Visualising phonetics at the neuromuscular level reveals the discrete pulsatile gated nature of speech

Please write XXX instead of the name(s) of the author(s)

Please write XXX instead of the affiliation(s)
please write XXX instead of the email address(es)

ABSTRACT

Glossometry, lip-video, photoglottography and a new understanding of the compartmental nature of lingual neuromuscular structure, provide the opportunity to study speech production from a motor control perspective. External photoglottography is used to visualise glottal opening/closing; pose-estimation of profile lip-video to visualise lip protrusion and closure; and pose-estimation of ultrasound to visualise the degree of doming/grooving of five independently controlled sectors of the tongue body and an independently controlled tongue blade/tip. A short utterance is analysed to demonstrate the potential of these new tools. Motor commands to agonist/antagonist teams of neuromuscular compartments appear to be initiated at regular discrete intervals of approximately 100ms. The velocity of each compartment contraction appears to be planned such that each phone, comprising a set of motor commands, is timed to reach its target in one or two 100ms periods. Thalamic or oliviocerebellar pulsing neurons are suggested as a mechanism for gating and synchronising motor commands.

Keywords: ultrasound, timing, photoglottography, glossometry, motor control

1. INTRODUCTION

This paper proposes a set of lingual measures which, it is claimed, represent underlying motor commands. In combination with a measure of glottal opening and a measure of lip movement, patterns of movement can be observed. Most apparent is the synchronicity and regularity of the motor commands. This in turn leads to the radical proposal that speech consists of sets of motor commands issued sequentially at discrete intervals regulated by a central “clock”.

2. THE COMPARTMENTAL TONGUE

A recent review of the anatomical structure of the tongue [14] indicates the genioglossus muscle may be divided into five independently controlled neuromuscular compartments (Fig 1) which work along with compartments of transversus and verticalis muscles to form five synergistic agonist/antagonist units that control the tongue body. The hypothesis proposes that antagonistic contraction of genioglossus and transversus neuromuscular compartments controls the amount of grooving or doming of each sector independently. This hypothesis is supported by the measurement of distances from EMA-type keypoints on the tongue surface to the origin of the genioglossus muscle which attaches to the mandible via a short tendon. These five points show independent coronal doming or grooving for different vowels and consonants.

A little-known subcompartment of the inferior longitudinal muscle (Oblique longitudinal) arises from the tongue root and extends to the anterior of the tongue body. It is proposed that its purpose is to constrain the rostral/caudal expansion of the genioglossus, thus enhancing and stiffening the bunching of the tongue body.

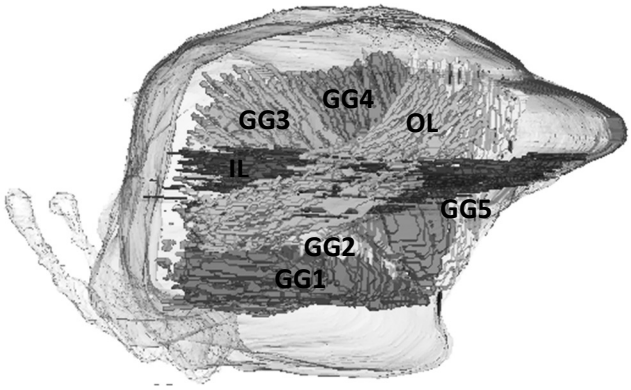

Figure 1: Sagittal view of tongue with 5 independently innervated sectors of genioglossus GG1-5; OL - Oblique longitudinal; IL - inferior longitudinal)

3. METHOD: INSTRUMENTATION AND ANALYSIS

3.1. Ultrasound with DeepLabCut pose estimation

DeepLabCut pose estimation was trained on approximately 2000 hand-labelled ultrasound images of speech and water swallows from five different ultrasound systems, seven different probe geometries, child and adult speakers. 14 keypoints are labelled. 11 points on the tongue, one on hyoid,

one on base of mandible (inferior tubercle) and one on the short tendon (superior tubercle).

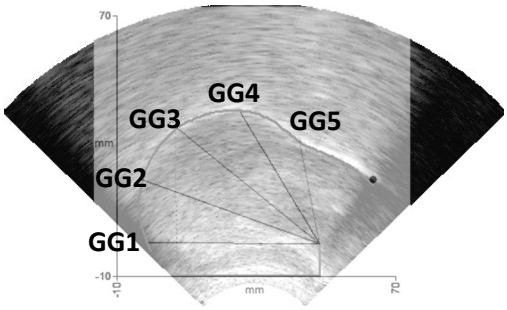

Figure 2: DeepLabCut pose estimation of tongue (11 keypoints) and short tendon with distances GG1-GG5 superimposed

From root to tip the first seven keypoints (Fig 3.) indicate the tongue body and the last four indicate the blade/tip. For the purpose of this study, the tongue body contour from keypoint 1-7 is then interpolated along its length to extract five equally spaced points to represent the five neuromuscular compartments and labelled GG1 (root)-GG5(anterior tongue body) (see Fig 2). Distances from these five points to the origin of the short tendon on the superior tubercle of the mandible are referred to herein as glossometric analysis.

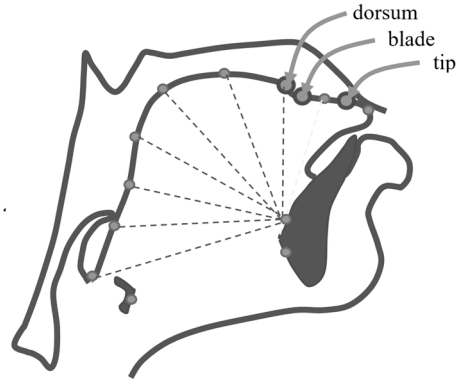

Figure 3: Sagittal diagram of the tongue with 11 pose estimation keypoints marked on the tongue contour, one on the hyoid one on each of the superior and inferior tubercles of the mandible. Three highlighted points are selected for correlation comparison with co-registered EMA data.

Validity of the pose-estimated lingual keypoints was evaluated on a co-registered dataset [8] with electromagnetic articulograph (EMA) dorsum, blade and tip sensors and synchronous ultrasound images (81Hz). Both systems recorded data in mm and the pose estimated data was rotated and translated to match the EMA data. The EMA sensors were correlated with the nearest pose-estimated keypoints (Fig 3.) for around 700 ultrasound frames taken from three phonetically balanced sentences from a single speaker (unseen in the pose-estimation training set). Table 1. Shows the x and y Pearson correlation scores. Y-correlations are high (>0.80)1. X-correlations are lower due to challenges in accurately training the tip position which is not always visible in the ultrasound image.

	tip	blade	dorsum
x	0.46	0.59	0.67
y	0.81	0.88	0.91

Table 1: Pearson correlation scores for tongue tip, blade and dorsum EMA sensors vs closest pose-estimated keypoints derived from co-registered ultrasound images.

3.2 Lip video with DeepLabCut pose estimation

Lip camera images we trained using DeepLabcut. Seven keypoints were labelled: superior and inferior vermillion border of upper and lower lips, commissure and upper and lower keypoints midway between lips and commissure. (see Fig 4.)

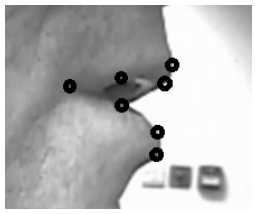

Figure 4: DeepLabCut pose estimation of Upper 4 keypoints and lower lip 4 keypoints

3.3 Photoglottograph

An external photoGlottograph (ePGG) was built similarly to the design by Honda & Maeda [5] and Amelot et al [1] and synchronised with audio, ultrasound and lip camera using sync pulses from the Micro ultrasound system. The ePPG signal is sampled at 2880Hz and reveals the glottal opening gesture with three gross states – open (high value), close (low value) and midway for voicing.

4. RESULTS: MOTOR COMMAND ANALYSIS OF THE PHRASE “THE CHALK”

Figure 5 shows a recording of the phrase “the chalk” /ðəˈtʃɔk/. Different patterns of coronal doming (high values) and grooving (low values) are observed for GG1-GG5. Vertical lines indicate time points where motor commands appear to initiate movement towards a following phonetic target. For example, motor command to transition from ʧ- ɔ starts at the beginning of region D. The glottis takes 1 period to reach its voicing target but GG2, 3 & 4 take 2

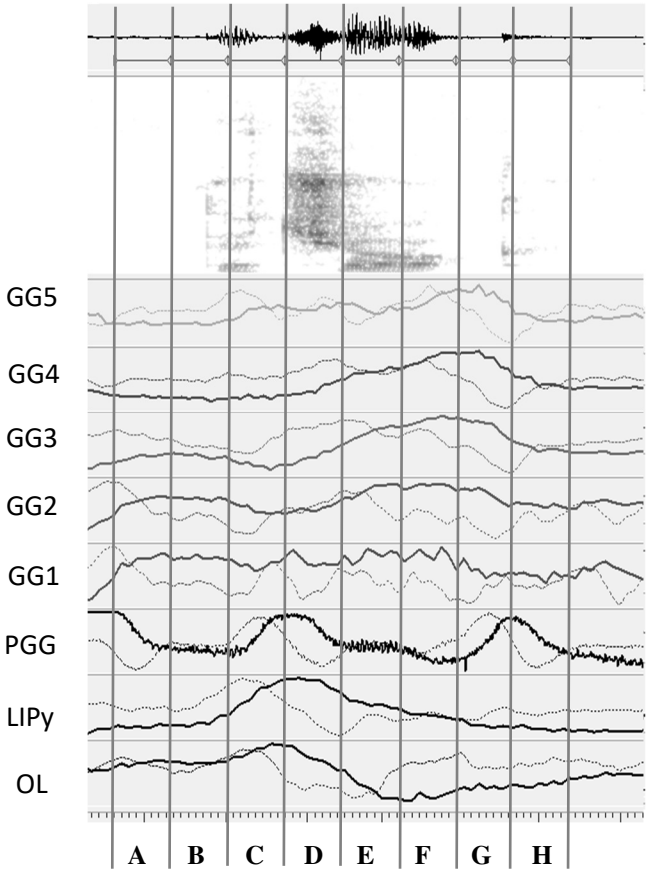

Figure 5: "the chalk" /ðəˈtʃɔk/ Vertical lines represent gating clock with clock periods labelled A-H. GG1-5 Genioglossus compartments; PGG Glotttal opening; LIPy lip commissure raising; OL oblique longitudinal compartment.
A # – ð **B&C** ð – ʧ˒ **D&E** ʧ- ɔ **F** ɔ – k **G&H** k release
Fine dotted lines are contraction velocities [94ms clock]

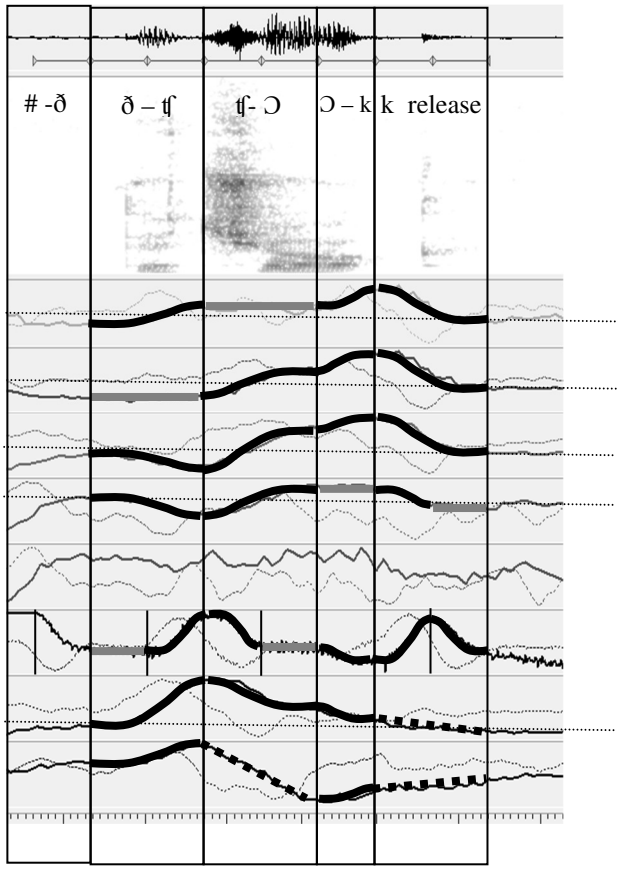

Figure 6: "the chalk" A duplicate of figure 5 but with the motor commands highlighted as sigmoid transitions – black; linear transitions – dotted; or sustained– grey. Horizontal dotted lines represent relaxed neutral positions for the synergistic unit. GG1 signal is too noisy to interpret.

periods to reach their target. This transition can also be observed in the audio spectrogram by the F2 formant movement. A command to retract the lip commissure also takes two periods. By contrast, a single period (F) is required to transition from Ɔ – k (closure target). This involves doming of GG4 &5 and closure of the glottis. Finally, the k closure release takes 2 periods (G&H) involving relaxation of GG1-5 back to a neutral position. The glottis opens in one period for a voiceless release and immediately closes again.

5. CONCLUSIONS

Although it is not prudent to draw any firm conclusions from such a small speech sample, several significant observations can be made.

Motor commands for all the articulators (tongue, lips, glottis) appear to be initiated at regular intervals of approximately 100ms. In this example the period between initiation of movement appears to be 94ms. (vertical lines in Fig 5)

Many of the transitions are sigmoid shape implying a bell-shaped velocity profile. This is most easily observed in the glottal signal where the bell-shaped velocity trace is clearly apparent in the plot (fine dotted line).

The velocity of the transition can be controlled. The GG4 transition for ʧ- ɔ takes ~200ms whereas a similar increase in doming of GG4 is achieved in ~100ms for the ɔ – k closure transition. The contraction velocity is apparently not an inherent property of a neuromuscular compartment synergy but instead can be programmed.

Transitions towards a target are not necessarily initiated at the same instant. In the ð – ʧ˒ transition both tongue and lip trajectories are initiated ~200ms before the target is reached but the glottis remains in voiced position (to realise the schwa) and movement toward the ʧ˒ target is initiated 100ms later.

Some acoustic segments arise as part of the transition towards a subsequent phonetic target. In the ð – ʧ˒ transition, the short central vowel of

"the" emerges as the tongue moves smoothly from dental closure for ð to palatal closure for ʧ.

The example exhibits no evidence of simultaneous initiation of transitions for two consecutive phones. The syllable onset consonant ʧ and the following vowel nucleus are not initiated synchronously. The transition to the ɔ vowel syllable nucleus starts at the beginning of the affrication and takes 200ms to reach the vowel target. It is perhaps easy to see why a vowel transition starting at the beginning of the affricate "segment" would be interpreted as simultaneous initiation. However, the ð – ʧ transition is initiated 200ms before the frication starts and reaches its target at the start of acoustic frication. The subsequent 100ms of frication is a consequence of the slow lingual transition to the vowel and the open glottis.

The majority of the speech signal is transitional. In this example, as soon as a target is reached the transition to the next target is initiated. [Caveat: the authors have viewed other examples where targets are sustained for at least one additional period perhaps for prosodic purposes]

6. DISCUSSION

6.1 Physiological tremor and rythmic neurons

In 1886 Horsely & Schäfer [6] observed the curve of voluntary muscle contraction invariably showed a series of undulations that succeed each other with almost exact regularity at about 10 times a second. This has become known as physiological tremor. This 8-12-Hz physiological tremor is thought to originate from the central nervous system, possibly thalamic or olivocerebellar, rhythmic neurons but the function of these central oscillations remains a subject of debate [11]. One function theory proposes that the tremor may have a role in the timing of voluntary commands [3]. Welsh and Llinás [13] have suggested that discontinuous timing of motor output would simplify the computational demand in comparison to computing commands for every instant of a movement. It has been further suggested that the 8-12Hz oscillation in the alpha wave band provides a gating mechanism that is important for synchronization of pulsatile motor signals, providing functional co-ordination of particular combinations of muscle actions [8][9][12].

Children under 10 have been observed to have a slower tremor frequency of around 6Hz compared with the adult rate of around 10Hz [10].

6.2 The pulsatile nature of movement and speech

If the premise that descending motor commands are gated at discrete intervals of approximately 100ms is accepted, then it follows that the descending motor commands are pulsatile. Put another way, the predominant idea that trajectories of movements are controlled by continuously updated descending commands and continuous feedback throughout the movement is incompatible with the concept of discrete gated commands.

In the example utterance, three of the transitions took two 100ms periods consisting of an initial pulsatile command with the option of a second corrective adjustment to that command 100ms later in the middle of the transition.

6.3 Phonological and phonetic link

This small example reveals the motor plan for the phrase "the chalk" as consisting of five discrete lingual/lip transitions to targets for ð ʧ ɔ k and a final neutral configuration. The schwa between ð and ʧ is efficiently planned to emerge as part of the ð - ʧ transition. These motor command targets map to locations in multidimensional auditory space and the cerebellum is capable of learning this mapping. As proposed by Guenther [4], phonemes may be learnt as a locus in this multidimensional auditory sensory space. More than one set of transitional commands may reach an auditory target within the locus for the desired phonological unit and the appropriate one must be selected so that the transition ends in or passes through that locus. The efficiency of the transition and how close to the centre of the phonemic auditory locus will influence which set of motor commands are selected and sequenced.

7. SUMMARY

A radical, disruptive, new theory of speech production is proposed through observations made using a new set of visualisation tools. Observation of the contraction patterns of 5 independent tongue compartments, the glottis and lips reveals motor commands that appear to be issued synchronously and discretely every ~100ms. An 8-12Hz gating of descending motor commands is proposed, generated by pulsing neurons. This central timekeeper synchronises the activation of all the articulators. It is further observed that these motor commands are programmed to reach their target contractions in either one or two cycles and in combination they specify an articulatory/auditory phone target. The tools to make these observations are widely available to anyone via the AAA software application. It is hoped that these tools can be much more widely used by other research groups to investigate many questions regarding timing, rhythm, stress and variation in speech production from this new perspective.

8. REFERENCES

[1] Amelot, A., Sathiyanarayanan, D., Maeda, S., Honda, K., Crevier-Buchman. L. 2018. Validation of a Non-Invasive System to Observe Glottal Opening and Closing: External Photoglottograph (EPGG), *11th International Conference on Voice Physiology and Biomechanics (ICVPB)* August 1-3, 2018, East Lansing, Michigan.

[2] Garcia-Rill, E., D'Onofrio, S., Luster, B., Mahaffey, S., Urbano, F. J., & Phillips, C. 2016. The 10 Hz Frequency: a fulcrum for transitional brain states. Translational brain rhythmicity, 1(1), 7.

[3] Goodman, D., & Kelso, J. A. S. 1983. Exploring the functional significance of physiological tremor: a biospectroscopic approach. *Experimental Brain Research*, 49(3), 419-431.

[4] Guenther, F. H. 2016. Neural control of speech. Mit Press.

[5] Honda K. and Maeda S. Non-invasive photoelectro-glottography method and device. *U.S. patent* 20100256503A1 (October 7, 2010).

[6] Horsley, V. A. H., & Schäfer, E. A. 1888. I. A record of experiments upon the functions of the cerebral cortex. *Philosophical Transactions of the Royal Society of London.(B.)*, (179), 1-45.

[7] Kirkham, S., Strycharczuk, P., Gorman, E., Nagamine, T., Wrench, A. Co-registration of simultaneous high-speed ultrasound and electromagnetic articulography for speech production research *In Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, Prague, 7-11 August 2023. International Phonetic Association.

[8] Llinás, R. 1991. The noncontinuous nature of movement execution, *in Motor Control: Concepts and Issues,* eds D. Humphrey and H. Freund (New York, NY: Wiley), 223–242.

[9] Llinás, R.R., 2014. The olivo-cerebellar system: a key to understanding the functional significance of intrinsic oscillatory brain properties. *Frontiers in neural circuits*, 7, p.96.

[10] Marshall, J. 1959. Physiological tremor in children. *Journal of Neurology, Neurosurgery, and Psychiatry*, 22(1), 33.

[11] McAuley, J. H., & Marsden, C. D. 2000. Physiological and pathological tremors and rhythmic central motor control. *Brain*, 123(8), 1545-1567.

[12] Welsh, J. P., Lang, E. J., Suglhara, I., & Llinás, R. 1995. Dynamic organization of motor control within the olivocerebellar system. *Nature*, 374(6521), 453-457.

[13] Welsh, J. P., & Llinás, R. 1997. Some organizing principles for the control of movement based on olivocerebellar physiology. *Progress in brain research*, 114, 449-461.

[14] Wrench, A., & Balch-Tomes, J. 2023 The compartmental tongue. *Journal of speech, language, and hearing research*, In Preparation