5

Articulatory phonetics

Bryan Gick, Murray Schellenberg, Ian Stavness, and Ryan C. Taylor

Articulatory phonetics is broadly concerned with understanding how humans use body structures to produce speech sounds. Articulatory phoneticians use a wide range of tools and techniques to study different aspects of these structures, from their physical form, function and control to their evolution, development and patterns of use. As such, the field of articulatory phonetics interacts with a wide range of other disciplines such as neuroscience, speech-language pathology, motor control, dentistry, cognitive science, biomedical engineering, developmental psychology, otolaryngology, and of course other subfields of phonetics and linguistics.

If articulatory phonetics as a field is indeed concerned with how speech is realized through physical actions, then it stands to reason that our models must include – or at least provide links to – parts of the body. Any attempt to study or describe a thing (such as a body part) is an attempt to reduce its complexity – and the speech apparatus is nothing if not complex. The human vocal tract is a continuous, high-dimensional space with practically unlimited degrees of freedom, and articulatory phoneticians have naturally come up with different ways of reducing the complexity of this space. Of course, when we reduce complexity we also remove or omit potentially important information, so it is important to consider carefully which parts our models and descriptions should include, and which they may leave out.

The present chapter explores some of the approaches traditionally used in the field of articulatory phonetics to describe the vocal tract, then goes on to consider the vocal tract from the lips to the larynx through the lens of a more embodied approach. We use the term "embodied" here in the sense of "situated in the body," where the body includes not just the "meat," but also all aspects of the nervous system including (but not limited to) both motor systems and internal and external sensory systems. Articulatory phonetics is in this sense inherently "embodied," as it attempts to make direct reference to body structures and their communicative functions. However, as articulatory phonetics is relatively young as a scientific (rather than descriptive) field, the broader implications of a more fully embodied approach to theories of speech sounds remain to be explored.

One of the predominant methods phoneticians have used to simplify the speech production system has been to reduce the vocal tract to a single central plane (the *midsagittal* plane). The midsagittal approach to reducing the dimensionality of the vocal tract has been

reinforced over the past century by the development of tools such as X-ray that provide images of the midline of the vocal tract (Russell, 1933) and by the utility of the midsagittal approximation of vocal tract area for acoustic modeling (Serrurier et al., 2012; though this method of approximation is not without its challenges, e.g., Anderson et al., 2015). Figure 5.1 shows a two-dimensional outline of the vocal tract. The midsagittal outline, familiar to any student of phonetics, is fairly simple and appears to include all of the major "parts" commonly referred to by phoneticians (tongue, lips, velum, etc.). A serious limitation of the midsagittal reduction method is that it indiscriminately omits everything outside of that plane – indeed, being only an outline of the outer surfaces of the major structures along the midline, this representation omits reference even to any internal structure that may intersect this plane. The resulting outline contains no information about structures, for example, at the sides of the vocal tract, and no reference to any neurally accessible components of the system such as muscles and nerves. While the midsagittal section is certainly helpful for an initial descriptive presentation of the vocal apparatus, it is important to question whether it contains sufficient information to be useful as a scientific model of how the human body is used to articulate speech sounds.

Another way articulatory phoneticians have reduced the dimensionality of the vocal tract is by describing it in terms of a set of fixed, anatomically defined structures, often referred to using familiar body-part terms such as "lips" or "tongue tip." These terms are seldom given technical definitions, leaving it to the reader to interpret their precise meaning and

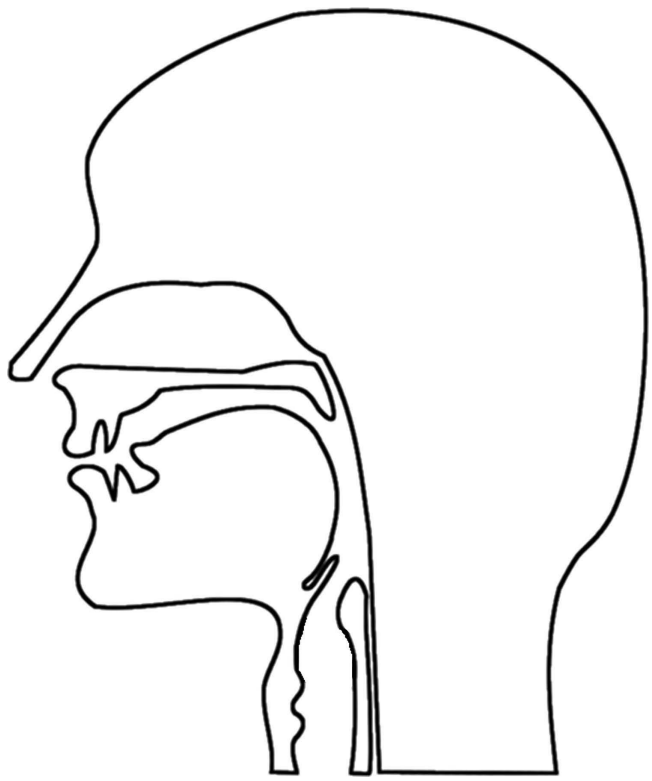

Figure 5.1 A two-dimensional outline of the vocal tract along the midsagittal plane.

scope. Not only does this leave our models open to significant cultural interpretation concerning the specific referents of body part terms (see, e.g., Enfield et al., 2006; Wierzbicka, 2007 for contrasting viewpoints on this subject), but dividing the vocal tract into anatomical articulators in this way raises the philosophical question of whether it is desirable or even possible to separate anatomical "body parts" from the muscular and neurological systems that control them. Wilson (1998) addresses this question in relation to the defining of hands:

> Should those parts of the brain that regulate hand function be considered part of the hand? . . . Although we understand what is meant conventionally by the simple anatomic term ['hand'], we can no longer say with certainty where the hand itself, or its control or influence, begins or ends in the body.
>
> *(Wilson, 1998: 9)*

This concept is by no means new to articulatory phonetics: In Cooper et al. (1958: 939) "action plans" were used as a way to link speech events to underlying muscular organization, with the goal "to describe speech events in terms of a rather limited number of muscle groups." Later, Liberman et al. (1967: 446) extended this concept to phonological features, arguing that "the distinctive features of a phoneme are closely linked to specific muscles and the neural commands that actuate them." Of course, while it is trivially true that any physical action can be realized only through neuromotor commands and concomitant muscle movements, the challenge is to explore which particular commands and movements are used, and more to the point, which (if any) are useful to us in understanding how speech works.

The following sections describe the "articulators" of speech in more embodied terms, allowing us to consider different regions of the vocal tract in turn, from the lips to the larynx. Each section develops concepts that emerge from a fuller description of the "parts" of speech as they help to deepen our understanding of the sound systems of human languages. It is important to note that, although the following sections have been given familiar anatomically based titles such as "Lips" and "Tongue," this usage should be taken as strictly descriptive.

Lips

The lips are the most visible part of the vocal tract, making them an easy place to start thinking about how the body makes speech sounds. Lips are associated with the production of bilabial speech sounds such as /b/, /m/ and /w/ and with labio-dental sounds such as /f/ and /v/. Despite the distinct structures and functions of the upper lip and the lower lip, the English language reserves a single term ("lip") for both structures, referring roughly to the fleshy parts immediately adjacent to the vermilion border (the often darker-hued bands immediately surrounding the mouth). In contrast, Jahai, a Mon-Khmer language spoken in Malaysia, uses a separate word for each lip, with *nus* "upper lip" including all the fleshy parts between the nose and mouth and *tnɨt* "lower lip" including all the fleshy parts of the lower lip and chin (Burenhult, 2006). Interestingly, defined in this way, the Jahai "articulators" could be said to reflect more accurately the underlying innervation, with the *nus* innervated by the buccal branches of the facial nerve and the *tnɨt* serviced by the marginal mandibular branches of the facial nerve. Anatomists have long since moved beyond trying to define the lips as a discrete and neatly contained anatomical structure (Lightoller, 1925). Rather, what we refer to as "lips" comprises a complex network of muscles and other tissues

extending roughly from the chest to the cranium (see Figure 5.2), along with all of its associated skeletal, neural and metabolic structures, taking on different forms depending on its immediate function. So, where does one draw the line around "lips"?

The notion of "lips" becomes still more complicated as soon as we begin to use them to produce speech: Consider that the set of muscles recruited to produce the flat lip closure for /b/ (mainly marginal orbicularis oris [OOm], risorius [RIS] and mentalis [MEN]) is completely different from that recruited for the rounding of /w/ (mainly peripheral orbicularis oris [OOp]) (Gick et al., 2011), while /f/ and /v/ only use inferior OO (OOi) muscles to

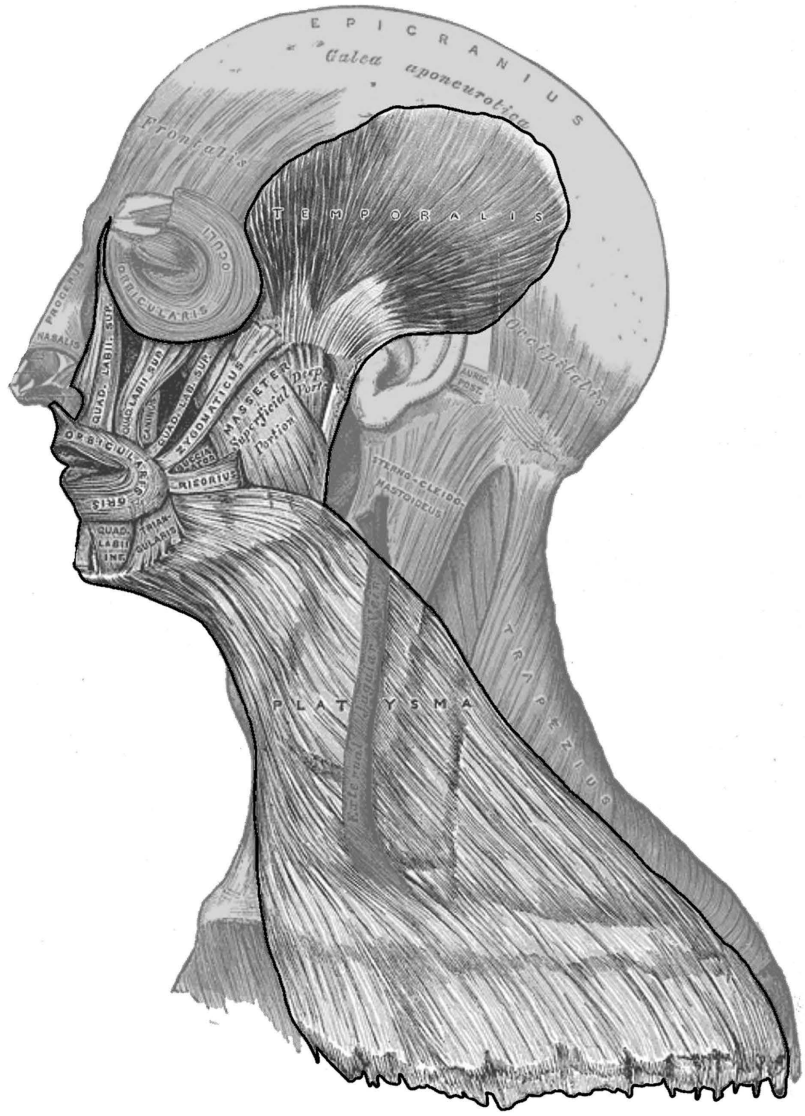

Figure 5.2 Muscles contributing to lip movement (outlined in black).[1]

Source: Adapted by B. Gick and R.C. Taylor, from Gray and Lewis (1918) Plate 378, [public domain]

close the bottom lip against the upper teeth (see Figure 5.3). That is to say, if articulators are defined so as to include the associated muscular and neural structures, then the "lips" we use to produce /b/ are quite literally a different body part from the "lips" we use to produce /w/ or the "lips" we use to produce /f/. At this point, the use of the term "lips" as a formal part of our phonetic description seems more of a liability than an aid. It would make more sense to define three separate structures that control these three separate sounds. Thus, if we really want to understand articulatory phonetics, we should ask not "what are lips?" or even "how do we control our lips?" but rather "how do the various systems of the body work together to produce the movements that result in the sound /w/?" This then raises other interesting questions, such as why do so many different languages realize /w/ (or /b/ or /f/) in similar ways?

When articulatory phonetics is viewed from an embodied point of view, we find that there are certain movements that our bodies can produce with particular efficiency and reliability – and that these robust movements appear in the phonetic inventories of language after language. These robust movements may be thought of as attractors or "sweet spots" in the action space that require less precise control while still producing consistent articulatory outputs. Examples of physical properties that permit consistent movement with imprecise control include situations of contact between body parts in which movement abruptly stops or changes, tissue stiffness that resists movement, and limits on muscle force-generating capacity. All of these are examples of the kind of "quantal" biomechanical-articulatory relations that have long been described for speech (e.g., Stevens, 1989; Fujimura, 1989; Schwartz et al., 1997) but seldom quantified; computer simulations representing the biomechanics of the body parts associated with speech articulation can be useful in elucidating these properties. Biomechanical simulations have shown quantal relations to correspond with the kinds of movements often used to make speech sounds (e.g., Buchaillard et al., 2009; Nazari et al., 2011; Gick et al., 2014, Moisik and Gick, 2017).

Three-dimensional biomechanical simulations of lip musculature and other structures can be useful in identifying the labial movements that show quantal properties, and the different groupings of muscles that drive these movements (e.g., Gick et al., 2011). The graph

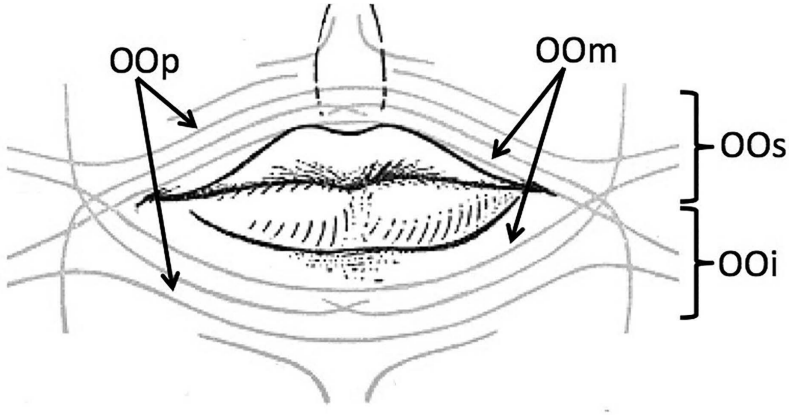

Figure 5.3 Schematic diagram illustrating the outer ring OOp (peripheral), inner ring OOm (marginal), upper half OOs (superior), and lower half OOi (inferior) of the orbicularis oris (OO) muscle.

Source: Adapted by B. Gick, from Gray and Lewis (1918), Plate 381 [public domain]

in Figure 5.4 illustrates how two different muscle groupings are used to produce reliable lip constrictions of different sizes (as for /b/ and /w/); in other words, each discrete lip movement (say, closure for /b/ or rounding for /w/) results from the action of a separate functional body structure, each having its own distinct musculature, innervation, and functional outcome. Semi-closed structures of this kind have sometimes been associated with the term "modules" or "muscle synergies" in the motor control literature (e.g., Safavynia and Ting, 2012), because they act as distinct and relatively autonomous functional units of action in the system. As these neuromuscular modules are built to take advantage of "biomechanical affordances" (Ting and Chiel, 2015), it is important to consider biomechanical properties of speech movements.

The term "modules" in this context refers to groupings of muscles and nerves whose activations result in robust, reliably produced movements (D'Avella and Bizzi, 2005). As Loeb et al. (2000:79) describe in their study of limb control:

> in the large space of simulated muscle combinations there exists a well-defined subset of synergies which will stabilize the limb despite activation noise, muscle fatigue, and other uncertainties – and these synergies stabilize the limb at predictable, restricted locations in the workspace.

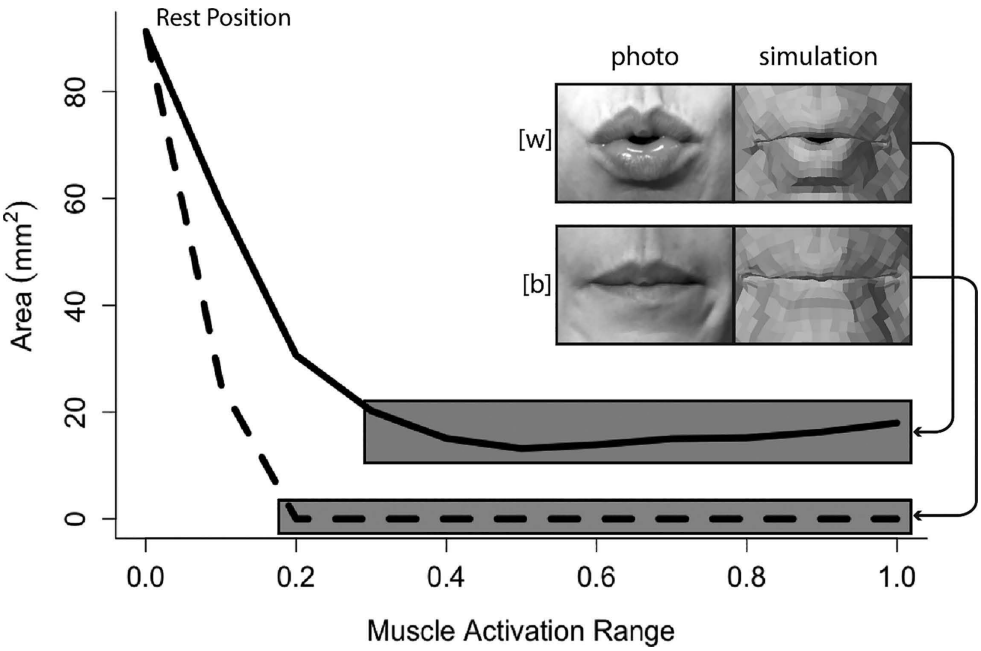

Figure 5.4 Area of lip opening as muscle activation increases; dashed line shows activation of a lip-closing module (OOM + RIS + MEN), solid line shows activation of a lip-rounding module (OOP). Note that each module stabilizes at its own natural size of opening after about 20–30% activation, with stable regions indicated in dotted boxes. Photos to the right show rounded and closed lip shapes for illustration.

Source: Graph adapted from (Gick et al., 2011) with permission. Images generated by C. Chiu using ArtiSynth; www.artisynth.org.

> So, according to our model, these *robust* muscle synergies can be depended on to get the limb to the right place.

Groupings of this kind – of nerves and muscles and the structures they move, tuned by use and feedback, and organized to produce reliable speech movements – form the basis of actionable body parts of speech; these are the "articulators" of articulatory phonetics (see Gick and Stavness, 2013). A prediction of this approach is that there should be gaps in phonetic inventories corresponding to the absence of robust, reliable mechanisms at particular vocal tract locations. Insofar as our phonetic transcription systems, e.g., the International Phonetic Alphabet (IPA; see Figure 5.5), are representative of the phonetic variation observed in the languages of the world, these provide a good sense of what options are or are not available to speakers of a language. In fact, the IPA contains many gaps that are not judged physically impossible, but that are nevertheless unattested. For example, while it is entirely possible to produce a labio-dental stop by compressing the lower lip into the upper teeth with sufficient pressure, as evidenced by its occasional appearance as a blend of [p] and [f] in Belgian Dutch (Verhoeven, 2005) or in clinical speech (e.g., Hagedorn et al., 2014), this sound is not used contrastively in any language, presumably because the upper teeth simply provide a comparatively poor closure surface for the lower lip.

Note that as the structures we colloquially call "lips" may be involved in a variety of different speech sounds, their range of possible actions will potentially be governed by a large number of distinct neuromuscular modules specialized to produce different speech movements (as well as other modules for non-speech functions relating to ingestion, respiration, facial expression, and so on). Such modules make no formal reference to any predefined anatomical structure such as "lips," "upper lip," "lower lip," "jaw," etc. Importantly, what this means for articulatory phonetics is that each "articulator" identified in our model is an actionable unit with a built-in reliable function, enabling us to link every observable or measurable phonetic output with its concomitant embodied properties, including its kinematics, biomechanics, musculature, neural control, and the sensory structures through which it is trained and reinforced.

CONSONANTS (PULMONIC) © 2015 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

Figure 5.5 Excerpt from the IPA Chart, www.internationalphoneticassociation.org/content/ipa-chart, available under a Creative Commons Attribution-Sharealike 3.0 Unported License.

Source: Copyright 2015 International Phonetic Association

Tongue

The tongue is commonly thought of as the primary mover in speech articulation. It is perhaps the body's most flexible structure, interacting with other parts of the oral and pharyngeal tracts to produce a wide range of vowel and consonant sounds, including coronals (e.g., /t/, /n/, /l/), palatals (e.g., /i/, /j/, /c/), velars (e.g., /k/, /g/, /x/) and uvulars (e.g., /g/, /ɴ/, /ʀ/), to mention a few. Taken as a whole anatomical object, the tongue acts as a muscular hydrostat; that is, the volume of the tongue remains constant as it changes shape, so that when the tongue is squeezed in one area by the action of muscles we expect a corresponding passive expansion in another area to maintain the volume, as with a water balloon. Because of its hydrostatic properties, the human tongue has sometimes been compared with other muscular hydrostats in the animal kingdom, such as the octopus arm or squid tentacle (Kier and Smith, 2002). However, while tongues and tentacles may appear similar, and both do indeed have many degrees of freedom (DoF) – so many that it is hard to conceive how they can be controlled – the analogy largely ends here. Human tongues and octopus arms have very different internal structures and are controlled very differently. One way octopuses solve their DoF problem is by bending their arms only at specific locations (Sumbre et al., 2006), almost as if the arm contained bones and joints that help constrain its possible movements. This simplifies the control problem for the octopus's nervous system.

The tongue is also very flexible, but unlike an octopus arm, the tongue is surrounded by hard structures, and it interacts with these to produce its speech movements. Contrary to a "tentacle" analogy, which might lead us to imagine that the sides of the tongue move more-or-less in concert with the midline, the lateral edges of the tongue actually serve their own quite independent and important function – they are responsible for "bracing" the tongue against the upper molars and sides of the hard palate (Figure 5.6). This bracing greatly reduces the

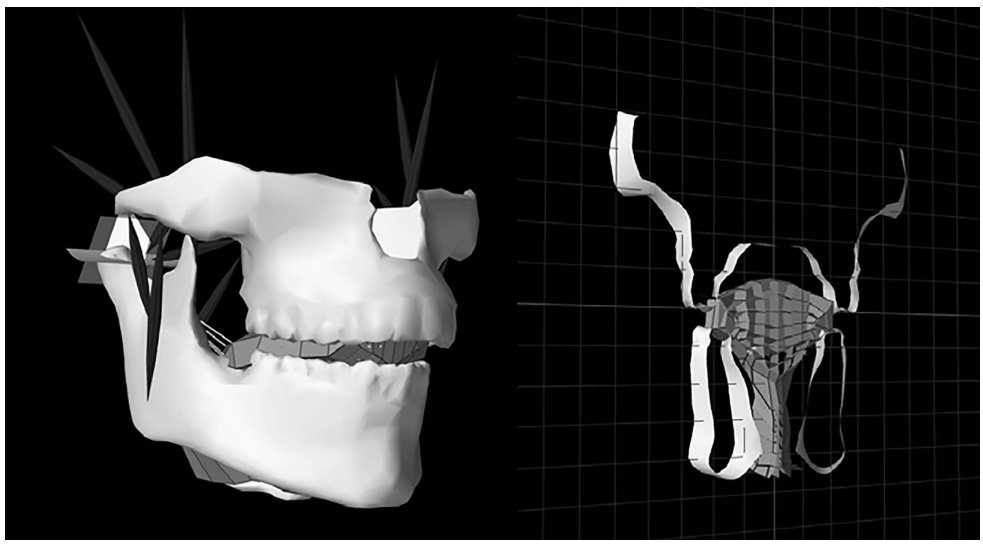

Figure 5.6 Computer simulation of oblique view of the tongue with surrounding skeletal structure (left) in the position to produce a coronal closure, with a coronal cross-section (right) showing the lateral tongue in contact with the molars.

Source: Images generated by I. Stavness using ArtiSynth; www.artisynth.org.

tongue's degrees of freedom and provides sensory feedback about the position of the tongue and, crucially, it forms the seal that keeps air from escaping into the cheeks, thus defining the aeroacoustic tube for speech. Because of its essential functions, with but few exceptions, bracing is maintained continuously throughout running speech (Gick et al., 2017).

Thus, for example, whenever the tongue tip moves upward to close for a /t/, the sides of the tongue are already braced against the teeth, holding in the air to allow pressure to build behind the tongue tip closure; likewise, in the case of a fricative like /s/, lateral bracing forms the central passageway through which air must travel, forcing the air forward through the narrow anterior passage to break against the teeth. This is not just true for coronal sounds, though, or for consonants: The tongue maintains this lateral bracing for vowels and consonants alike, suggesting a possible dedicated module for bracing that is kept constantly activated throughout speech, presumably as part of a language's pre-speech posture (Perkell, 1969) or articulatory setting (Gick et al., 2004). The notable exception to lateral bracing, of course, is the class of lateral sounds, such as English /l/, which are commonly associated with intentional lowering of the sides of the tongue. Acoustically, this lowering engages the side resonators of the buccal cavities, giving /l/ its characteristic sound. It is interesting to note that, when the tongue loses its lateral bracing during /l/, the anterior tongue normally makes contact with the alveolar ridge, providing anterior bracing for the tongue and ensuring that bracing at some location is maintained between the tongue and palate throughout running speech.

Bracing is an essential part of understanding tongue articulation, as it delineates the tongue's independently controllable functional regions (see Stone and Lundberg, 1996). For example, when the tongue is braced laterally, the tongue tip can function as a relatively simple valve. This simple valve is so efficient that the tongue tip is even capable of producing sequences of fast upward and downward motions in an oscillatory fashion, without active control; in speech, this dynamic can be seen in trilling of the tongue tip, as well as in sequential flaps in the word "Saturday" as produced by speakers of many dialects of English (Derrick et al., 2015). More posteriorly, as long as the lateral tongue remains braced, the mass of the tongue body cannot be raised and lowered as a single object by extrinsic muscles; rather, it must be raised and lowered through reshaping of the tongue, primarily through the action of muscles intrinsic to the tongue. Still farther back, the tongue dorsum interacts with the soft palate in complex ways to produce uvular sounds (such as the rhotic sounds in French or German); this complex interaction will be discussed in the following section on the soft palate. The tongue root likewise uses its own mechanisms to move anteroposteriorly (forwards and backwards) through combinations of intrinsic reshaping and as part of more general constriction maneuvers of the pharynx.

Considering the mechanics of the tongue in this light, its function looks less like that of a free-floating tentacle and more like a set of different – and tightly constrained – local mechanisms, some of which act like "valves" (Edmondson and Esling, 2006) or sphincters (Gick et al., 2013c). As with the lips, each such locality may be associated with a small number of dedicated neuromuscular modules, with the number depending on how many degrees of freedom (i.e., how many distinct configurations) need to be exploited at that location to realize a particular language's phonetic inventory.

Soft Palate

The soft palate (also referred to interchangeably as the velum) acts as the primary "gatekeeper" separating the oral, nasal, and pharyngeal cavities. Different parts of this highly

flexible structure are responsible for controlling the velopharyngeal port (VPP; the passageway between the nasal and pharyngeal cavities) and the oropharyngeal isthmus (OPI; the passageway between the oral and pharyngeal cavities). This division of cavities by the soft palate is illustrated in Figure 5.7. In speech, the soft palate is sometimes characterized as functioning like a "trapdoor," which can adopt one of two positions: raised/open, to produce nasal sounds, or lowered/closed, to produce oral sounds; it is thus involved in creating both the oral/nasal distinction (as in /d/ vs. /n/ or /a/ vs. /ã/) and in interfacing with the tongue to form uvular constrictions for sounds such as /q/, /ɴ/ or /ʀ/ (Fujimura and Lindqvist, 1971).

The trapdoor model, however, does not account for the dual functions of the soft palate/velum as a simultaneously nasal and oral articulator. That is, if this structure can only be raised or lowered trapdoor-style, how can it participate in synchronous constrictions of the OPI and VPP? The answer lies in the complex structure and function of the soft palate (shown in Figure 5.7): These two distinct functions (closing the VPP to create oral sounds and interacting with the tongue dorsum to make various uvular constrictions for sounds such as the English /w/ or the French /ʁ/) are realized using independently controllable parts of the soft palate. Specifically, X-ray research has shown that, while the upper portion of the soft palate holds the VPP closed, the lower portion (the "veil" or "traverse" that hangs down in the back of the mouth, terminating with the uvula) functions independently from the rest of the soft palate, bending toward the tongue to form the uvular constriction for French /ʁ/

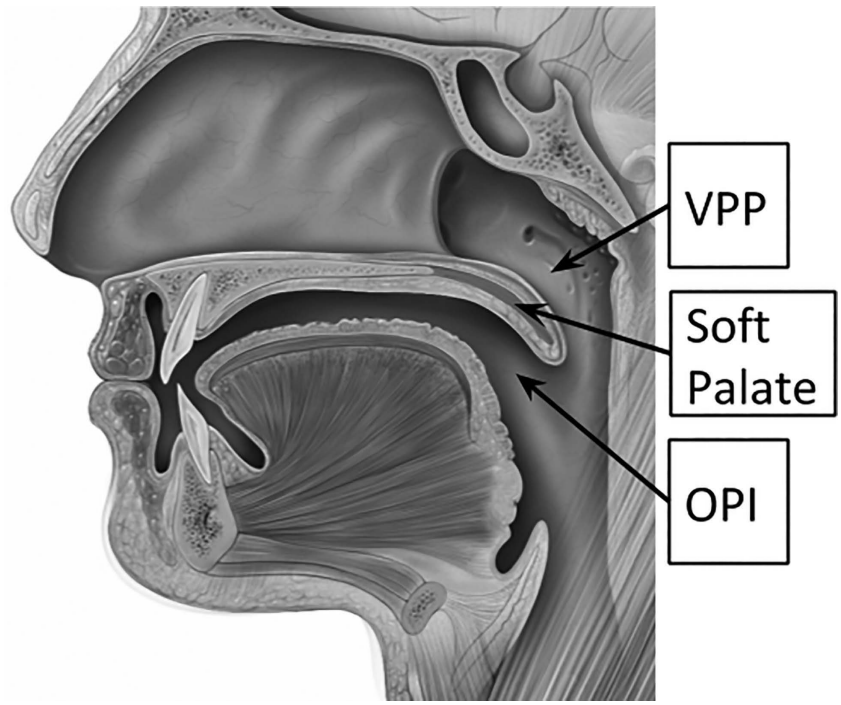

Figure 5.7 Midsagittal view of velopharyngeal port (VPP) and oropharyngeal isthmus (OPP) separated by the soft palate.

Source: Adapted from an image by Patrick J. Lynch, medical illustrator (CC BY 2.5 [http://creativecommons.org/licenses/by/2.5]), via Wikimedia Commons.

(Gick et al., 2014, 2013a). Considering its distinct structure and function – and the etymology of the term "velum" (Latin "veil") – it would make sense to reserve the term "velum" to refer exclusively to the veil-like structure that descends to interact with the tongue to form "velar" sounds.

An important lesson about articulatory phonetics that is perhaps most clearly demonstrated by the velum is that different speakers can use different – sometimes dramatically different – mechanisms for achieving similar outcomes. Well-known observations of this kind have been made about non-velar sounds, such as categorical variants of /r/ (e.g., Stavness et al., 2012) and /s/ (Dart, 1998). Similarly, the closure of the VPP has been shown to be achieved in one of four distinct ways (Biavati et al., 2009), as illustrated in Figure 5.8. Just over half of people are found to close the VPP with a relatively simple raising ("trapdoor") method where the velum raises mainly through constriction of the levator palati muscle, lifting the velum to the pharyngeal wall to close off the passageway; roughly 20% of people use not only the levator palati but also squeeze the VPP from the sides using the superior pharyngeal constrictor muscle; a further 15–20% of people add in constriction from the Passavant's Ridge in the pharyngeal wall to provide constriction from all four directions; finally, a small percentage of people close only laterally using the superior pharyngeal constrictor. Thus, distinct modules are not only important for describing different functions, but also for describing the various mechanisms individuals use to perform similar functions.

In the most widely attested VPP closure mechanism (Figure 5.8a), a distinctive "hump" is formed in the central posterior region of the velum, assisting in VPP closure. This hump is one example of how the soft palate functions more like a complex, tongue-like structure than like a simple trapdoor. Using computed tomography (CT) and magnetic resonance imaging (MRI) scans, Serrurier and Badin (2008) built a three-dimensional reconstruction of the soft palate that clearly shows the soft palate undergoing this intrinsic reshaping, apparently using intrinsic muscles to raise and retract the soft palate towards the rear pharyngeal wall. Subsequent simulation studies (Anderson et al., 2016) show that it is compression of the intrinsic portion of the levator veli palatini (LVP) muscle that hydrostatically produces this characteristic palatal hump (visible in Figure 5.9), similar to how the tongue raises to produce a palatal or dorsal constriction for sounds such as [j] or [k].

Our knowledge of the muscles involved in soft palate control has expanded considerably in recent years through the use of computer simulations. Simulations are an important tool for understanding a structure that is so difficult to observe in action using standard imaging techniques. Simulations such as the one shown in Figure 5.9 start with a detailed reconstruction of the three-dimensional geometry (shape) of bones, cartilage, ligaments,

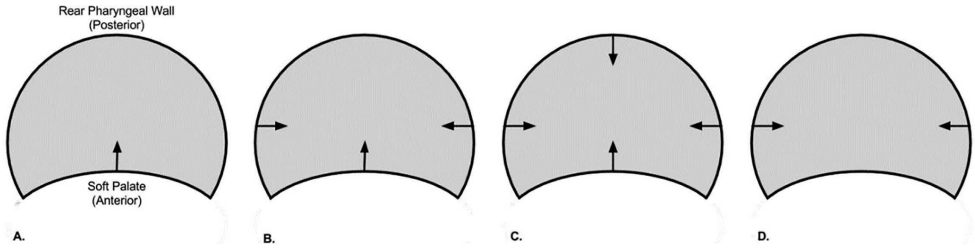

Figure 5.8 Variation in VPP closure: A. simple raising ("trapdoor") method; B. raising + lateral compression; C. raising + lateral compression + Passavant's Ridge; D. lateral compression only. Image by R. Taylor.

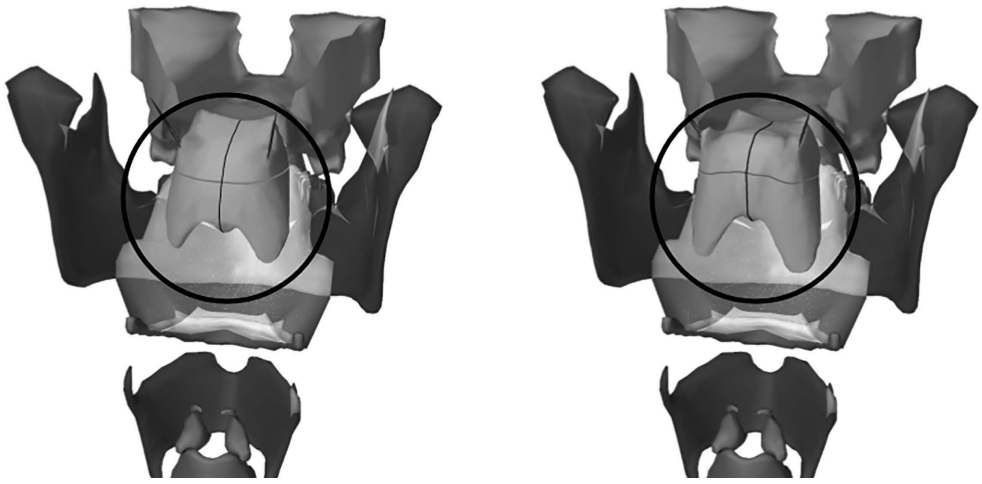

Figure 5.9 Computer simulation of posterior oblique views of soft palate, in relaxed (left) and raised/humped (right) position. The extrinsic portion of LVP muscles are shown as lines extending from the top corners of the soft palate. A horizontal gray intersect line roughly distinguishes the lower veil/velum portion from the rest of the soft palate, while a black line traces the mid-sagittal plane, highlighting the contour of the hump.

Source: Figure adapted by R.C. Taylor and Yadong Liu from images generated by P. Anderson using ArtiSynth (www.artisynth.org).

muscles, and other structures based on available medical image data from sources such as cryosection, CT and MRI; the resulting reconstructed structures are then broken into smaller elements that are assigned properties such as elasticity and stiffness based on known material properties. Simulations built from these structures can then be used to predict how, for example, contracting a particular muscle might modify the shape of that muscle and surrounding structures (see, e.g., Anderson et al., 2016; Gick et al., 2014). Computer simulations of the OPI have also shown that, as with the lips, different sizes and types of constrictions are produced using qualitatively different mechanisms, implying distinct neuromuscular modules. Gick et al.'s. (2014) simulation, for example, showed how the soft palate is an active contributor to making uvular sounds, with larger-area OPI constrictions (such as those for /u/ or /w/) produced using a mechanism depending mainly on the action of the palatoglossus muscle, while smaller-area constrictions (such as those for uvular fricatives) employ a strategy combining both palatoglossus and palatopharyngeus, which together form a "sling" effect that narrows the OPI as it pulls the velum toward the tongue.

Far from resembling a simple trapdoor, the soft palate/velum can be seen to exhibit the same qualitative properties as the tongue, with functionally independent substructures, essential speech and non-speech roles in multiple vocal tract cavities (oral, nasal, and pharyngeal), a high degree of speaker-specific variation, distinct functions for intrinsic and extrinsic musculature, hydrostatic control of shape, and modular neuromuscular organization. These properties enable the soft palate to participate in producing nearly every sound in the languages of the world.

Larynx

If there is any structure in the body that has a range and complexity of motion to rival the tongue, it is the larynx. While laryngeal phonetics has traditionally concentrated on the vocal folds, which act as the primary noise source for most speech vocalizations, a fuller picture of the larynx must include not just the vocal folds but also the epilarynx (which encompasses the epiglottis and the aryepiglottic folds; see Moisik, 2013). Different laryngeal states, or states of the glottis (Esling, 2006), have often been described as occurring on a continuum of glottal closure, from open to closed (Gordon and Ladefoged, 2001; Ladefoged, 1971):

(1) [open] voiceless – breathy – modal – creaky – glottal closure [closed]

The interpretation in (1) suggests that the larynx is controlled by a single "opening/closing" mechanism that produces different outputs by activating to different degrees. If this were correct, it would run counter to a modular approach. Note that, under this "single mechanism" view, the most widely attested type of phonation – modal voice – employs a potentially less stable intermediate state along a mechanical continuum of this kind (think of a light switch, with stable *up* and *down* "end state" positions at either end of its motion, but with any number of less stable positions in between).

Contrary to this view, laryngoscopic evidence (Esling, 1999) suggests that each state along the apparent continuum is produced by a quite distinct mechanism (Esling and Harris, 2005; Moisik and Esling, 2011), similar to those described earlier for the supralaryngeal articulators. These mechanisms may map to some degree onto the structures Edmondson and Esling (2006) term "valves." Moisik and Gick (2017) attribute these distinct laryngeal mechanisms to the actions of different muscle groupings, each of which is built to take advantage of a stable "sweet spot" in laryngeal biomechanics: By simulating the actions of a number of these structures, they find that each generates a different reliable outcome corresponding to a different speech sound, producing such varied output states as modal voice, creaky voice, glottal stop, and aryepiglotto-epiglottal fricative. Thus, each different degree of laryngeal aperture is the output of a distinct module rather than of gradual fine adjustments in aperture, so that every laryngeal state may be viewed (both in terms of description and potential stability) as an "end state" along its own continuum. In this way, "glottal" sounds are described here in the same terms as any other laryngeal and supralaryngeal sounds. To give an example, Moisik and Gick (2017) show that the ventricular folds are involved in producing a reinforced version of a glottal stop. Rather than relying simply on the true vocal folds for the closure, the ventricular folds act as a mass that supports the closure and dampens the vibration of the true folds (like a hand resting on a handbell to stop its ringing). This view of the larynx is analogous to our earlier description of the labial sounds, which result from discrete articulatory configurations as opposed to degrees along a continuum. The physical stability of each of these states is what allows speakers to accurately produce a specific articulatory configuration time and again in rapid speech.

The larynx is unlike the other structures we have observed in that it provides an exception to this "mechanical endpoint" control type when it is involved in one particular activity: inspiration. Moisik and Gick (2017) find that, unlike the vocal fold abduction that is used for speech sounds such as /h/ and plosive aspiration, the wider abduction used for inspiration operates in a variable, scalar way. That is, the more muscle activation that is used to open the vocal folds, the wider they open, in a more or less linear relationship. This flexibility in

degree of glottal opening for inspiration allows the body to take in exactly as much oxygen as needed from one breath to another – an essential property for survival. The only specifically speech-related function that appears to possibly share this property of variable control is pitch (Moisik et al., 2017), though further research is needed to determine the mechanisms used by populations adept at producing a specific pitch target, such as professional singers and speakers of tone languages.

Putting articulations together

The previous sections have focused on identifying the basic structures and movements that make up speech. However, describing individual movements is of course insufficient for a model of how speech works, as even the simplest utterances combine strings of multiple overlapping actions. Coarticulation – how overlapping speech movements interact with one another – must thus be a central part of a complete model of speech.

Joos (1948) proposed an early embodied model of coarticulation in which he described local interactions between speech movements as the outputs of neurally mediated combinations of overlapping muscle activations. Joos argued that speech movements can be thought of as basically additive, but that the brain is needed to mediate such interactions to handle the complexity of coarticulation. The even-more-basic idea of unmediated additive muscle activations in motor control has since gained currency in the neurophysiology literature with Bizzi et al.'s. (1991) discovery of "superposition" of muscle activations in spinalized frogs. They show that overlapping movements can combine through simple addition of muscle activations, unmediated by higher structures in the nervous system. In this way, activations combine in different proportions to produce a large variety of movements. Subsequent work on humans has shown that, thus combined, only a half-dozen or so superposed modules are needed to describe a wide range of actions of the arm (D'Avella et al., 2006), the hand (Overduin, 2012), or even complex whole-body postures (Torres-Oviedo and Ting, 2007). This same mechanism has been proposed as a basis for combining speech movements in coarticulation (Gick and Stavness, 2013).

Some limited work has been done toward testing whether superposition is a plausible way to model coarticulation. Gick and colleagues (2013b) used biomechanical modeling to test superposition through simulations of coarticulatory interactions in VCV sequences. Comparing these simulation results with electromagnetic articulometry (EMA) results reported by Recasens and Espinosa (2009) resulted in characteristic patterns of coarticulation, obtained simply by temporally overlapping activations for canonical consonants and vowels. On the surface, these coarticulatory patterns can appear very complex: Not only do sounds coarticulate with one another differently under different conditions (at varying speech rates, for example), but each segment also seems to follow its own rules of coarticulation, with some segments described as more "resistant" to coarticulatory effects, or more "aggressive" in exerting their coarticulatory effects on nearby segments (e.g., Bladon and Al-Bamerni, 1976; Fowler and Brancazio, 2000).

It has often been observed that these two properties of coarticulation – resistance and aggressiveness – are positively correlated (e.g., Farnetani, 1990; Fowler and Saltzman, 1993; Recasens et al., 1997; Recasens and Espinosa, 2009). In biomechanical terms, this correlation is not surprising: Both "resistance" (i.e., the extent to which a movement may, or may not, be perturbed in response to another movement) and "aggressiveness" (i.e., the extent to which a movement perturbs other movements) are direct functions of the stiffness associated with each of the overlapping movements – and variable stiffness in the body is

a function of muscle activation. Thus, a movement that employs more muscle activation – particularly intrinsic muscle activation – causes an increase in stiffness, which will effect an increase in both "resistance" and "aggressiveness" in coarticulation. Recasens and Espinosa (2009) observe that palatal segments exhibit both increased resistance and increased aggressiveness in their coarticulatory relations. Because of the absence of extrinsic muscles in the anterior mouth, palatal sounds have often been associated with comparatively high levels of intrinsic muscle activation (e.g., Stavness et al., 2012). While this intrinsic stiffness possibility remains to be more fully tested, an embodied approach offers a testable path forward towards a deeper understanding of how speech sounds work together.

An embodied approach to coarticulation treats instances of superposition the same, irrespective of the timescale of overlap. Thus, as muscle activations can be maintained over long periods of time (as in the tonic activations used in controlling body posture), we can construe any case of basic coarticulation as superposition, whether it appears to occur locally (as with immediately adjacent speech sounds), non-locally (as with the long-distance interactions observed in harmony systems), or globally (as with articulatory settings, where a setting may affect every sound in a language, see Gick et al., 2004). Of course, the extent to which any instance of coarticulation may be seen as "basic" is not known, but as technology has made these proposals testable, future work will enable researchers to uncover the extent to which additional factors come into play in determining the patterns of coarticulation in speech. It is likewise possible to construe other linguistic structures such as the syllable as being governed by the same basic principles of overlapping elements, as with other cases of coarticulation, provided these principles are accompanied by a theory specifying the relative timing of elements. Superposition with non-speech actions may also be captured using this approach (as with talking while chewing or smiling), providing a unified model for interactions both within and between speech and non-speech movements. This approach suggests that the built-in mechanics of the human body can go a long way towards handling local coarticulatory interactions in ways that may look quite complex on the surface, without reference to advance planning, contextual information or extrinsic models – indeed, with no specified model of coarticulation at all.

Conclusion

While there is a perennial appeal to looking for answers to questions about speech and language in the brain, every speech sound we produce is necessarily the result of the moving body. Humans are able to produce many of our most complex vocal tract behaviors (e.g., swallowing, vocalizing, suckling, breathing) at birth without experience beyond the womb, and indeed without a brain above the brainstem, as evidenced by studies of anencephalic newborns (Radford et al., 2019). Such observations reveal in vivid terms the degree to which the biomechanical and neural structures needed for complex vocal tract action appear to be built into the body, as is the case with other functions that are present from birth, such as locomotion (Dominici et al., 2011). It has long been argued that embodied structures of this kind are the basis not just of reflexive movement, but of all volitional movement – that is, that these are the only controllable discrete "body parts" our nervous systems can employ (e.g., Easton, 1972). Casting speech behavior in terms of these body-based structures offers researchers new ways of approaching some of the defining problems of articulatory phonetics (Gick, 2016; Gick and Stavness, 2013).

An apt model for speech sounds in such an embodied approach is not that we learn to control an inventory of sounds per se, but rather that we learn to build and control an

inventory of highly specialized body parts, each of which is constructed and optimized to serve a specific phonetic function. Physical articulators thus cannot be divorced from the tasks they perform, combining both representation and action into primitives of the speech system. Identifying the movement primitives that make up this system has been a central enterprise in articulatory phonetics, where generations of phoneticians, phonologists, and speech researchers have described and cataloged in detail the minimal elements of the speech motor system; this tradition contrasts starkly with that of movement research in other areas, such as locomotion and posture, where primitives have been studied much less frequently, systematically, and comprehensively. This rich tradition in phonetics has grown out of the long-observed fact that humans have harnessed these physically stable movement primitives to form the basis of nature's most complex communication system. The field of articulatory phonetics thus deeply underlies theories of phonetics and phonology, speech evolution and acquisition, and sound change.

Acknowledgments

The authors wish to thank the many collaborators who have contributed to the studies and simulations that have fed into this paper. This work has been supported by NSERC Discovery grant RGPIN-2015-05099 to the first author, and by National Institutes of Health grant DC-002717 to Haskins Laboratories.

Note

1 Note that muscles that control the jaw (e.g., temporalis, masseter) are included here as these muscles are important in determining lip position. As the role of the jaw as a primary articulator in adult speech is controversial (see, e.g., Redford and van Donkelaar, 2008), the jaw will not be addressed independently in the present work.

References

Anderson, P., Fels, S., Stavness, I., Gick, B., 2016. Intrinsic and extrinsic portions of soft palate muscles in velopharyngeal and oropharyngeal constriction: A 3D modeling study. *Canadian Acoustics*. 44, 18–19.

Anderson, P., Harandi, N.M., Moisik, S.R., Stavness, I., Fels, S., 2015. A comprehensive 3D biomechanically-driven vocal tract model including inverse dynamics for speech research, in: *Proceedings of Interspeech 2015: 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, pp. 2395–2399.

Biavati, M.J., Sie, K., Wiet, G.J., Rocha-Worley, G., 2009. Velopharyngeal insufficiency. *Emedicine*: *Otolaryngology Facial Plastic Surgery*. pp. 1–21.

Bizzi, E., Mussa-Ivaldi, F.A., Giszter, S., 1991. Computations underlying the execution of movement: A biological perspective. *Science*. 253, 287–291.

Bladon, R.A.W., Al-Bamerni, A., 1976. Coarticulation resistance in English /l/. *Journal of Phonetics*. 4, 137–150.

Buchaillard, S., Perrier, P., Payan, Y. 2009. A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning. *The Journal of the Acoustical Society of America*. 126, 2033–2051.

Burenhult, N., 2006. Body part terms in Jahai. *Language Science*. 28, 162–180.

Cooper, F.S., Liberman, A.M., Harris, K.S., Grubb, P.M., 1958. Some input-output relations observed in experiments on the perception of speech. *Proceedings of the 2nd International Congress on Cybernetics*. Namur, Belgium: International Association for Cybernetics. pp. 930–941.

Dart, S.N., 1998. Comparing French and English coronal consonant articulation. *Journal of Phonetics*. 26, 71–94.

D'Avella, A., Bizzi, E., 2005. Shared and specific muscle synergies in natural motor behaviors. *Proceedings of the National Academy of Science U. S. A*. 102, 3076–3081.

D'Avella, A., Portone, A., Fernandez, L., Lacquaniti, F., 2006. Control of fast-reaching movements by muscle synergy combinations. *Journal of Neuroscience*. 26, 7791–7810.

Derrick, D., Stavness, I., Gick, B., 2015. Three speech sounds, one motor action: Evidence for speech-motor disparity from English flap production. *The Journal of the Acoustical Society of America*. 137, 1493–1502. https://doi.org/10.1121/1.4906831

Dominici, N., Ivanenko, Y.P., Cappellini, R.E., G. d'Avella, A., Mondi, V., Cicchese, M., Fabiano, A., Silei, T., Di Paolo, A., Giannini, C., Poppele, R.E., Lacquaniti, F., 2011. Locomotor primitives in newborn babies and their development. *Science*. 334, 997–999.

Easton, T.A., 1972. On the normal use of reflexes. *American Scientist*. 60, 591–599.

Edmondson, J.A., Esling, J.H., 2006. The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology*. 23, 157–191.

Enfield, N.J., Majid, A., Van Staden, M., 2006. Cross-linguistic categorisation of the body: Introduction. *Language Sciences*. 28, 137–147.

Esling, J.H., 1999. The IPA categories,"Pharyngeal" and "Epiglottal" Laryngoscopic observations of Pharyngeal articulations and Larynx height. *Language and Speech*. 42, 349–372.

Esling, J.H., 2006. States of the glottis, in: Brown, K. (Ed.), *Encyclopedia of Language and Linguistics*. Oxford: Elsevier. pp. 129–132.

Esling, J.H., Harris, J.G., 2005. States of the glottis: An articulatory phonetic model based on laryngoscopic observations, in: Hardcastle, W.J., Beck, J. (Eds.), *A Figure of Speech: A Festschrift for John Laver*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 347–383.

Farnetani, E., 1990. VCV lingual coarticulation and its spatiotemporal domain, in: Netherlands, S. (Ed.), *Speech Production and Speech Modelling*, Dordrecht: Kluwer Academic Publishers, pp. 93–130.

Fowler, C.A., Brancazio, L., 2000. Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech*. 43, 1–41.

Fowler, C.A., Saltzman, E., 1993. Coordination and coarticulation in speech production. *Language and Speech*. 36, 171–195.

Fujimura, O. 1989. Comments on "On the quantal nature of speech," by K. N. Stevens. *Journal of Phonetics*. 17, 87–90.

Fujimura, O., Lindqvist, J., 1971. Sweep-tone measurements of vocal-tract characteristics. *The Journal of the Acoustical Society of America*. 49, 541–558. https://doi.org/10.1121/1.1912385

Gick, B., 2016. Ecologizing dimensionality: Prospects for a modular theory of speech production. *Ecological Psychology*. 28, 176–181. https://doi.org/10.1080/10407413.2016.1195195

Gick, B., Allen, B., Roewer-Despres, F., Stavness, I., 2017. Speaking tongues are actively braced. *Journal of Speech, Language and Hearing Research*. 60, 494–506. https://doi.org/10.1044/2016_JSLHR-S-15-0141

Gick, B., Anderson, P., Chen, H., Chiu, C., Kwon, H.B., Stavness, I., Tsou, L., Fels, S., 2014. Speech function of the oropharyngeal isthmus: A modelling study. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. 2, 217–222.

Gick, B., Francis, N., Klenin, A., Mizrahi, E., Tom, D., 2013a. The velic traverse: An independent oral articulator? *The Journal of the Acoustical Society of America*. 133, EL208–EL213.

Gick, B., Stavness, I., 2013. Modularizing speech. *Frontiers in Psychology*. 4, 977.

Gick, B., Stavness, I., Chiu, C., 2013b. Coarticulation in a whole event model of speech production, in: *The Journal of the Acoustical Society of America – Proceedings of Meetings on Acoustics ICA2013*. ASA, 060207–060211.

Gick, B., Stavness, I., Chiu, C., Fels, S., 2011. Categorical variation in lip posture is determined by quantal biomechanical-articulatory relations, *Canadian Acoustics*. 39(3) pp. 178–179.

Gick, B., Wilson, I., Derrick, D., 2013c. *Articulatory Phonetics*. Oxford: John Wiley & Sons.

Gick, B., Wilson, I., Koch, K., Cook, C., 2004. Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica* 61, 220–233.
Gordon, M., Ladefoged, P., 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics*. 29, 383–406.
Gray, H., Lewis, W., 1918. *Anatomy of the Human Body*. New York, NY: Lea & Febiger.
Hagedorn, C., Lammert, A., Bassily, M., Zu, Y., Sinha, U., Goldstein, L., Narayanan, S.S., 2014. Characterizing post-glossectomy speech using real-time MRI, in: *Proceedings of the International Seminar on Speech Production*. Cologne.
Joos, M., 1948. Acoustic phonetics. *Language* 24, 5–136.
Kier, W.M., Smith, A.M., 2002. The structure and adhesive mechanism of octopus suckers. *Integrative and Comparative Biology*. 42, 1146–1153.
Ladefoged, P., 1971. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.
Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. *Psychological Review*. 74, 431–461.
Lightoller, G.H.S., 1925. Facial muscles: The Modiolus and Muscles surrounding the Rima Oris with some remarks about the Panniculus Adiposus. *Journal of Anatomy*. 60, 1–85.
Loeb, E.P., Giszter, S.F., Saltiel, P., Bizzi, E., Mussa-Ivaldi, F.A., 2000. Output units of motor behavior: An experimental and modeling study. *Journal of Cognitive Neuroscience*. 12, 78–97.
Moisik, S., 2013. *The Epilarynx in Speech*. Doctoral dissertation, University of Victoria.
Moisik, S.R., Esling, J.H., 2011. The "Whole Larynx" approach to laryngeal features, in: *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 1406–1409.
Moisik, S., Gick, B., 2017. The quantal larynx: The stable regions of laryngeal biomechanics and implications for speech production. *Journal of Speech, Language, and Hearing Research*, 60(3), 540–560. https://doi.org/10.1044/2016_JSLHR-S-16-0019
Moisik, S., Gick, B., Esling, J.H., 2017. The quantal larynx in action: Smooth and abrupt aspects of laryngeal motion observed in laryngoscopic videos, in: *Proceedings of the 11th International Seminar on Speech Production*. Tianjin, China, pp. 95–96.
Nazari, M.A., Perrier, P., Chabanas, M., Payan, Y., 2011. Shaping by stiffening: A modeling study for lips. *Motor Control* 15, 141–168.
Overduin, E., d'Avella, S.A., Carmena, A., Bizzi, J.M., 2012. Microstimulation activates a handful of muscle synergies. *Neuron* 76, 1071–1077.
Perkell, J.S., 1969. *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Cambridge, MA: MIT Press.
Radford, K., Taylor, R., Hall, J., Gick, B., 2019. Aerodigestive and communicative behaviours in anencephalic and hydranencephalic infants. *Birth Defects Research* 111(2). 41–52.
Recasens, D., Espinosa, A., 2009. Acoustics and perception of velar softening for unaspirated stops. *Journal of Phonetics*. 37, 189–211.
Recasens, D., Pallarès, M.D., Fontdevila, J., 1997. A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America*. 102, 544–561.
Redford, M.A., van Donkelaar, P., 2008. Jaw cycles and linguistic syllables in adult English, in: Davis, B.L., Zajdo, K. (Eds.), *The Syllable in Speech Production: Perspectives on the Frame/Content Theory*. London: Taylor & Francis, pp. 355–376.
Russell, G.O., 1933. First preliminary X-ray consonant study. *The Journal of the Acoustical Society of America*. 5, 247–251.
Safavynia, S.A., Ting, L.H., 2012. Task-level feedback can explain temporal recruitment of spatially fixed muscle synergies throughout postural perturbations. *Journal of Neurophysiology*. 107, 159–177.
Schwartz, J-L., Boë, L-J., Vallée, N., Abry, C. 1997. The dispersion-focalization theory of vowel systems. *Journal of Phonetics*. 25, 255–286.
Serrurier, A., Badin, P., 2008. A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data. *The Journal of the Acoustical Society of America*. 123, 2335–2355.

Serrurier, A., Badin, P., Barney, A., Boë, L-J., Savariaux, C., 2012. The tongue in speech and feeding: Comparative articulatory modeling. *Journal of Phonetics*. 40, 745–763.

Stavness, I., Gick, B., Derrick, D., Fels, S., 2012. Biomechanical modeling of English /r/ variants. *The Journal of the Acoustical Society of America*. 131, EL355–EL360.

Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics*. 17, 3–45.

Stone, M., Lundberg, A., 1996. Three-dimensional tongue surface shapes of English consonants and vowels. *The Journal of the Acoustical Society of America*. 99, 3728–3737.

Sumbre, G., Fiorito, G., Flash, T., Hochner, B., 2006. Octopuses use a human-like strategy to control precise point-to-point arm movements. *Current Biology*. 16, 767–772.

Ting, L.H., Chiel, H.J., 2015. Chapter 12: Muscle, biomechanics, and implications for neural control, in: Hooper, S.L., Büschges, A. (Eds.), *The Neurobiology of Motor Control: Fundamental Concepts and New Directions*. New York, NY: Wiley.

Torres-Oviedo, G., Ting, L.H., 2007. Muscle synergies characterizing human postural responses. *Journal of Neurophysiology* 98, 2144–2156. https://doi.org/10.1152/jn.01360.2006

Verhoeven, J., 2005. Belgian standard Dutch. *Journal of the International Phonetic Association*. 35, 243–247.

Wierzbicka, A., 2007. Bodies and their parts: An NSM approach to semantic typology. *Language Sciences*. 29, 14–65.

Wilson, F.R., 1998. *The Hand: How Its Use Shapes the Brain, Language, and Human Culture*. New York, NY: Pantheon Books.