DOES SWALLOWING BOOTSTRAP SPEECH LEARNING?

Connor Mayer[1], François Roewer-Despres[2], Ian Stavness[2], and Bryan Gick[1,3]
[1]Department of Linguistics, University of British Columbia
[2]Department of Computer Science, University of Saskatchewan
[3]Haskins Laboratories, New Haven, CT

1 Introduction

Although speech movements must be learned, many of our most complex oral behaviours, such as swallowing, suckling, vocalizing, and breathing, can be produced at birth. This indicates a degree to which the biomechanical and neural structures needed for complex action appear to be built in. It has been hypothesized that phylogenetically-encoded structures such as these are used to bootstrap speech learning [1][2]: complex, innate movements may be broken down into constituent submodules that are then recruited for use in speech. This is consistent with developmental studies that have found that features of speech movements in infants are qualitatively similar to suckling movements, but these patterns are subsequently refined [3].

The learning of speech movements can be modeled as a search of a high-dimensional muscle activation space for sets of activations that satisfy task-specific criteria relevant to the speech learner. Even when considering a single speech movement in isolation, the dimensionality and size of the search space are large enough to provide significantly problems for an unstructured search: the number of sets of activations that result in a solution for a given task has been shown to be very small relative to the number of possible sets of activations [4], and the number of redundant sets of activations that are solutions for a given task makes predicting muscle activation difficult [5].

In this paper we explore these ideas using the 3D biomechanical modelling platform Artisynth (www.artisynth.org; e.g., [6][7]), specifically in the context of tongue bracing. The sides of the tongue are in contact with the upper molars during the vast majority of speech, and this bracing of the tongue against the molars requires dedicated muscle activation [4]. We compare the sets of neuromuscular activations that result in tongue bracing with the sets that result in the full oral closure stage of a swallow, which has the tongue pressed against the roof of the mouth as well as the upper molars. Although there is no guarantee that in such a high-dimensional, non-linear activation space the set of activations resulting in the full oral swallowing closure will overlap with those that result in tongue bracing, our hypothesis generates the prediction that these activation sets will overlap, providing a plausible starting point for learners' searches of the activation space.

2 Methods

We used muscle-driven tongue simulations to examine the muscle effort required to establish various types of tongue-palate contact. All possible sets of muscle activations in the tongue model were generated at three activation levels (0%, 20%, and 50%) for a group of ten muscles: superior longitudinal (SL), inferior longitudinal (IL), transverse (TRANS), verticalis (VERT), hyoglossus (HG), mylohyoid (MH), styloglossus (STY), and the posterior, medial, and anterior genioglossus (GGP, GGM, and GGA respectively). This generated 3^{10}, or approximately 60,000, activations. Virtual contact sensors were placed on the hard palate and upper teeth of the model to detect contact between the tongue and these surfaces. This allowed us to partition the activation space into four different types of contact. Bilateral contact is defined as lateral contact on both sides of the palate. Anterior contact is contact in the anterior region of the palate. Anterior-bilateral contact is bilateral contact as well as anterior contact: this is a representative example of tongue bracing during speech. Swallowing contact is bilateral, back, and mid contact, representing the end of the oral phase of swallowing, immediately after the tongue has moved the bolus into the pharynx. A more detailed methodological description of the simulations and sensor regions can be found in section 4.1 of [4].

3 Results

Of the 60,000 different combinations of muscle activations, 1000 resulted in bilateral contact, 247 in anterior contact, 81 in anterior-bilateral contact, 11 in swallowing contact, and 57,829 in other types of contact or no contact. Only about 2% of the total activations matched any of our defined contact types.

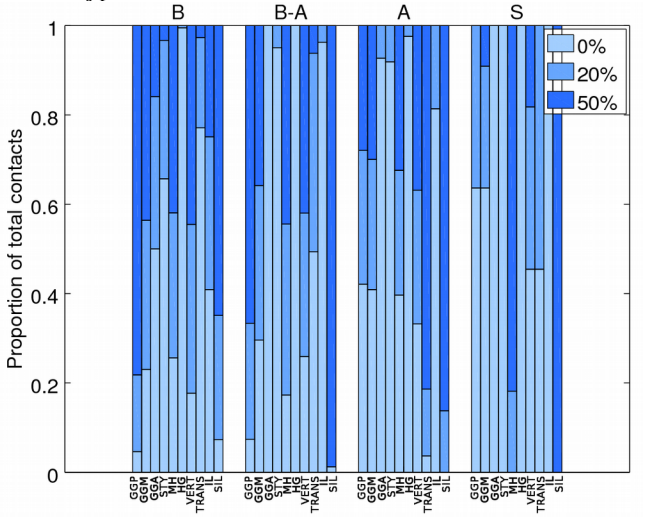

Figure 1: The proportion of the total number of each contact type (B: bilateral, B-A: bilateral-anterior, A: anterior, S: swallowing) with each muscle at 20% and 50% activation.

Figure 1 shows the distribution of excitation levels for each muscle in each type of contact. The swallowing contacts are the simplest, consisting primarily of contributions from the SL and MH. This is broadly consistent with what is known about muscle use in the oral phase of swallowing: the MH plays an important role in raising the tongue body, and the SL further elevates the surface of the tongue to make contact with the hard palate. The activations for the bilateral and anterior-bilateral conditions still show a prominent role for these two muscles, but include many other activations necessitated by the more complex tongue shapes required. t-Distributed Stochastic Neighbor Embedding (t-SNE) [8] was used to visualize the distributions of the different contact types in activation space in Figure 2. t-SNE is an iterative algorithm that maps from high-dimensional to low-dimensional space using an optimization function that prioritizes maintaining the Euclidean distance between each point and its neighbours. As a result, this visualization captures the relationships between the points in the original ten-dimensional space. The many clusters in Figure 2 are the result of our sampling methodology: because there were only three activation levels used for each muscle, it resulted in an uneven distribution in the activation space, and this is reflected in the figure.

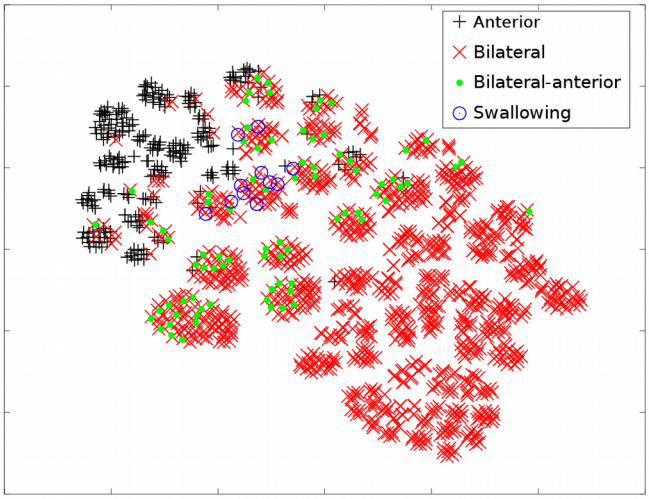

Figure 2: A two-dimensional t-SNE plot of the activation space.

There is a fairly clear boundary between contact types without bilateral contact (anterior) and contact types with bilateral contact (bilateral, anterior-bilateral, and swallow). There are a large number of redundant solutions for bilateral and, to a lesser degree, anterior-bilateral and swallowing contact, mirroring the redundancy typically observed in biomechanical simulations [5]. However, each cluster of activations resulting in swallowing contact is contiguous with clusters of activations resulting in anterior-bilateral and bilateral contacts, indicating that they share similar activations.

4 Discussion

The muscle activations found in the swallowing contacts are a subset of the activations found in the anterior-bilateral and the bilateral conditions: the SL and MH play the most significant role in swallowing contacts, and they play significant roles in anterior-bilateral and bilateral contact as well, but with additional muscle activations needed to produce the more complex tongue shapes for these types of contact, such as depressing the mid-sagittal region of the tongue and raising the tip. This is consistent with other biomechanical studies that have found that simpler, innate movements are subsequently refined to produce the more complex movements used in lip and jaw control in speech [5] and locomotion [9], and indicates that movements used in swallowing are suitable candidates for bootstrapping the learning of subtler speech movements such as bilateral tongue bracing.

The distribution of the swallowing contacts in the activation space also supports this idea. Even within the limited activation space defined by three activation levels per muscle, only a small number of activations result in any of the contact types we are investigating (about 2% of activations). An unstructured search of the full activation space would thus prove costly and inefficient. Although it is impossible to speculate on which activations are favoured by speakers based on this simulation alone, it is telling that each activation resulting in swallowing contact is contiguous with a cluster of bilateral and anterior-bilateral contact activations, and that the cluster containing the majority of swallowing contacts contains no instances of anterior contact. This indicates that swallowing activations could be a useful starting point for learning bilateral and anterior-bilateral speech movements by acting as a heuristic starting point for a search of the activation space.

Acknowledgments

We acknowledge funding from NSERC Discovery Grants to the third and fourth authors.

References

[1] Peter MacNeilage. (2008). The Origin of Speech. Oxford.

[2] M. Studdart-Kennedy and L. Goldstein (2003). Launching language: The gestural origin of discrete infinity. In M. Christiansen and S. Kirby (eds.), *Language Evolution*. Oxford.

[3] J.R. Green, C.A. Moore, M. Higashikawa, and R.W. Steeve. (2000). The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination. *J. Sp., Lang. & Hear. Res.*, 43:239.

[4] B. Gick, B. Allen, F. Roewer-Despres, and I. Stavness (in pr.). Speaking tongues are actively braced. *J. Sp., Lang. Hear. Res.*

[5] Gerald E. Loeb. (2012). Optimal isn't good enough. *Biological Cybernetics*, 106:757.

[6] I. Stavness, J.E. Lloyd, and S.S. Fels (2012). Automatic Prediction of Tongue Muscle Activations Using a Finite Element Model. *J. Biomechanics*, 45:2841.

[7] B. Gick, P. Anderson, H. Chen, C. Chiu, H.B. Kwon, I. Stavness, L. Tsou, and S. Fels (2014). Speech function of the oropharyngeal isthmus: A modeling study. *Compu. Meth. Biomech. & Biomed. Eng.: Imaging & Visualization*, 2:217.

[8] L.J.P. van der Maaten and G.E. Hinton (2008). Visualizing High Dimensional Data Using t-SNE. *J. Mach. Lear. Res.*, 9:2570.

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