). Grasping a large object, requiring relatively large opening between the thumb and the index finger, simultaneously increases the lip aperture in comparison to conditions in which the target object is small, requiring narrower finger opening. Corresponding phenomenon was observed in infants aged between 11 and 13 months. When infants were allowed to grasp and manipulate large objects in comparison to small objects, the first formant (F1) of their concurrent vocalization increased (Bernardis, Bello, Pettenati, Stefanini, & Gentilucci, 2008). Given that higher F1 is related to larger internal mouth aperture (Fant, 1960), it appears that the effect in which mouth opens symmetrically with opening of fingers during grasping and manipulation is a robust phenomenon. Moreover, when participants are required to vocalize closed ([i]) or open ([a]) vowel while grasping an object using the precision grasp, the open and closed vowels

induced larger and smaller finger opening, respectively (Gentilucci & Campione, 2011). In other words, the automatic tendency to open and close the mouth and fingers in tight synchrony seems to operate from mouth to hand and from hand to mouth.

The evidence showing interplay in opening and closing movements between mouth and fingers can be considered to provide further evidence for so called “motor overflow” phenomenon (Yensen, 1965) showing, in general, involuntary contractions of muscles that are not directly related to the performed action. Most commonly this motor overflow is observed during unimanual tasks in involuntary movements of contralateral homologous muscles (Todor & Lazarus, 1986). For example, it has been shown that it is easier to move both arms upwards than move one arm upwards and other downwards (e.g., Swinnen, Jardin, Meulenbroek, Dounskaia, & Hofkens-Van Den Brandt, 1997; Serrien, Bogaerts, Suy, & Swinnen, 1999). As another example of motor overflow between hands, finger opening (i.e., distal extension) movement performed with one hand is facilitated by simultaneous proximal extension movement of the other hand in comparison to the condition in which this other hand is performing proximal flexion movements (Vainio & Tiainen, 2017). Moreover, Gentilucci et al. (2001) showed that when participants are asked to grasp different sized objects using the precision grip and simultaneously open the thumb and the index finger of the other hand, the finger opening of the non-grasping hand was increased when the grasped object was large, and hence required relatively large finger opening. The above introduced evidence showing tight symmetry in movement kinematics between hand grasping and mouth opening (e.g., Gentilucci et al., 2001; Gentilucci & Campione, 2011) shows that similar movement symmetry effects that have been observed between two hands can be also observed between hand and mouth movements. As such, it could be assumed to reflect the same mechanisms that underlie the phenomenon recognized, for example, by Darwin (1872) who reported simultaneous opening and closing movements of jaws symmetrically with the blades of the scissors when an individual is cutting a paper.

Mirroring of opening and closing movements between mouth and hand, however, is not something that would be necessarily always expected to occur when observing interaction between opening/closing of the mouth and grasp-related finger movements. Indeed, direct electrical stimulation of the precentral gyrus causes the opening of the mouth and the closing of the hand, which is moved toward the mouth through contraction of flexor muscles of the hand and wrist (Desmurget, Richard, Harquel, Baraduc, Szathmari, Mottolese, et al. 2014). This effect was assumed to underlie ethologically relevant hand-mouth synergies for the purpose of bringing a hand-held object or food into the mouth - behavior that has been shown to occur even in fetuses and newborns (Butterworth & Hopkins, 1988; Myowa-Yamakoshi & Takeshita, 2006; Rochat, 1993). Accordingly, in so-called Babkin effect (Babkin, 1958), which is similarly assumed to be closely related to combined hand-mouth coordination operating for food intake, pressing palms of newborns do not only result in the palmar grasp reflex but it can also trigger opening of the mouth (Futagi, Yanagihara, Mogami, Ikeda, & Suzuki, 2013). In addition, behavioral studies in adults have also shown interaction between power grasp closure and vocal tract opening. When participants are observing someone else reaching-to-grasp an apple using the power grip, and they are asked to pronounce the syllable /ba/ at the time of the grip closure, vocal tract opening is increased, indicated by relatively high F1 values in the voice spectra, in comparison to the condition in which the observed hand is grasping a cherry using the precision grip (Gentilucci, Stefanini, Roy & Santunione, 2004).

Regarding the relationship between opening of the mouth and the power grip closure, research has also shown that the power grip closure can be facilitated by simultaneous production of the open vowel [a] in comparison to the closed vowel [i]. It is noteworthy that similar power grip closure execution has been shown to be processed within the premotor network that is also involved in execution of manipulative hand actions (e.g., watering a plant or cracking a nut) and observing the corresponding hand actions (Arnstein, Cui, Keysers, Maurits, & Gazzola, 2011). In the study reported by Vainio, Schulman, Tiippana, and Vainio (2013), participants were presented with the vowel [i] or [a], in green or blue color, on the computer

monitor. The participants were holding precision and power grip devices in their hand. They were required to perform either the power grip closure or the precision grip closure according to the color of the letter, and simultaneously articulate the vowel (see Figure 1 for description of this dual-action protocol). It was found that the power grip closure was performed relatively rapidly together with the open vowel while the precision grip closure was performed relatively rapidly together with the closed vowel. In addition, this vowel-grip effect can be also observed in vocal reaction times: the production of the closed vowel [i] is faster when the simultaneous manual response is produced with the precision grip device while the production of the open vowel [a] is faster when the manual response is produced with the power grip device (Tiainen, Tiippana, Vainio, Komeilipoor, & Vainio, 2017a). Correspondingly, the production of the open vowel [a] is facilitated by observing an image of the hand that is shaped into the power grip closure whereas the production of the closed vowel [i] is facilitated by observing an image of the hand that is shaped into the precision grip closure (Vainio, Rantala, Tiainen, Tiippana, Komeilipoor, & Vainio, 2017). These findings are in line with the observations showing increased vocal tract opening when the participant is presented with the power grasp closure stimulus in comparison to the precision grasp closure stimulus at the time of vocalization (Gentilucci, Stefanini, Roy, & Santunione, 2004). Hence, it appears that the relationship between the power grip closure and opening of the mouth is an innate tendency (Babkin, 1958) that primarily might serve for food intake functions (Desmurget et al. 2014) but that also seem to operate in the context of production of open vowels (Gentilucci et al., 2004; Vainio et al., 2013; Tiainen et al., 2017a).

In contrast, the precision grip closure appears to be linked to relatively closed vocalizations (Gentilucci et al., 2004; Vainio et al., 2013; Tiainen et al., 2017a). Vainio et al. (2013) proposed that the precision grip is associated with the closed vowel [i] because the articulatory gesture for producing this vowel is kind of articulatory counterpart for the precision grip: it is analogously formed by a “small” mouth shape in which the tongue blade (i.e., the area just behind the tongue tip) is pushed into a high-anterior position, which in turn significantly

narrows the vocal tract. Similarly, Ramachandran and Hubbard (2001) proposed that the small pincer gesture might be mimicked by oral gestures for articulating the vowel [i], which in turn is based on tight cross-wiring between motor processes that program hand and mouth actions for gesturing.

1.4. Behavioral evidence for interaction between grasping and consonant-related mouth movements

It has been shown that not only are the mechanisms responsible for articulating open and closed vowels connected to precision and power grip closures, but the mechanisms responsible for articulating certain consonants are also connected to these grip actions. Vainio et al. (2013) used the same dual-action protocol that showed connections between open and closed vowels and the different grip types in order to explore this issue. However, instead of pronouncing the open and closed vowels, the participants were required to pronounce the apical stop consonant [t] or the dorsal stop consonant [k] in the context of consonant-vowel (CV) syllables (i.e., [te] and [ke]). It was found that the production of the consonant [t] was associated with facilitation of the precision grip closure while the production of the consonant [k] was associated with facilitation of the power grip closure. Similarly to the vowel-grip effect, this consonant-grip effect can be also observed in the vocal responses (Tiainen et al., 2017a).

Vainio et al. (2013) originally proposed that this consonant-grip effect is something that might be predicted by mouth-hand mimicry theories according to which there is an innate tendency to mimic manipulative hand actions by tongue, lips and mouth (Wallace, 1881; Paget, 1963; Hewes, 1973). As such, the effect between the precision grip and the consonant [t] might be observed because the final shape of the tongue, when articulating the consonant [t], and the final shape of the hand, when producing the precision grip, are both achieved by moving the tip of the distal effector (thumb/tongue) into contact with other part of the distal effector (tip of the index finger/alveolar ridge and back of upper-central incisors). That is, the effect is observed

because the overlapping motor representations are involved in planning analogous movement goals for two separate but tightly interacting effectors. In non-speech context, a corresponding effect, which is based on these shared motor representations, might be observed, for example, when an individual protrudes the tip of the tongue whilst performing a manual task requiring precise object manipulation with the precision grip (Mason & Proffitt, 1974).

The consonant-grip effect associating the power grip with the consonant [k] was also proposed to be based on similar overlaps in planning analogous final shapes of mouth and hand. In this power-grip effect, the tongue body is progressively moved back on the velum analogously to the power grip closure which is produced by flexing intermediate and proximal components of fingers—rather than finger tips—against the palm base; that is, both actions use the base part of the effectors in order to produce the final shape. Earlier, Ramachandran and Hubbard (2001) have speculated similar analogy between the shape of articulation and a power grip-type of hand movement. They suggested that “the flexion of the fingers and palmar crease in ‘come hither’ is mimicked by the manner in which the tongue goes back progressively on the palate” (p. 21). Explanation of the consonant-grip effect is, however, further discussed and elaborated below in the sections related to development and evolution of speech.

In line with findings showing that chimpanzees increasingly perform protrusion of the lips during precision grip performance (Waters & Fouts, 2002), and that activation of the muscle that executes corresponding lip protrusion in humans is increased during precision grasping (Higginbotham et al., 2008), Vainio et al. (2013) showed that the vowel [u], which is executed by protruding the lips, is linked to facilitated precision grip responses. In addition, the study showed that articulation of the consonant [m] is associated with facilitated power grip responses. This finding is in line with the evidence showing that people have a tendency to associate this consonant with large objects (Newman, 1933; Thompson & Estes, 2011). Moreover, Vainio, Tiippana, Tiainen, Rantala and Vainio (2018) showed that articulation of the consonant [r] is linked to facilitated precision grip responses. This finding was also expected because the

consonant [r] is an alveolar trill that is produced by fronting the tip of the tongue, similar to the consonant [t].

The articulation-grip effects introduced above that link certain vowels ([i] and [ɑ]) and consonants ([t] and [k]) to the precision and power grips were originally observed in Finnish speakers (Vainio et al., 2013). However, the same effects have been, so far, also replicated in Czech speakers (Tiainen, Lukavský, Tiippana, Vainio, Šimko, Felisberti, & Vainio, 2017b). The fact that the same effects can be observed in two separate languages that belong to differentiated language families (Finnish is a Finno-Ugric language, and Czech is a Slavic language) suggests that people have an implicit and language-independent tendency to link the precision grip with the closed vowel [i] and the apical stop consonant [t] and the power grip with the open vowel [ɑ] and the dorsal stop consonant [k]. In addition, the effect observed in the facilitated precision and power grip responses can be observed when participants are silently reading meaningless CV syllables consisting of these letters just before the grip response or when they are aurally presented with these syllables at the time of the grip response (Vainio, Tiainen, Tiippana, & Vainio, 2014). Moreover, the effect can be observed in facilitated vocal responses even when the grip is only prepared but not performed, suggesting that merely planning a grip response primes the corresponding articulatory response (Tiainen et al., 2017a). In addition, in line with these behavioral observations, TMS (transcranial magnetic stimulation) elicited motor-evoked potentials of the hand muscles associated with the precision and power grip exhibit significantly increased amplitudes during articulation of the syllables [ti] and [kɑ] (Komeilipoor, Tiainen, Tiippana, Vainio, & Vainio, 2016).

Finally, the preparation of these grip responses appears to modulate the categorization (Tiainen, Tiippana, Vainio, Peromaa, Komeilipoor, & Vainio, 2016) and perception (Tiainen, Tiippana, Paavilainen, Vainio, & Vainio, 2017c) of heard syllables that are congruent with the grip type, which suggests that the motor preparation of these grip types is connected to the processing of speech sounds. As a consequence, this finding has some parallels with the motor theory of speech perception (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967),

which suggests that the perception of speech sounds is partially formed by representing the speech sounds in the listener's own articulatory representations. Although this theory is criticized by many researchers (Lotto, Hickok, & Holt, 2009), it is nevertheless noteworthy that a great number of evidence supports this view. It has been shown, for example, that listening to speech (e.g., syllables) recruits specific motor representations of the tongue (e.g., D'ausilio, Maffongelli, Bartoli, Campanella, Ferrari, Berry & Fadiga, 2014), and that these motor structures provide a functional contribution to the perception of noise masked speech sounds (e.g., D'Ausilio, Pulvermüller, Salmas, Bufalari, Begliomini, & Fadiga, 2009). As such, the evidence showing that grip preparation systematically modulates the categorization and perception of heard syllables (Tiainen et al., 2016; Tiainen et al., 2017) suggests that similarly to the articulatory motor processes, the grip processes can be also involved in the decoding speech sounds. In sum, the evidence introduced above suggest that the articulation-grip effects are very robust, implicitly operating and language-independent phenomena that might be linked to the perception of speech sounds congruent with these grip types, and that are primarily based on the processes shared for planning and execution of grasp actions and articulatory gestures.

1.5. Behavioral evidence for interaction between arm movements and articulatory gestures

Although the studies exploring interaction between articulatory gestures and manual actions have focused on manual grasping (e.g., Gentilucci et al., 2001; Vainio et al., 2013), there is one effect showing interaction between articulatory gestures and reach-related manual actions (Vainio, Tiainen, Tiippana, Komeilipoor, & Vainio, 2015). This effect was found using a similar dual-action paradigm that was also used when exploring the vowel- and consonant-grip effects with an exception that the participants were required to manually respond by pushing or pulling a joystick at the time of articulation. This effect shows that front-back tongue movements required to articulate front and back vowels are associated with processes preparing forward-backward hand movements, respectively (see Figure 1 for the illustration of this method and

effect). For example, the hand movement and the vowel production are both facilitated when the simultaneously performed dual-action requires forward-directed manual movement and articulation of the rounded front-mid vowel [ø]. Similar facilitation effect is observed between backward-directed manual movement and articulation of the rounded back-mid vowel [o] (Vainio et al., 2018). Moreover, similarly to the articulation-grip effect, this reach effect can be found in Finnish and Czech speakers (Vainio et al., 2015; Tiainen, et al., 2017b).

The vowel-reach effect has some similarities with the so-called articulatory In-Out effect originally reported by Topolinski, Maschmann, Pecher, and Winkielman (2014). This effect shows that pseudo-words constructed by a consonant-vowel structure whose articulation requires oral movement from front to back of the mouth (e.g., PEKA) are rated more likeable than the pseudo-words constructed by a consonant-vowel structure whose articulation requires oral movement from back to front of the mouth (e.g., KEPA). This effect was assumed to be an oral version of the approach-avoidance effect in which a hand is typically moved faster backwards when processing affectively positive stimuli (i.e., it is approached by moving a lever toward oneself) and moved faster forwards when processing affectively negative stimuli (i.e., it is avoided by pushing the lever away from oneself) (e.g., Chen & Bargh, 1999). However, in contrast to the In-Out effect, the vowel-reach effect cannot be observed when articulation requires moving the tongue forwards or backward in order to produce the coronal consonant [t] produced by fronting the tongue tip towards high-front position or the dorsal consonant [k] produced by raising the tongue body towards the back of the velum (Tiainen et al., 2017b). In fact, Vainio et al., (2018) suggested that the vowel-reach effect is exclusively linked to fronting and backing the tongue body for vowel production rather than shaping the tongue in order to articulate the consonants such as [t] and [k] that require relative forward and backward tongue movements. Furthermore, the fact that the In-Out effect is not systematically observed in relation to vowel fronting and backing (Topolinski & Boecker, 2016) unlike the vowel-reach effect, suggests that these two effects are based on different underlying mechanisms.

This vowel-reach effect has similarities with the commonly observed movement symmetry effects showing tendencies for moving, for example, finger, wrist and arm of both hands in the same direction rather than different directions (e.g., Swinnen, et al., 1997; Serrien, et al., 1999). These movement symmetry effects are likely to be based on representing locational reach coordinates in terms of multi-sensory spatial codes that are involved in coordinated movements of multiple effectors such as hand, eyes and head (Cohen & Andersen, 2002). Hence, given that the same premotor regions appear to be involved in spatial programming of hand and tongue movements (Kawashima, Itoh, Ono, Satoh, Furumoto, Gotoh, et al., 1996; Watanabe, Sugiura, Miura, Watanabe, Maeda, Matsue, et al., 2004; Petrides & Pandya, 2009), the vowel-reach effect finding suggests that spatial programming of fronting and backing the tongue body for vowel production is processed in interaction with spatial programming of moving the hand forwards and backwards. Hence, the vowel-reach effect provides one more example of how motor programming between non-communicative hand actions and movements of mouth, lips and tongue is integrated.

To summarize thus far, a wide variety of evidence shows close connections between articulatory and non-articulatory mouth actions and manipulative hand movements such as grasping. That is, manual processes are not connected to speech only due to co-speech gestures (McNeill, 1992), but these connections appear to operate even at the very basic and systematic level of different grasp actions and articulatory gestures (Gentilucci & Campione, 2011; Vainio et al., 2013; Vainio et al., 2015). The next sections will focus on reflecting these hand-mouth effects on findings and theories related to development and evolution of speech and language as well as the phenomenon of sound symbolism. These sections will assess to what extent these effects are in line with different theories of speech evolution, and how compatible they are in relation to developmental aspects of speech. Based on these evaluations, the aim is also to explore whether these effects can provide some novel insights on evolutionary and developmental accounts of speech.

---Figure 1 about here---

2. Developmental perspectives

Learning to master speech is the most complex motor action. It involves combined coordination of distinct articulatory segments used for speech including the larynx, lips, teeth, alveolar ridge, hard palate, velum (soft palate), uvula, and various parts of the tongue (Knight, 2012). Within the first year of life, the mastery of speech sounds develops from around seven phoneme types, at the age of 2 months, to around nineteen phoneme types (Irwin, & Chen, 1946). In general, vowels develop earlier than consonants (Wellman, Case, Mengert, & Bradbury, 1931). Babbling (i.e., repetitively produced speech-like syllables) precedes the first real words (Oller, Wieman, Doyle, & Ross, 1976; Kent, & Murray, 1982) and begins typically at the age around 5 to 6 months, having a CV shape in which the consonants are most likely labial or alveolar stops or nasals and the vowels are most likely central or low- to mid-front in place (e.g., [bʌ], [da], [mæ]) (Rvachew, Alhaidary, Mattock, & Polka, 2008). Babbling begins to diminish when the child learns to produce his/her first words by 12 months of age (Fenson, Dale, Reznick, Bates, Thal, Pethick, et al., 1994).

The other everyday action that has to be learned during infancy and early childhood, which might be assumed to require close to the same level of complexity as speech, is manual manipulation. It employs fine combined coordination of five fingers – in some cases even requiring the combined coordination of fingers on two hands. At the developmental level, manual skills progress from reflexive power-type of grips to object-directed grasps that are first achieved using power grips and then precision grips (Butterworth, Verweij, & Hopkins, 1997; Halverson, 1931; Wallace & Whishaw, 2003). There is a few months transition period starting on average at the age of 8 months when the power grip, which does not involve the thumb, is replaced by the precision grip when small objects are grasped (Butterworth et al., 1997). In addition, during early grasping and manipulation, the arm, hand and fingers move mostly as a single unit (Jeannerod, 1988). Gradually, the child works toward gaining independent movement of these segments for grasping and manipulation (Schuster & Ashburn, 1992; Trevarthen, 1984). This kind of limited independence of anatomically distinct segments is a common phenomenon in

immature motor systems (Missiuro, 1963). In fact, it has been proposed that development of motor control over articulatory organs may progress in a similar manner so that gradually increased sophistication of speech during development requires increasingly independent control of distinct articulatory organs (Green, Moore, Higashikawa, & Steeve, 2000). That is, speech and manual skills appear to progress through similar developmental motor stages.

In addition to similarities in the motor complexity and in developmental requirements for increased independency related to motor control of distinct organs, mouth- and hand-related actions appear to develop in close interaction. There are many reported ways, at different levels of development, in which they are developmentally connected to each other. One of the earliest manifestations of the hand-mouth interaction in infants is the Babkin effect (Babkin, 1953), as already shortly mentioned above. It is elicited by pressing thumbs against palms of newborns lying on a flat surface in the supine position. This pressing technique results in opening of the mouth as well as flexion of the fingers and the forearms (Futagi et al., 2013; Pedroso & Rotta, 2004). This reflex can be observed more frequently before feeding than after feeding suggesting that it is closely related to food intake behavior in which the piece of food is first grasped and then moved toward the opening mouth by flexion of the forearm (Futagi et al., 2013). Although this reflex is typically suppressed by the third or the fourth month of life (Futagi et al., 2013), the motor network that controls this reflex appears to still exist in adulthood given that electrical stimulation of the precentral gyrus of adult participants elicits involuntary hand-to-mouth movement that is very similar to the Babkin reflex (Desmurget et al. 2014). In addition, it has been also found that, in infants as young as 9 weeks of age, the finger flexion behavior appears to be also associated with vocalization (Fogel & Hannan, 1985). That is, when infants' behavior is observed during spontaneous face-to-face interaction with their mother, it appears that they are likely to flex their fingers while they open their mouth for vocalization. Correspondingly, in adults, flexing fingers around an object in order to produce the power grip facilitates simultaneous vocalization of open vowels in comparison to closed vowels (Vainio et al., 2013). Hence, there appears to be hardwired connections between power grip-type of flexion of fingers

and mouth opening that might primarily function for primitive food intake behavior, but that might also operate in relation to vowel production that require shaping of the vocal tract in a particular manner.

It is possible that the connections that integrate motor programs that control specific types of hand and mouth movements might even play some role in the development of speech. Indeed, it has been shown that the development of manual movements is functionally related to speech development. Firstly, it has been proposed that manual babbling can facilitate speech development (Iverson & Thelen, 1999; Goldin-Meadow & Alibali, 2013). These rhythmical movements of the arms and hand (e.g., swaying, waving and banging) occurs about the same age, averaging about 27 weeks, as the emergence of canonical babbling. According to Iverson (2010) this connection between manual and canonical babbling is an important developmental step for full-blown emergence of co-speech gesturing, which typically begins between 8 and 12 months initially using deictic gestures (Bates, 1976; Iverson & Goldin-Meadow, 2005). The fact that the co-speech gesturing at 14 months is relatively accurate predictor of a child's vocabulary at 42 months (Rowe & Goldin-Meadow, 2009), and that the development of co-speech gesturing can, in turn, be boosted by the connection between manual and canonical babbling (Iverson & Goldin-Meadow, 2005) shows a functional link between the manual and articulatory processes in relation to speech development.

Secondly, development of manipulation and grasping behavior also seem to be associated with speech development. It has been shown that, in general, poor manual motor skills in childhood related to precise object manipulation and grasping (e.g., peg moving, posting coins, fastening buttons and threading beads) are often associated with specific language impairments such as relatively poor articulatory movement skills (Owen & McKinlay, 1997; Hill, 2001; Bishop, 2002; Finlay & McPhillips, 2013). Hence, one might speculate that development of grasping and manipulation is an important precursor and facilitator of developing oral skills for speech. This view is also supported by the finding that language development appears to correlate with lateralization of grasping and manipulation behavior: infants who show right-hand

preference for grasping and manipulation show relatively advanced development of language skills measured a few months later (Nelson, Campbell, & Michel, 2014).

The cortical basis for the above mentioned functional connection between the development of manual and speech skills is not entirely understood, however. As mentioned in the Section 1.2., in monkeys, the core of the mouth-hand connections might be grounded on neural overlaps between the premotor networks in the area F5, which provides a basis for representing manual grasping and certain mouth movements in a relatively integrated manner. Similarly, in humans, the pars opercularis of the inferior frontal gyrus, the homologue of this F5 area (Rizzolatti & Arbib, 1998), might primarily program the mouth and hand movements for self-eating behavior but might also be adapted for refining articulation in connection to manipulation behavior (Gentilucci & Corballis, 2008). As such, one might speculate that hand-mouth connections, particularly in this part of Broca's area, might have a central role in the developmental relationship between the manual and speech skills. These connections might serve the facilitatory influence of manual development on speech development. As an example, an infant's improved manual skills for precise object manipulation and precision grasping might improve the articulatory control of the consonant [t] via the given grasp- and articulation-specific frontal connections.

The networks mediated by the central pattern generators (CPG) in the brainstem might also play an important role in connecting the development of manual processes to the development of speech. These networks have been proposed to support the rhythmic features of babbling, as well as the oromotor rhythms in speech (Barlow, Radder, Radder, & Radder, 2010). In addition, manipulative hand movements show similar rhythmical patterns to oculomotor movements related to speech, and might therefore be similarly controlled by CPG-related mechanisms, at least to some extent (Kunesch et al., 1989). It would be tempting to propose that the same CPG-related mechanisms that are initially involved in connecting manual and canonical babbling in speech development also operate for connecting manipulative hand movements to the speech development.

2.1. Proposal for mutual development of grasping and articulation

One of the primary proposals of the current article is to suggest that, similarly to the facilitatory influence of chewing and sucking on speech motor development (Moore & Ruark, 1996), mastery of control over separate speech organs is to some extent also facilitated by the development of manual manipulation and grasping skills through connections between mouth and hand motor representations. This view assumes that the developmental origin of these connections rely partially on grasp-related motor representations that are shared by processes controlling certain mouth and hand movements. Although these innate connections can be initially observed, for example, in Babkin reflex (1953) and in the finger flexion during mouth opening for vocalization (Fogel & Hannan, 1985), they are still present –even though mostly suppressed– in adulthood and can be observed, for example, in the vowel-grip effect linking the power grip with the vowel [ɑ] (Vainio et al., 2013). These connections might construct a basis for the facilitatory influence of development of the precision and power grips on the development of certain articulatory gestures to be discussed next.

I propose that the connections between grasp-related motor programs and certain articulatory gestures, observed in several behavioral studies (e.g., Gentilucci & Campione, 2011; Vainio et al., 2013; Vainio et al., 2015), provide some level of functional value for speech development. Vainio et al. (2018) have previously provided a description of how motor representations controlling individual tongue muscles involved in producing certain consonants might be connected to motor representations controlling the muscles involved in precision and power grip movements. This suggestion can be also used as a basis for an example of how manual processes can facilitate development of articulation at very precise levels. First, relevant to this proposal and roughly phrasing, the extrinsic tongue muscles are responsible for fronting/backing and raising/lowering of the tongue body for vowel production whereas the intrinsic tongue muscles are mostly responsible for shaping the tongue for consonant production (Perkell, 1969; Sauerland & Mitchell, 1975; Jürgens, 2002). Correspondingly, Öhman (1966) has suggested that vowels are mostly performed by moving the tongue body while consonants

are mostly performed by moving the tongue articulators of apex and dorsum. This view assumes that the apex-related consonants such as [t] are achieved by tongue tip movements while the dorsum-related consonants such as [k] are achieved by arching of the tongue body. As already stated, production of the apical ([t]) and dorsal ([k]) consonants are associated with production of the precision and power grips, respectively (Vainio et al., 2013). Hence, Vainio et al. (2018) have recently suggested that these consonant-grip effects might be based on relatively systematic connections between motor representations of hand and tongue. As such, this view assumes that, for example, development of the apex-related articulatory gestures might be facilitated by development of the manual grasping and manipulation skills that utilize the tips of the thumb and the index finger. In contrast, development of the dorsum-related articulatory gestures might be facilitated by development of the manual grasping and manipulation skills that require flexion movements of fingers into contact with the palm. However, further investigation is warranted for supporting this view.

2.2. Speech development and mouthing

The hand-to mouth behavior that can also boost the development of speech is called mouthing. Right after infants learn to grasp objects, they start to perform mouthing behavior in which a grasped object is brought into contact with the mouth, lips and/or tongue (Rochat, 1989; Belsky & Most, 1981). By the age of 5 months, there is a significant increase in frequency and duration of mouthing behavior (Rochat, 1989). It is assumed that infants primarily use mouthing as a means to explore object characteristics and affordances such as shape and size relevant for planning actions (Rochat, 1989; Ruff, Saltarelli, Capozzoli, & Dubiner, 1992). However, it has been proposed that mouthing can also influence development of speech (Fagan & Iverson, 2007). Indeed, it has been shown, for example, that when children whose age varies between 6 and 9 months are observed for 20 minutes, they produce more lip and tongue-related consonants during mouthing than during non-mouthing (Fagan & Iverson, 2007). That is, because mouthing of objects results in closure of the vocal tract, which is a central feature of

supraglottal consonants, and because inserting an object into the mouth during vocalization, for example, when an infant is babbling, pushes the tongue and the lips into various positions that are in turn associated with production of various consonants (Iverson, 2010). As such, mouthing of objects provides infants with information not only about object characteristics, but also about their own vocalizations. In line with this view, the time when mouthing peaks (i.e., between 6 and 9 months) is also an important time in consonant development (Holmgren, Lindblom, Aurelius, Jalling, & Zetterstrom, 1986; Ruff et al., 1992; Belsky & Most, 1981). A corresponding technique is, in fact, intentionally used even by speech therapists: for example, when a child is taught to pronounce the consonants [k], the tip of his/her tongue is pushed downwards using a tongue depressor while s/he is asked to pronounce the consonant, which in turn forces the tongue shape into the right position for pronouncing it.

In addition to vision and mouthing, the other important way to explore object characteristics and affordances such as size and shape is by grasping and fingering it (Rochat, 1989). Additionally, grasping is always required for mouthing an object. Hence, it is possible that if learning some consonants is boosted by mouthing behavior, this learning might also include building implicit association between the consonant whose articulation is boosted by mouthing, the affordance (e.g., size) of the object that is mouthed, and the grasp type that is used for mouthing of that object. In other words, it is possible that, for example, learning to pronounce the consonant [k] is facilitated by bringing a relatively large object into mouth using the power grip, which in turn implicitly associates this grip type with the consonant [k]. If this is the case, it is possible that the consonant-grip effect that links power grip performance to facilitated production of the consonant [k] (Vainio et al., 2013) reflects this kind of association that is initially acquired by mouthing relatively large objects. Moreover, given that production of a wide variety of supraglottal consonants, such as [k] and [t], has been observed during mouthing (Fagan & Iverson, 2007), it is in principle possible that the consonant-grip effect related to the consonant [t] (Vainio et al., 2013) is also primarily based on mouthing behavior. If so, mouthing of a small or narrow object, which is held using the precision grip, may shape the

tongue for producing the consonant [t]. However, the fact that the consonant-grip effect has been observed between the coronal consonant [r] and the precision grip (Vainio et al., 2018), and that the consonant [r] is a later developing consonant that is not typically observed during mouthing (Fagan & Iverson, 2007), suggests that the consonant-grip effects are not likely to be solely based on mouthing behavior. Rather, I prefer the view that the grip effects are more likely to be based on an innate overlap between motor representations that program specific mouth and grasp movements.

2.3. Additional functions of hand-mouth interactions on development of language

In addition to potential influence on the development of certain articulatory gestures, previous research has suggested that hand-mouth interactions might also contribute to the development of more high-level language skills. Perhaps the most famous proposal linking hand movements to language processes, at the developmental level, assumes that hand-mouth interactions provide a basis for the development of grammar (Gleenfield, 1991). According to this view, manipulative and vocal motor functions are poorly separated in the infant brain. At this early stage, the neighboring orofacial and hand motor areas are connected via a large number of short-range connections resulting in conjoint movements of hand and mouth that have been already introduced above. Similarly, early in development, Broca's area mostly programs manual manipulative actions and language production in an undifferentiated manner. That is, it organizes manipulative manual actions (e.g., reaching-to-grasp a spoon in order to pick up a piece of food and move it into the mouth) as well as combining sounds and words for spoken language by using overlapping networks within Broca's area. Later, maturation of the prefrontal cortex, which processes linguistic grammar and action sequences in separate areas (Ojeman, 1983; Roland, 1985), and maturation of its connections to Broca's area, triggers a chain of developmental neural events that gradually separates motor programming between manual and orofacial movement as well as increases differentiation of combinatorial manual object programs from linguistic grammar programs.

Another example of how hand-mouth interactions might contribute to development of high-level language skills is related to the WAT (words as social tools) proposal (Borhi, Barca, Binkofski & Tummolini, 2018). This view is largely based on the account of embodied cognition according to which concepts are grounded in the sensorimotor systems (e.g., Barsalou, 2008; Ghio, Vaghi & Tettamanti, 2013). Indeed, research has shown, for example, that concrete hand-related concepts are partially represented in the hand-motor area while concrete leg-related concepts are represented in the leg-motor area (Hauk, Johnsrude, & Pulvermüller, 2004). However, studies reported by Borghi and her colleagues (e.g., Borghi, Flumini, Cimatti, Marocco & Scorolli, 2011; see review in Borhi, et al., 2018) suggest that motor processes are also involved in representing abstract concepts (e.g., “truth” and “logic”). According to this view, the mouth motor system in particular appears to contribute to representing abstract concepts. Given that infants generally acquire the concrete words before abstract words (Gleitman, Cassidy, Nappa, Papafragou & Trueswell, 2005), it is possible that acquiring abstract concepts is partially grounded on multimodal networks that represent already acquired concrete concepts (Borghi, Barca, Binkofski, Castelfranchi, Pezzulo & Tummolini, 2018). Particularly mouth motor processes might contribute to the abstraction of concepts because these processes are tightly engaged in linguistic operations that are presumable more necessary for acquiring and representing abstract concepts than concrete concepts (Borhi et al., 2018). This view is also supported by the recent brain-imaging findings showing that the face motor region is increasingly involved in representing abstract words (Dreyer & Pulvermuller, 2018). As an example of how hand-mouth connections might support these kinds of abstraction processes, it has been shown that hand motor processes are involved in representing the meaning of, for example, the sentence “Close the drawer” (Glenberg & Kaschak, 2002). However, acquiring abstraction of the word “close” in order to use it in a more abstract manner (e.g., “Close the meeting”) might require that mouth motor processes are also harnessed for representing the abstracted meaning of this word. Perhaps even mimicking the closing action with the lips, mouth and tongue might provide the critical input to this abstraction process.

3. Evolutionary perspectives

As discussed above, evidence suggest that there exist innate anatomical and functional connections between mouth and hand movement, and that these connections appear to facilitate speech development. In fact, it could be presumed that their facilitatory role for speech development is one of the main reasons why these connections have remained until today. Some scientists have argued, however, that these connections might even have a fundamental role in the origin of speech. These kinds of gestural theories of language evolution emphasize the role of manual actions in speech evolution (see review in the paper of Wacewicz, Żywiczyński & Orzechowski, 2016). In addition to these gestural theories, there are a wide variety of evolutionary accounts that emphasize an importance of aspects not related manual processes on language evolution. As an example, some accounts highlight the role of music skills and rhythm processing in language evolution (e.g., Kotz, Ravignani & Fitch, 2018), while other accounts highlight the evolution of specific frontotemporal connections (e.g., the arcuate fascicle) in evolution of language (e.g., Pulvermüller & Fadiga, 2010; Friederici, 2018). Hauser and Fitch (2003) underlined that speech has evolved from primate calls, and gradually the lowered larynx and improved abilities to use vocal organs for spectral shaping has allowed humans to produce a wider range of formant patterns than other animals. In contrast, in their version of the vocal theory of speech evolution, Goldstein, Byrd and Saltzman (2006) emphasize that speech is composed by actions of vocal tract organs (i.e., articulatory gestures) that are planned in terms of reaching their target (i.e., specific shape of the vocal tract formed by contracting, for example, tongue and lips muscles). As such, articulatory gestures can be considered as other skilled actions such as grasping, which is similarly a goal-directed action that is reached by precisely controlling individual units of action such as the arm, hand and fingers. Indeed, this view assumes that evolution of speech is largely based on gradually improved abilities to control individual speech organs such as the tongue tip for producing articulatory gestures, which has increased repertoire of articulatory phonemes and enabled coupling of vowels and consonants in a synchronized manner. In contrast, MacNeilage (1998)

has emphasized the adaptation of ingestive cyclicities (i.e., rhythmical open-close alternations of the vocal tract in the actions such as chewing, sucking and licking) for the evolution of syllabic structure of speech, which similarly consists of opening and closing phases for producing vowels and consonants, respectively. According to this view, these initially ingestive vocal tracts actions have been gradually adapted for communicative purposes as lip-smacks, tongue-smacks and teeth chatters – orofacial actions that are used for communicative purpose in many non-human primates (Redican, 1975). The view assumes that appearance of these orofacial communicative gestures can be considered as a precursor to development of consonants. In turn, accompanying them increasingly by phonation has triggered usage of consonant-vowel syllables for communication. In line with this MacNeilage's evolutionary account, it has been proposed that the same brainstem circuits of a central pattern generator are involved in producing rhythmical mouth movements for mastication and vocalization (Lund & Kolta, 2006; Jürgens, 2002).

3.1. Gestural theories of speech evolution

Although the 'gestural theories' of speech evolution do not, in general, deny the importance of the above-mentioned evolutionary aspects related to vocal organs and their control, they mostly emphasize that manual gesturing is a precursor to speech. One of the earliest attempts to demonstrate the gestural origin of speech was disclosed by Alfred Wallace (1881) who noticed that "many savages point with the lips as we do with the finger, signifying there, by protruding the lips in the direction to be indicated" (p. 245), and hence in many languages the word "go" includes contraction of the lips as the lips would mimic manual pointing gesture to show direction. Later, Paget (1930) has similarly suggested that the actions of articulators often echo hand movements, and that speech has largely evolved in the process of articulators involuntarily mimicking an individual's own hand gesticulations. This view has been revived by Gordon Hawes (1973; 1977) who has agreed with the view that speech might have originated first from a tendency to mimic an individual's own communicative and manipulative manual actions by

orofacial gestures. Although these views presume that hand movements were initially mimicked by visible orofacial gesturing, Paget (1944) in particular emphasized the role of the acoustic consequences of these orofacial gestures in speech evolution. Indeed, as stated by Wacewicz, et al. (2016), an important phase in speech evolution might have occurred when individuals began to combine these orofacial gestures with phonatory sounds enabling the use of invisible gestures in the vocal tract for communicative purpose.

Given that babies learn their first words largely by imitating utterances of other people, the gestural as well as the vocal theories of speech evolution mostly accept that capability for sophisticated imitation is a necessary precondition for speech evolution (Christiansen & Kirby, 2003). Indeed, imitative capabilities at 1.5 years have been found to predict language production at 2 years (Laakso, Poikkeus, Katajamaki & Lyytinen, 1999). Modern gestural theories of speech evolution have raised imitation processes into the core of the theory mostly because of the mirror neuron findings introduced earlier in this review. According to Rizzolatti and Arbib (1998), the mirror neuron system, which enables recognizing and imitating actions made by others, has provided the neural prerequisite for the development of inter-individual communication and finally of speech. This mirror system hypothesis of speech evolution was later elaborated by Arbib (2002; 2005) who emphasized the role of evolution of a mirroring system, which initially allowed imitation of grasping and manual manipulation movements, in the development of speech. This view also underlined the importance of finding that mirror neurons do not only allow motor processing of observed hand movements, but observed communicative orofacial gestures are also processed in the monkeys' mirror neuron area of F5 (Ferrari et al., 2003). Briefly, the theory assumes that transition from manually based communication to orally based communication worked its way through the development of these kinds of mirror structures that allowed increasingly flexible and sophisticated imitation and motor control processes, first, for developing manually produced proto-signs and, later, for developing proto-speech. Corballis (2003), who has also provided a comprehensive account for the role of hand-mouth interactions in the putative gesture-speech transition, has similarly emphasized the role of mirror neuron

system in this transition of communication from hand to mouth. As stated, for example, by Gentilucci and Corballis (2006), this transition might be based on hand-mouth connections that allow representing observed grasp actions regardless of whether the action is performed with the hand or mouth (Rizzolatti et al., 1988; Gallese et al., 1996). According to this view, these connections might have originally operated for self-feeding behavior, but were later co-opted for adapting communicative and manipulative hand gestures in orofacial gesturing.

3.2. Proposal for the role of hand-mouth connections to speech evolution

Although different theories of speech evolution emphasize different aspects of this evolution, it is likely that multiple different factors have together influenced speech evolution. Some of these factors might be exclusively related, for example, to the evolution of speech organs and the neural processes that control them – apart from any influences of manual processes – while some of them might be linked to interactions between mouth and manual processes. Hence, in the present review, I do not take a strong stance for or against the view that appearance of speech was preceded by manual gesturing. However, the evidence discussed above showing anatomical (e.g., Rizzolatti et al., 1988; Gallese et al., 1996), behavioral (e.g., Gentilucci & Campione, 2011; Vainio et al., 2013) and developmental (e.g., Hill, 2001; Finlay & McPhillips, 2013) connections between manipulative hand actions and mouth actions might be taken – at least – to support the view that these connections have provided one ingredient for speech evolution.

If one considers the evidence for interaction between articulatory and non-articulatory mouth actions and non-communicative hand movements (e.g., manual grasps and manipulation actions), it would be difficult to deny some influence of processes controlling manipulative hand movement on the evolution of speech. If these connections really are innate as the evidence introduced in the section of “developmental perspectives” suggests (e.g., Babkin, 1953; Fogel & Hannan, 1985; Iverson & Fagan, 2004), and if they really influence speech development from

very early infancy as the evidence also appears to suggest (e.g., Nelson et al., 2014; Hill, 2001; Finlay & McPhillips, 2013), it would be plausible to assume that they have also played some role in the development of speech in the scale of evolution. Hence, in the light of the above mentioned evidence, I prefer to propose that evolutionarily there is old overlap between processes that program orofacial actions and different kinds of manual grasp and manipulation actions. Agreeing with the view of Gentilucci and Corballis (2006), this overlap might have existed even before emergence of the first orally based language, and they might have initially served for self-feeding function. However, these connections have remained until today, and they still might continue to serve function of speech development by facilitating shaping of articulatory gestures during early stages of speech development. While agreeing with the view that the speech development in children to some extent mirrors how speech was originally started by our ancestors and evolved through history (Imai & Kita, 2014), however, these connections might have also had a facilitatory role in the development of some articulatory gestures in the scale of evolution. These interactions might have even facilitated shaping of the first consonants and vowels of the early orally based languages. Indeed, as proposed by MacNeilage (1998), the early orally produced languages presumably had a very limited number of consonants, and the consonants [t] and [k] – included in most languages and that are frequently used consonants in babbling – were likely to be among these first consonants included in the initial orally-produced languages. The present view assumes that one reason why these consonants are among most widely and frequently used ones, is because their shaping is facilitated by grasping and manipulative motor processes – recall that they are associated with the precision and power grip, respectively (Vainio et al., 2013) – that in turn advances their evolvment in the developmental scale, but which has also presumable advanced their evolvment in the evolutionary scale. That is, the present view assumes that the grip-mouth connections might play a facilitatory role in the development of specific articulatory gestures (such as certain apical and dorsal consonants) and they might have accelerated their appearance and/or boosted their occurrence during speech evolution.

4. Sound symbolism and mouth-hand connections

In addition to evolution and development of speech, one manifestation of the interactions between articulatory gestures and different grasp-related and manipulative hand shapes might be linked to a phenomenon called sound symbolism (Vainio et al., 2013). Link between sound and meaning seems relatively arbitrary for most words in the lexicons of languages (de Saussure, 1983). Increasing evidence, however, suggests that a significant proportion of words also include various types of non-arbitrary sound-meaning relationships. For example, when Berlin (1994) asked their English-speaking participants to classify an aurally presented list of bird and fish names taken from a language spoken in north central Peru, the participants were able to classify the words into the correct fish-bird categories significantly above chance. More recently, Blasi, Wichmann, Hammarström, Stadler, and Christiansen (2016) analyzed word lists covering nearly two-thirds of the world's languages, and found that a considerable proportion of basic vocabulary items carry strong sound-meaning associations.

In some cases, the sound-meaning association evidently reflects a relationship between the speech sound and the concept. In the phenomenon called onomatopoeia, for instance, a word phonetically resembles the sound of the referent concept (e.g., boom, meow). In some other examples of sound-meaning associations, however, the relationship between the sound and the meaning is not that explicit. For instance, although sound-meaning associations between the phoneme [i] and the concept of 'small' as well as the phoneme [a] and the concept of 'large', are one of the most frequently reported phenomena of sound symbolism (Sapir, 1929), it is not entirely clear what is so 'small' or 'large' in the sounds [i] and [a], respectively. As such, the onomatopoetic phenomena can be considered to present instances of direct iconicity, in which form maps directly onto meaning via resemblance, whereas the size-sound mapping can be considered to present an instance of indirect iconicity, in which associations linked to sound are mapped onto meaning (Masuda, 2007; Sidhu & Pexman, 2018). Sidhu and Pexman (2018) have recently reported that there are several possible mechanisms that map speech sounds indirectly onto meaning. One possible way that speech sounds are indirectly mapped onto

meaning might be related to acoustic properties of speech sounds. It is known that front vowels such as [i] have relatively high frequency F2 and consequently they also have a higher pitch, while back vowels such as [ɑ] have lower frequency F2 and consequently they also have a lower pitch (Fant, 1960). Hence, it has been proposed that the size-sound symbolism, which associates front vowels with small objects and back vowels with large objects, might be due to the fact that smaller objects tend to resonate at higher fundamental frequency while larger objects tend to resonate at lower fundamental frequency (Ohala & Eukel, 1987). Indeed, recent evidence shows that formant F2 plays a central role in sound-symbolic mappings of size-sound correspondence when participants judge whether a heard speech sound is small or large (Knoeferle, Li, Maggioni & Spence, 2017).

Given that there might be several different mechanisms that can provide a basis for sound symbolic phenomena (Sidhu & Pexman, 2018), another factor that might explain some of these kinds of sound-meaning associations is that people have a tendency to mimic event and object properties in the external world with movements of lips and the tongue (Vainio, Tiainen, Tiippana, Rantala, & Vainio, 2017b; Ramachandran & Hubbard, 2001; Sidhu & Pexman, 2018). For example, regarding the relationship between [i]-‘small’ and [ɑ]-‘large’, the association might be also based on the size of an oral cavity mimicking the referent object size. As such, this view has obvious similarities with the view assuming that people have an implicit tendency to mimic communicative and manipulative hand actions with movements of the mouth, lips and tongue (e.g., Hewes, 1973). It has been, in fact, proposed that some sound-meaning associations, such as the relationship between [i] and ‘small’, might be based on the shape of the articulatory organs mimicking the manual action used to grasp and manipulate the referent object (Ramachandran & Hubbard, 2001; Vainio et al., 2018). This view is supported by the finding that the vowels [i] and [ɑ] are not only associated with the execution of the precision and power grip responses, respectively (Vainio et al., 2013), but they are also systematically associated with these grip types when participants are required to pronounce these vowels while viewing images of these grip types (Vainio et al., 2017a). This account, however, presumes that this

mimicking process operates mostly implicitly through neural connections that link grasp-related motor processes to articulatory motor processes rather than intentionally attempting to copy these hand actions with articulatory organs.

Another sound symbolism phenomenon that has been proposed to reflect mimicking of manual action with orofacial gesturing was put forward by Wallace (1881) who suggested that in many languages articulating the word “go” involves the lip protrusion, required for producing a rounded vowel, as if the manual pointing gestures would be replaced by a lip-pointing gesture. More recently, Ramachandran and Hubbard (2001) have similarly proposed that in many languages articulating the word “you” involves the lip protrusion and the word “me” involves spreading ones lips inwards, as if these articulatory gestures would mimic manual pointing that is directed outwards or inwards, respectively. According to my knowledge, however, there is only one study that has explored this issue specifically. Wichmann, Holman and Brown (2010) have reported the cross-linguistic study in which they did not find evidence for a relationship between vowel roundness and deictic words. Instead, they found that the back vowel [ɑ] was relatively frequently included in the pronoun that points to the speaker whereas the front vowel [i] was more frequently included in the pronoun that points to the hearer. Although this study provides only preliminary evidence for this phenomenon, and more investigation on this topic is needed, it is however in line with the reach-vowel effect (Vainio et al., 2015) that associates the forward directed hand movements with pronunciation of front vowels and the backward directed hand movements with back vowels. Hence, it is possible that Wallace (1881) was partially right with his intuition. In the light of empirical evidence, however, it seems more likely that if manual pointing is mirrored in the articulatory organs, the mirroring is more related to tongue movements than lip movements.

It has been proposed that one of the primary advantages of sound-meaning associations is that they might facilitate word learning in infants (Imai & Kita, 2014). Indeed, people learn the actual English equivalents for Japanese words more accurately using such associations than when learning randomly paired meanings (Nygaard, Cook & Namy, 2009). According to this

view, even pre-speech infants are sensitive to sound-meaning associations, and consequently regularities in sound-meaning pairings of words that children hear in their environment facilitate acquiring new words. As a consequence, if it were the case that some of the sound-meaning associations were indeed based on the shape of the articulatory organs mimicking the manipulative (e.g., a grip type) or communicative (e.g., pointing) manual actions, it could be assumed that learning new words would also be facilitated by these hand-mouth connections. That is, word learning during early childhood might be to some extent facilitated by regularities between the shape of the articulatory organs required to pronounce a specific word and the manual action associated with that word.

In addition to influencing the language development, sound symbolic mappings might be linked to evolution of language. It has also been proposed that sound symbolism might, for example, be a vestige of protolanguage showing that evolutionarily early speech was largely based on symbolic use of cross-modal mappings (Kita, 2008). As stated above, this view presumes that sound symbolic mappings are partially based on innate tendencies to mimic events, actions and objects. Initially, this mimicking might have occurred in relation to hand gestures that, in turn, were gradually replaced by mimicking based on the movements of lips and the tongue (Hewes, 1973). In turn, these oral mimicking tendencies might have been adapted into words (Ramachandran & Hubbard, 2001). Hence, given the tendency of mapping objects, events and actions to phonetic sounds, it is possible that the connection between mouth and hand processes might have boosted the transmitting mimicry-related components of communication from hand to mouth that, in turn, might have facilitated the evolution of speech.

5. Conclusion

Mounting evidence demonstrates anatomical and functional connections between mouth movements and manipulative hand movements. These connections can operate at the non-verbal as well as verbal level associating, for example, a specific grasp type with a specific

articulatory gesture. In addition, evidence suggests that these connections are already operational in neonates and can influence early development of speech. The present review paper emphasizes that these kinds of systematic connections between different grip types and articulatory gestures might facilitate development of certain articulatory gestures during early speech development. As an example, development of precision and power grip might facilitate development of specific apical and dorsal consonants, respectively. In addition, these grip-mouth connections could have even played a role in evolution of language by facilitating the emergence of certain articulatory gestures (such as [t] and [k]) at the initial stages of speech evolution as well as boosting the occurrence of these articulatory gestures in languages developed after the emergence of speech. The present account also proposes that some phenomena that have been previously explained in terms of sound-meaning pairings can be, in fact, partially based on interactions between processes that program manual and articulatory actions. Although the present paper presents some novel proposals, these proposals mostly concur with several previously proposed theories of speech evolution and development. As such, the present views can be taken as a continuation and elaboration of accounts assuming that motor processes related to manual manipulation and grasp actions play a significant role in speech evolution and development. In summary, the primary message of this paper is that processes related to articulatory gestures – the primary speech units on which complex human languages are built – are interacting with processes related to manual grasping and manipulation movement, and these interactions might have specific functional roles in the development and evolution of speech.

Acknowledgments

I would like to thank Martti Vainio, Mikko Tiainen, Kaisa Tiippana, Naeem Komeilipoor and Aleksi Rantala for their contribution to many views presented in this article

References

Arbib, M. A. (2002). fO *The Mirror System, Imitation, and the Evolution of Language. Imitation in animals and artifacts*, 229.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and brain sciences*, 28(2), 105-124.

Arnstein, D., Cui, F., Keysers, C., Maurits, N. M., & Gazzola, V. (2011). μ -suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal, and SI cortices. *Journal of Neuroscience*, 31, 14243-14249.

Babkin, P. S. (1958). Early postnatal establishment of reflex activity in man. *Fiziologicheskii zhurnal SSSR imeni IM Sechenova*, 44(10), 922-927.

Barlow, S. M., Radder, J. P. L., Radder, M. E., & Radder, A. K. (2010). Central pattern generators for orofacial movements and speech. In *Handbook of behavioral neuroscience* (Vol. 19, pp. 351-369). Elsevier.

Borghgi, A. M., Flumini, A., Cimatti, F., Marocco, D., & Scorolli, C. (2011). Manipulating objects and telling words: a study on concrete and abstract words acquisition. *Frontiers in psychology*, 2, 15.

Borghgi, A. M., Barca, L., Binkofski, F., & Tummolini, L. (2018). Abstract concepts, language and sociality: from acquisition to inner speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 20170134.

Borghini, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2018). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of life reviews*.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.

Bates, E. (1976). *Language and context: The acquisition of pragmatics*. Academic Press.

Belsky, J. & Most, R.K. (1981). From Exploration to Play: A Cross-Sectional Study of Infant Free Play Behavior. *Developmental Psychology*, 17, 5,630-639

Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols & J. Ohala (Eds.), *Sound symbolism* (pp. 76–93). New York: Cambridge University Press.

Bernardis, P., Bello, A., Pettenati, P., Stefanini, S., & Gentilucci, M. (2008). Manual actions affect vocalizations of infants. *Experimental Brain Research*, 184(4), 599-603.

Binkofski, F., Buccino, G., Stephan, K. M., Rizzolatti, G., Seitz, R. J., & Freund, H. J. (1999). A parieto-premotor network for object manipulation: evidence from neuroimaging. *Experimental Brain Research*, 128(1-2), 210-213.

Bishop, D. V. (2002). Motor immaturity and specific speech and language impairment: Evidence for a common genetic basis. *American Journal of Medical Genetics*, 114(1), 56-63.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818-10823.

Borra, E., Belmalih, A., Gerbella, M., Rozzi, S., & Luppino, G. (2010). Projections of the hand field of the macaque ventral premotor area F5 to the brainstem and spinal cord. *Journal of Comparative Neurology*, 518, 2570-2591.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., ... & Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. *Journal of cognitive neuroscience*, 16(1), 114-126.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., ... & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European journal of neuroscience*, 13(2), 400-404.

Butterworth, G., & Hopkins, B. (1988). Hand-mouth coordination in the new-born baby. *British Journal of Developmental Psychology*, 6(4), 303-314.

Butterworth, G., Verweij, E., & Hopkins, B. (1997). The development of prehension in infants: Halverson revisited. *British Journal of Developmental Psychology*, 15(2), 223-236.

Cerri, G., Shimazu, H., Maier, M. A., & Lemon, R. N. (2003). Facilitation from ventral premotor cortex of primary motor cortex outputs to macaque hand muscles. *Journal of Neurophysiology*, 90(2), 832-842.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and social psychology bulletin*, 25(2), 215-224.

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in cognitive sciences*, 7(7), 300-307.

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience*, 3(7), 553.

Cohen, A. H., Rossignol, S., & Grillner, S. (1988). Neural control of rhythmic movements in vertebrates. Wiley.

Corballis, M. C. (2003). *From hand to mouth: The origins of language*. Princeton University Press.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Marry.

D'ausilio, A., Maffongelli, L., Bartoli, E., Campanella, M., Ferrari, E., Berry, J., & Fadiga, L. (2014). Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. *Phil. Trans. R. Soc. B*, 369(1644), 20130418.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381-385.

Davare, M., Lemon, R., & Olivier, E. (2008). Selective modulation of interactions between ventral premotor cortex and primary motor cortex during precision grasping in humans. *The Journal of physiology*, 586(11), 2735-2742.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., ... & Fazio, F. (1994). Mapping motor representations with positron emission tomography. *Nature*, 371(6498), 600.

Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., ... & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115(6), 1753-1768.

de Saussure F. 1983 *Course in general linguistics* [Transl. R. Harris]. La Salle, IL: Open Court

Desmurget, M., Richard, N., Harquel, S., Baraduc, P., Szathmari, A., Mottolese, C., & Sirigu, A. (2014). Neural representations of ethologically relevant hand/mouth synergies in the human precentral gyrus. *Proceedings of the National Academy of Sciences*, 111(15), 5718-5722.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1), 176-180.

Dreyer, F. R., & Pulvermüller, F. (2018). Abstract semantics in the motor system?—An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex*, 100, 52-70.

Fagan, M. K., & Iverson, J. M. (2007). The influence of mouthing on infant vocalization. *Infancy*, 11(2), 191-202.

Fant C. (1960). *Acoustic Theory of Speech production*. Mouton's: The Hague.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... & Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i-185.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, 17(8), 1703-1714.

Finlay, J. C., & McPhillips, M. (2013). Comorbid motor deficits in a clinical sample of children with specific language impairment. *Research in Developmental Disabilities*, 34(9), 2533-2542.

Fogel, A., & Hannan, T. E. (1985). Manual actions of nine-to fifteen-week-old human infants during face-to-face interaction with their mothers. *Child Development*, 1271-1279.

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

Friederici, A. D. (2017). Evolution of the neural language network. *Psychonomic bulletin & review*, 24(1), 41-47.

Futagi, Y., Yanagihara, K., Mogami, Y., Ikeda, T., & Suzuki, Y. (2013). The Babkin reflex in infants: clinical significance and neural mechanism. *Pediatric Neurology*, 49(3), 149-155.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2), 593-609.

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasping with hand and mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology* 86: 1885–1699.

Gentilucci, M., Stefanini, S., Roy, A. C., & Santunione, P. (2004). Action observation and speech production: study on children and adults. *Neuropsychologia*, 42(11), 1554-1567.

Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: a gradual transition. *Neuroscience & Biobehavioral Reviews*, 30(7), 949-960.

Gentilucci, M., & Volta, R. D. (2008). Spoken language and arm gestures are controlled by the same motor control system. *The Quarterly Journal of Experimental Psychology*, 61(6), 944-957.

Gentilucci M, Campione GC (2011) Do Postures of Distal Effectors Affect the Control of Actions of Other Distal Effectors? Evidence for a System of Interactions between Hand and Mouth. *Plos ONE* 6, 5: e19793.

Ghio, M., Vaghi, M. M. S., & Tettamanti, M. (2013). Fine-grained semantic categorization across the abstract and concrete domains. *PloS one*, 8(6), e67090.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to language via the mirror neuron system*, 215-49.

Gleenfield, P. (1991). Language, tools and brain: the ontogeny and phylogeny of hierarchically organised sequential behaviour. *Behavioural and Brain Sciences*, 14, 531-595.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, 1(1), 23-64.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic bulletin & review*, 9(3), 558-565.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual review of psychology*, 64, 257-283.

Graziano, M. S., Taylor, C. S., & Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron*, 34(5), 841-851.

Green, J. B. (1967). An electromyographic study of mirror movements. *Neurology*, 17(1), 91-91.

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research*, 43(1), 239-255.

Halverson HM (1931) An experimental study of prehension in infants by means of systematic cinema records. *Genetic Psychology Monographs* 10: 107–283.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2), 301-307.

Hauser, M. D., & Fitch, W. T. (2003). What are the uniquely human components of the language faculty?. *Studies in the Evolution of Language*, 3, 158-181.

Hewes, G. W., Andrew, R. J., Carini, L., Choe, H., Gardner, R. A., Kortlandt, A., ... & Rumbaugh, D. G. (1973). Primate communication and the gestural origin of language [and comments and reply]. *Current Anthropology*, 14(1/2), 5-24.

Hewes, G. W. (1977). A model for language evolution. *Sign Language Studies*, (15), 97-168.

Higginbotham, D.R., Isaak, M.I., & Domingue, J.M. (2008). The exaptation of manual dexterity for articulate speech: an electromyogram investigation. *Experimental Brain Research* 186: 603–609.

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, 36(2), 149-171.

Holmgren, K., Lindblom, B., Aurelius, G., Jalling, B., & Zetterstrom, R. (1986). On the phonetics of infant vocalization. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 51-63). New York Stockton Press.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286(5449), 2526-2528.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Phil. Trans. R. Soc. B*, 369(1651), 20130298.

Irwin, O. C., & Chen, H. P. (1946). Development of speech during infancy: curve of phonemic types. *Journal of experimental psychology*, 36(5), 431.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 367-371.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, 37(2), 229-261.

Jeannerod, M. (1988). *The neural and behavioural organization of goal-directed movements*. Clarendon Press/Oxford University Press.

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, 26(2), 235-258.

Kawashima, R., Itoh, H., Ono, S., Satoh, K., Furumoto, S., Gotoh, R., ... & Yanagisawa, T. (1996). Changes in regional cerebral blood flow during self-paced arm and finger movements. A PET study. *Brain Research*, 716(1-2), 141-148.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, 72(2), 353-365.

Kita, S. (2008). World-view of protolanguage speakers as inferred from semantics of sound symbolic words: a case of Japanese mimetics. *Origins of language* (ed. & Masataka N), pp. 25–38. Tokyo, Japan: Springer.

Knight, R. A. (2012). *Phonetics: A coursebook*. Cambridge University Press.

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific reports*, 7(1), 5562.

Komeilipoor, N., Tiainen, M., Tiippana, K., Vainio, M., & Vainio, L. (2016). Excitability of hand motor areas during articulation of syllables. *Neuroscience Letters*, 620, 154-158.

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in cognitive sciences*, 22(10), 896-910.

Kunesch, E., Binkofski, F., & Freund, H. J. (1989). Invariant temporal characteristics of manipulative hand movements. *Experimental Brain Research*, 78(3), 539-546.

Laakso, M. L., Poikkeus, A. M., Katajamäki, J., & Lyytinen, P. (1999). Early intentional communication as a predictor of language development in young toddlers. *First Language*, 19(56), 207-231.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, 74(6), 431.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in cognitive sciences*, 13(3), 110-114.

Lund, J. P., & Kolta, A. (2006). Brainstem circuits that control mastication: Do they have anything to say during speech?. *Journal of communication disorders*, 39(5), 381-390.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21(4), 499-511.

Mason, R. M., & Proffit, W. R. (1974). The tongue thrust controversy: background and recommendations. *Journal of Speech and Hearing Disorders*, 39(2), 115-132.

Masuda, K. (2007). The physical basis for phonological iconicity. *Insistent images*, 57-71.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Meguerditchian, A., Plouvier, M., Pruetz, J. D., & Hopkins, W. D. (2014). From hand to mouth: fine precision grip during mutual grooming elicited wide lip movements in wild fongoli chimpanzees. In *Evolution of Language: Proceedings of the 10th International Conference (EVOLANG10)* (pp. 487-488).

Missiuro, W. (1963). Studies on developmental stages of children's reflex reactivity. *Child Development*, 33-41.

Moore, C. A., & Ruark, J. L. (1996). Does speech emerge from earlier appearing oral motor behaviors?. *Journal of Speech, Language, and Hearing Research*, 39(5), 1034-1047.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology*, 20(8), 750-756.

Myowa-Yamakoshi, M., & Takeshita, H. (2006). Do human fetuses anticipate self-oriented actions? A study by four-dimensional (4D) ultrasonography. *Infancy*, 10(3), 289-301.

Nelson, E. L., Campbell, J. M., & Michel, G. F. (2014). Early handedness in infancy predicts language ability in toddlers. *Developmental Psychology*, 50(3), 809.

Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology*, 45(1), 53-75.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112(1), 181-186.

Ohala, J. J., & Eukel, B. W. (1987). Explaining the intrinsic pitch of vowels. *In honor of Ilse Lehiste*, 207-215.

Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of child language*, 3(1), 1-11.

Owen, S. E., & McKinlay, I. A. (1997). Motor difficulties in children with developmental disorders of speech and language. *Child: Care, Health and Development*, 23(4), 315-325.

Paget, R. A. (1944). The origin of language. *Science*, 99(2558), 14-15.

Papoutsis, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. *Cerebral Cortex*, 19(9), 2156-2165.

Pedroso, F. S., & Rotta, N. T. (2004). Babkin reflex and other motor responses to appendicular compression stimulus of the newborn. *Journal of Child Neurology*, 19(8), 592-596.

Perkell, J. S. (1996). Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modeling. *Journal of Phonetics*, 24(1), 3-22.

Petrides, M., Cadoret, G., & Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature*, 435(7046), 1235.

Petrides, M., & Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS biology*, 7(8), e1000170.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature reviews neuroscience*, 11(5), 351.

Redican, W. K. (1975). Facial expressions in nonhuman primates. *Primate Behavior*., 103-194.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of Consciousness Studies*, 8(12), 3-34.

Rizzolatti, G., Camarda, R., Fogassi, I., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey: II. area F5 and the control of distal movements. *Experimental Brain Research*, 71:491–507.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., & Fazio, F. (1996). Localization of grasp representations in humans by PET: 1. Observation versus execution. *Experimental Brain Research*, 111(2), 246-252.

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

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

Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalography and clinical neurophysiology*, 106(4), 283-296.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron*, 31(6), 889-901.

Rochat, P. (1989). Object manipulation and exploration in 2-to 5-month-old infants. *Developmental Psychology*, 25(6), 871.

Rochat, P. (1993). Hand-mouth coordination in the newborn: Morphology, determinants, and early development of a basic act. *In Advances in psychology* (Vol. 97, pp. 265-288). North-Holland.

Rowe, M. L., & Goldin - Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science*, 12(1), 182-187.

Ruff, H. A., Saltarelli, L. M., Capozzoli, M., & Dubiner, K. (1992). The differentiation of activity in infants' exploration of objects. *Developmental Psychology*, 28(5), 851.

Rvachew, S., Alhaidary, A., Mattock, K., & Polka, L. (2008). Emergence of the corner vowels in the babble produced by infants exposed to Canadian English or Canadian French. *Journal of Phonetics*, 36(4), 564-577.

Salmelin, R., & Sams, M. (2002). Motor cortex involvement during verbal versus non-verbal lip and tongue movements. *Human Brain Mapping*, 16(2), 81-91.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12, 239-255.

Sauerland, E. K., & Mitchell, S. P. (1975). Electromyographic activity of intrinsic and extrinsic muscles of the human tongue. *Texas reports on biology and medicine*, 33(3), 444-455.

Schuster, C. S., & Ashburn, S. S. (1992). *The process of human development: A holistic life-span approach*. Lippincott Williams & Wilkins.

Serrien, D. J., Bogaerts, H., Suy, E., & Swinnen, S. P. (1999). The identification of coordination constraints across planes of motion. *Experimental Brain Research*, 128(1-2), 250-255.

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, 25(5), 1619-1643.

Swinnen, S. P., Jardin, K., Meulenbroek, R., Dounskaia, N., & Den Brandt, M. H. V. (1997). Egocentric and allocentric constraints in the expression of patterns of interlimb coordination. *Journal of Cognitive Neuroscience*, 9(3), 348-377.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology*, 64(12), 2392-2404.

Topolinski, S., Maschmann, I. T., Pecher, D., & Winkielman, P. (2014). Oral approach-avoidance: Affective consequences of muscular articulation dynamics. *Journal of personality and social psychology*, 106(6), 885.

Topolinski, S., & Boecker, L. (2016). Minimal conditions of motor inductions of approach-avoidance states: The case of oral movements. *Journal of Experimental Psychology: General*, 145(12), 1589.

Tiainen, M., Tiippana, K., Vainio, M., Peromaa, T., Komeilipoor, N., & Vainio, L. (2016). Selective influences of precision and power grips on speech categorization. *PloS one*, 11(3), e0151688.

Tiainen, M., Tiippana, K., Vainio, M., Komeilipoor, N., & Vainio, L. (2017a). Interaction in planning vocalizations and grasping. *The Quarterly Journal of Experimental Psychology*, 70(8), 1590-1602.

Tiainen, M., Lukavský, J., Tiippana, K., Vainio, M., Šimko, J., Felisberti, F., & Vainio, L. (2017b). Connections of grasping and horizontal hand movements with articulation in Czech speakers. *Frontiers in Psychology*, 8, 516.

Tiainen, M., Tiippana, K., Paavilainen, P., Vainio, M., & Vainio, L. (2017c). Mismatch negativity (MMN) to speech sounds is modulated systematically by manual grip execution. *Neuroscience Letters*, 651, 237-241.

Todor, J. I., & Lazarus, J. A. C. (1986). Exertion level and the intensity of associated movements. *Developmental Medicine & Child Neurology*, 28(2), 205-212.

Trevarthen, C. (1984). How control of movements develops. In H. T. A. Whiting (Ed.), *Human motor actions: Bernstein reassessed*. North-Holland: Elsevier Science Publishers B.V.

Vainio, L., Schulman, M., Tiippana, K., & Vainio, M. (2013). Effect of syllable articulation on precision and power grip performance. *PloS one*, 8(1), e53061.

Vainio, L., Tiainen, M., Tiippana, K., & Vainio, M. (2014). Shared processing of planning articulatory gestures and grasping. *Experimental Brain Research*, 232(7), 2359-2368.

Vainio, L., Tiainen, M., Tiippana, K., Komeilipoor, N., & Vainio, M. (2015). Interaction in planning movement direction for articulatory gestures and manual actions. *Experimental Brain Research*, 233(10), 2951-2959.

Vainio, L., Rantala, A., Tiainen, M., Tiippana, K., Komeilipoor, N., & Vainio, M. (2017a). Systematic influence of perceived grasp shape on speech production. *PloS one*, 12(1), e0170221.

Vainio, L., Tiainen, M., Tiippana, K., Rantala, A., & Vainio, M. (2017b). Sharp and round shapes of seen objects have distinct influences on vowel and consonant articulation. *Psychological Research*, 81(4), 827-839.

Vainio, L., & Tiainen, M. (2018). Inter-limb coupling of proximal and distal hand actions. *Experimental Brain Research*, 236(1), 153-160.

Vainio, L., Tiippana, K., Tiainen, M., Rantala, A., & Vainio, M. (2018). Reaching and grasping with the tongue: Shared motor planning between hand actions and articulatory gestures. *Quarterly Journal of Experimental Psychology*, 1747021817738732.

Waciewicz, S., Żywiczyński, P., & Orzechowski, S. (2016). Visible movements of the orofacial area. *Gesture*, 15(2), 250-282.

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

Wallace, P. S., & Wishaw, I. Q. (2003). Independent digit movements and precision grip patterns in 1–5-month-old human infants: hand-babbling, including vacuous then self-directed hand and digit movements, precedes targeted reaching. *Neuropsychologia*, 41(14), 1912-1918.

Watanabe, J., Sugiura, M., Miura, N., Watanabe, Y., Maeda, Y., Matsue, Y., & Kawashima, R. (2004). The human parietal cortex is involved in spatial processing of tongue movement—an fMRI study. *Neuroimage*, 21(4), 1289-1299.

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

Wellman, B. L., Case, I. M., Mengert, I. G., & Bradbury, D. E. (1931). Speech sounds by young children.

Wichmann, S., Holman, E. W., & Brown, C. H. (2010). Sound symbolism in basic vocabulary. *Entropy*, 12(4), 844-858.

Woodward, A. I. (2004). Infants' Use of Action Knowledge to Get a Grasp. *Weaving a lexicon*, 149.

Yensen, R. (1965). A factor influencing motor overflow. *Perceptual and Motor Skills*, 20(3), 967-968.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256(5058), 846-849.

Zelaznik, H. N., & Goffman, L. (2010). Generalized motor abilities and timing behavior in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 53(2), 383-393.

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39, 151–168.

Figure

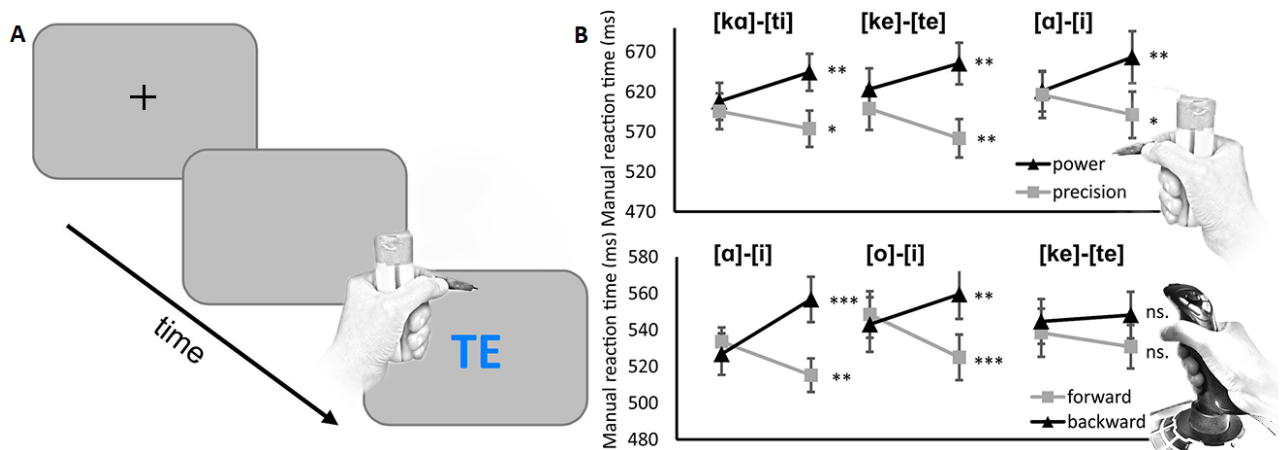


Figure 1. A) The paradigm used for investigating the articulation-grip effect (Vainio et al., 2013) and the vowel-reach effect (Vainio et al., 2015). In this dual-action paradigm, participants are holding the grip devices (regarding the grip effects) or the joystick (regarding the reach effect) in their right hand, and they are visually presented with, for example, the meaningless syllable 'TE' or the vowel 'A' in the green or blue color on the monitor. The participants are required to perform either precision or power grip response (regarding the grip effects) or move their hand forward or backward (regarding the reach effect) according to the color of the stimulus and simultaneously pronounce the vowel/syllable. B) Manual responses are performed faster when there is a congruency between the vowel/syllable and the response (Tiainen et al., 2017b). A similar effect is also observed in vocal responses. In the consonant-grip effect (the upper graph), the precision grip response is facilitated when the consonant in the consonant-vowel syllable is [t] in comparison to the consonant [k] whereas the power grip response is facilitated when the consonant is [k] in comparison to the consonant [t]. In the vowel-grip effect (the upper graph), production of the open vowel [a] facilitates the power grip responses and production of the closed vowel [i] facilitates precision grip responses. In the vowel-reach effect (the lower graph), the forward directed hand movements are facilitated by simultaneous pronunciation of front vowels (e.g., [i]) whereas the backward directed hand movements are facilitated by simultaneous pronunciation of back vowels (e.g., [o]).