937

Speech motor control

Anders Löfqvist and Björn Lindblom

Haskins Laboratories, New Haven, USA and Stockholm University, Stockholm, Sweden

The development of new measurement techniques and improved models of the larynx and the vocal tract have significantly advanced our understanding of speech motor control. Recently, several groups have been using electromagnetic transduction techniques to record tongue movements. The laryngeal vibrations have been modeled and studied using techniques from non-linear dynamics. Computational models of supraglottal movements have been proposed and tested. A connectionist model that synthesizes the results obtained from observing the effects of variations in rate, stress, and phonetic context on speech kinematics has recently been proposed.

Current Opinion in Neurobiology 1994, 4:823–826

Introduction

Spoken language is a uniquely human means of communicating. By varying the positions and trajectories of the lips, the jaw, the velum, and the posture of the larynx, a speaker creates variations in air pressure and air flow in the vocal tract. These variations in pressure and flow produce the acoustic signal that we hear when listening to speech. For this signal to be not only audible but also structured in such a way that it can transmit linguistic information, parts of the vocal tract must be controlled and coordinated so that the acoustic variations in the signal conform to the language being spoken. Like any other skilled movement, speech production requires coordination of several different subsystems, that is the respiratory system, the larynx, and the different parts of the vocal tract. Overviews of issues in speech motor control are presented in [1,2]. In this review, we discuss new techniques for recording articulatory movements and recent advances in understanding speech motor control using experimental and computational approaches.

Instrumental techniques

To advance our understanding of speech motor control, transduction systems are necessary that can track movements of the different parts of the vocal tract, in particular, the tongue. Studies using X-ray have provided a wealth of information, but its inherent risks impose limitations on the amount of data that can be recorded. The development of electromagnetic transduction as an alternate measurement technique offers a number of advantages compared with other methods, provided that the proper care is taken during data collection [3,4,5•]. Several investigators are currently applying this technique and are beginning to produce useful data [6•–8•].

Respiratory system

During speech, the role of the respiratory system is to provide the air pressure that drives the laryngeal vibrations and the noise sources in the vocal tract for sounds like ‘s’ and ‘t’. Speech breathing differs from tidal breathing in that the expiratory phase is considerably longer than the inspiratory phase, and the inspirations have to be adjusted to occur at the appropriate places to serve the demands of the communication. For example, in a study of breathing patterns during reading, Winkworth *et al.* [9••] have shown that speakers are likely to time inspirations to occur at linguistic boundaries, such as sentence breaks. The amount of air taken in during speech is related to the duration of the utterance being produced, as well as to the lung volume at the end of the preceding utterance. That is, utterance duration covaries with the inspired air volume, and a low lung volume, at the end of the preceding utterance, is associated with a deeper breath for the upcoming one.

The resistance to the air in the larynx and in the vocal tract is constantly changing during speech. While proposals have been made that speakers actively compensate for such variations based on the assumption that the respiratory system is a constant flow source, a recent study by Moon *et al.* [10••] suggests that the respiratory system is a constant pressure source. Hence, no active compensations appear to be necessary, as the mechanics of the system will provide a constant pressure.

Abbreviation
EMG—electromyographic.

Laryngeal vibrations

The laryngeal vibrations that form the source for voiced sounds result from an interplay between the properties of the vocal fold tissues and the pressure distribution within the glottis resulting from the air flow through it. An improved understanding of the physics of the laryngeal vibrations has been obtained by applying analysis procedures from non-linear dynamics [11•,12••]. While there are many degrees of freedom of the vibrating glottal system, two eigenfunctions appear to explain 98% of the variance of the nodal trajectories in a model of the glottis. Thus, in spite of its complexity, the human voice source produces a fairly periodic acoustic output because only a few modes are excited and all the modes are entrained. This type of analysis and modeling work may prove useful for understanding laryngeal vibration in pathological cases. Modeling the laryngel control of fundamental frequency in speech has also clarified the contribution and interaction of the laryngeal muscles controlling the tension of the vocal folds [13•].

Modeling tongue and jaw movements

Maeda and Honda [14•] have recently calculated the acoustic characteristics of vowels from patterns of muscle activity by using EMG (electromyographic) recordings from the orbicularis oris, genioglossus anterior and posterior, the geniohyoid, the styloglossus and the hyoglossus muscles, and positional data on the mandible for 11 American English vowels. The procedure first involved converting EMG measures into values for the parameters used to control vocal tract geometry in the Maeda model [15]. A second step consisted in deriving the acoustic output (the formant frequencies) associated with the specific vocal tract configuration derived for each vowel. The conversion of EMG to positional variables involved calculating an estimate of 'force' as the sum of scaled EMG activity measures. For instance, the tongue position parameter was computed as a combination of measures from the hyoglossus, the genioglossus posterior and the geniohyoid, whereas tongue shape was computed as a sum of contributions from the styloglossus and the genioglossus anterior. Despite the simplicity of the hypothesized rules, surprisingly good agreement between observed and computed formant values was obtained.

Wilhelms-Tricarico [16•] applied finite element methods to simulate the behavior of the tongue during speech production. Movements and shape deformations are derived as the solutions of a non-linear second-order constraint system of ordinary differential equations. Modeled on the basis of cardiac biomechanical data, muscle fibers are represented as vector fields specifying the directions in which active and passive tensile stress is produced. In a preliminary implementation, an eight-muscle tongue model was tested with normalized EMG data as input [16•]. Although these efforts cannot yet be seen as having resulted in a definitive tongue model for speech, they do seem to offer a novel and promising approach to the problem of representing soft-tissue articulators.

Jaw movements in speech have been analyzed in terms of a seven-muscle model of jaw and hyoid motion [17] based on the equilibrium point hypothesis (the λ-model [18]): muscle lengths (λs) at which motoneuron recruitment begins are controlled by central neural commands specifying the jaw's equilibrium position. Laboissière *et al.* [17] found that combinations of λs are invoked to produce specific patterns of rotation, translation and muscle co-contractions.

Perturbation studies

A useful experimental paradigm for studying movement coordination is the introduction of unexpected perturbations during ongoing motor acts. The rationale for this research is that the nature and time course of the responses to the load are thought to reveal the motor organization and reflex structure of the ongoing act. Studies using this methodology have generally found that compensations are rapid and functional in the sense that the goal of the act is achieved, that is, a correctly perceived signal is generated. This experimental paradigm has been extended to the more remote articulators, the lower lip and the larynx, in a study where a perturbation was applied to the lower lip in the production of the stop consonant 'p' [19•]. The lips and the larynx can be considered remote because the biomechanical coupling between them is minimal. For voiceless sounds such as 'p', 't' and 'k', the closure in the vocal tract has to be coordinated with the opening (abduction) and closing (adduction) movement of the vocal folds that is made to momentarily arrest the glottal vibrations. Experimental evidence suggests that the temporal coupling between the upper articulators and the larynx is particularly tight in making the closure, but less so at the release of the closure [20•]. The results of the perturbation study indicated that the onset of the glottal abduction was delayed in perturbed trials, possibly to maintain the phasing between the lips and the larynx [19•]. At the same time, the duration of the oral closure was shortened by the perturbation, while the duration of the glottal opening/closing movement increased. As a consequence, acoustic differences were found between perturbed and control trials [19•].

Context, speech rate and stress

A central problem in speech motor control is to understand how the articulatory attributes of individual sounds change as a function of context, speech rate and stress. For example, a velar stop consonant such as 'k' requires a closure between the tongue body and the hard palate. Both the place of contact and the movement trajectory into and out of the closure differ depending on the quality of the surrounding vowel, for example, between the sequences 'ika' and 'aki'. Generally referred to as co-articulation, co-production, or aggregation, such contextually conditioned variability has provided a rich testing ground for theories of speech production.

Based on kinematic studies of lip and jaw movements for consonants produced with a closure of the lips, such as 'p', 'b' and 'm', and occurring in different vocalic contexts, Gracco [21•] hypothesizes that observable speech movements reflect the combination of stored representations and sensorimotor processes that produce complex vocal tract actions from simple operations; this hypothesis is in good agreement with the findings from studies applying mechanical perturbations to speech articulators [19•]. It thus suggests that a distinction can be made between speech motor programs, as goal-directed actions, and dynamic programming processes that provide adaptive adjustments to speech motor sequences.

Studies of velocity profiles of both speech and non-speech movements have suggested that they may have a stable topology across changes in speaking rate, reflecting a strategy for motor control. A study of lip and tongue movements across a wide variety of speaking rates by Adams *et al.* [22•] provides experimental evidence that several topological measures of velocity profiles do in fact change with rate. These measures included symmetry, kurtosis, number of velocity peaks, as well as a measure of the profile's geometry based on peak velocity, movement amplitude, and movement duration.

Trajectories of tongue movements for velar stop consonants, such as 'k' and 'g', generally show curved paths in extrinsic space, in contrast to manual reaching and pointing movements. The same is the case for jaw movements in both speech and mastication [23•]. When plotted in joint coordinates, jaw motion paths during speech also show straight lines, but with different slopes for different speech sounds. The intercepts of the jaw paths also varied as a function of speaking volume. Together, these findings suggest that rotation and horizontal translation of the jaw can be independently controlled. These data bear on the question of the coordinate spaces used in speech motor control. While a joint-based strategy is consistent with the data, it is also reasonable to assume that speech movements are planned and controlled in extrinsic space, as the vocal tract shape is directly related to the acoustic properties of the sound being produced.

Further evidence that speech motor control is prospectively organized (i.e. output-oriented) comes from acoustic phonetic work on speaking styles by Moon and Lindblom [24•] showing that words spoken more 'clearly' are not merely louder but involve articulatory processes that have been modified to compensate for context-dependent phenomena such as 'formant undershoot' (vowel reduction). These findings indicate that in clear speech, speech movements are less reduced and tend to be more intelligible.

A number of experimental findings regarding the effects of variations in speaking rate and phonetic context on speech movement kinematics have been integrated and given a unified treatment in a connectionist model [25••]. In this model, the movement targets of different articulators for a given sound are expressed as spaces using convex hulls. The size of the hulls serves as an index of the precision needed for producing a given sound. The place of constriction or closure for consonants is associated with smaller hulls than the targets for vowels. A change in speaking rate can be modeled by making the hull for a given sound segment larger or smaller; a shrinking hull is associated with a decrease in speaking rate. The change of size of the hull is greater for vowels than for consonants. Consequently, in the model, an increase in speaking rate is associated with an increased movement velocity for vowels but no change, or even a decrease in velocity, for consonants, as has been found in several experimental studies.

Conclusion

Significant progress in understanding the coordination and control of speech movements has been made as a result of two different but essentially complementary research approaches. First, improved measurement techniques (e.g. electromagnetic transduction) have appeared that have provided new valuable information on articulatory processes. Second, investigators have been able to build computational models that exhibit much greater physiological and physical sophistication than earlier frameworks. These tools have been used to shed new light on how the vocal folds vibrate and how patterns of muscular activity influence articulatory parameters, such as the rotation and translation of the mandible and the position and shape of the tongue. Connectionist modeling has been successfully applied to present a unified theoretical account of a wide range of experimental facts about the rate- and context-dependence of speech movements.

Acknowledgements

Manuscript preparation supported by Grant DC-00865 from the National Institute on Deafness and other Communication Disorders to A Löfqvist, and by Grant BNS-9011894 from the National Science Foundation, by funds from the Centre National de Recherche Scientifique (CNRS), France, and from the HSFR, the Swedish National Council for Research in the Humanities and the Social Sciences awarded to B Lindblom.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest
•• of outstanding interest

1. Gracco V: **Characteristics of speech as a motor control system.** In *Cerebral control of speech and limb movements.* Edited by Hammond G. Amsterdam: Elsevier; 1990:3–28.

2. Löfqvist A: **Speech as audible gestures.** In *Speech production and speech modelling.* Edited by Hardcastle W, Marchal A. Dordrecht: Kluwer; 1990:289–322.

3. Schönle P: *Elektromagnetische artikulographie.* Berlin: Springer; 1988.

4. Perkell J, Cohen M, Svirsky M, Matthies M, Garabieta I, Jackson M: **Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements.** *J Acous Soc Am* 1992, 92:3078–3096.

5. • Hoole P, Tillman H: *Proceedings of the ACCOR workshop on electromagnetic articulography in phonetic research,* vol 31. München: Forschungsberichte, Institut für Phonetik und Sprachliche Kommunikation der Universität München; 1993.
A collection of papers describing technical and signal processing issues, as well as empirical studies of speech production.

6. • Perkell J, Matthies M, Svirsky M, Jordan M: **Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: a pilot motor equivalence study.** *J Acoust Soc Am* 1993, 93:2948–2961.
Examines the possibility that speakers use less lip rounding and more tongue retraction in producing the rounded vowel [u] (and vice versa) as these articulatory gestures have similar acoustic effects; some experimental evidence in favor of this notion is presented.

7. • Löfqvist A, Gracco V: **Tongue body kinematics in velar stop production: influences of consonant voicing and vowel context.** *Phonetica* 1994, 51:52–67.
A quantitative study of tongue movements as a function of phonetic context suggesting that the influence of stop consonant voicing on articulatory kinematics is specific to the articulator.

8. • Nguyen N, Hoole P, Marchal A: **Regenerating the spectral shapes of [s] and [ʃ] from a limited set of articulatory parameters.** *J Acoust Soc Am* 1994, 96:33–39.
Explores constraints on tongue movements in the production of voiceless fricatives.

9. •• Winkworth A, Ellis E, Adams R: **Variability and consistency in speech breathing during reading: lung volumes, speech intensity, and linguistic factors.** *J Speech Hear Res* 1994, 37:535–556.
A study of respiratory patterns during reading showing differences within and across subjects in respiratory volumes, but little variation in the timing of inspirations in relation to linguistic factors.

10. •• Moon J, Folkins J, Smith A, Luschei E: **Air pressure regulation during speech production.** *J Acoust Soc Am* 1993, 94:54–63.
An empirical and modeling study of respiratory control suggesting that the respiratory system is a constant pressure source, rather than being a constant flow source, as has been assumed in the past.

11. • Lucero J: **Dynamics of the two-mass model of the vocal folds: equilibria, bifurcations, and oscillation region.** *J Acoust Soc Am* 1993, 94:3104–3111.
An examination of oscillation properties and chaotic behavior in a model of the vocal folds.

12. •• Berry D, Herzel H, Titze I, Krischer K: **Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions.** *J Acoust Soc Am* 1994, 95:3595–3604.
The authors used a model to examine the modes of vocal fold vibrations. Their results suggest that only a few modes are excited and that all the modes are entrained.

13. • Farley G: **A quantitative model of voice F_0 control.** *J Acoust Soc Am* 1994, 95:1017–1029.
A detailed model of the influences of laryngeal muscles on the fundamental frequency of phonation.

14. • Maeda S, Honda K: **From EMG to formant patterns of vowels: the implication of vowel spaces.** *Phonetica* 1994, 51:17–29.
Derives the acoustic properties of vowels by modeling tongue movements using EMG recordings of tongue muscles.

15. Maeda S: **Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model.** In *Speech production and speech modelling.* Edited by Hardcastle W, Marchal A. Dordrecht: Kluwer; 1990:131–149.

16. • Wilhelms-Tricarico R: **Physiological modeling of speech production: methods for modeling soft-tissue articulators.** *J Acoust Soc Am* 1994, in press.
Presents a model of the tongue using finite element methods.

17. Laboissière R, Ostry DJ, Feldman AG: **A model of jaw movement control based on the equilibrium point hypothesis.** *J Acoust Soc Am* 1994, 95:2923.

18. Feldman AG: **Once more on the equilibrium-point hypothesis for motor control.** *J Mot Behav* 1986, 18:17–54.

19. • Munhall K, Löfqvist A, Kelso J: **Lip-larynx coordination in speech: effects of mechanical perturbations to the lower lip.** *J Acoust Soc Am* 1994, 95:3605–3616.
Examines the effects on lip and jaw kinematic and on lip/larynx phasing from mechanical perturbations to the lower lip and their acoustic consequences.

20. • Gracco V, Löfqvist A: **Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements.** *J Neurosci* 1994, in press.
A detailed study of lip, jaw, and larynx coordination for different voiceless sounds.

21. • Gracco V: **Some organizational characteristics of speech movement control.** *J Speech Hear Res* 1994, 37:4–27.
An empirical study of lip movement control in different phonetic contexts.

22. • Adams S, Weismer G, Kent R: **Speaking rate and speech movement velocity profiles.** *J Speech Hear Res* 1993, 36:41–54.
The authors studied the velocity profiles of lip and tongue movements across variations in speaking rate and showed changes in the topology of the profiles as a function of rate.

23. • Ostry D, Munhall K: **Control of jaw orientation and position in mastication and speech.** *J Neurophysiol* 1994, in press.
Suggests that jaw rotation and horizontal translation can be controlled independently in speech and chewing.

24. • Moon S-J, Lindblom B: **Interaction between duration, context, and speaking style in English stressed vowels.** *J Acoust Soc Am* 1994, 96:40–55.
Reports findings compatible with a revised and biomechanically motivated version of the so-called 'formant undershoot model' that predicts formant patterns from numerical information on three variables: consonant context, vowel duration and rate of formant frequency change.

25. •• Guenther F: *Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production.* Boston: Boston University Center for Adaptive Systems; 1994. [Technical Report CAS/CNS-94-012].
A connectionist model of speech motor control incorporating a learning, 'babbling', stage and providing a unified framework for understanding a wide variety of experimental findings.

A Löfqvist, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695, USA.

B Lindblom, Department of Linguistics, Stockholm University, Stockholm S-10691, Sweden.