Iskarous, K. (2003). Task dynamics of the tongue. In S. Palethorpe & M. Tabain (Eds.), *Proceedings of the 6th International Seminar on Speech Production* (pp. 107-112). Sydney: Macquarie University.

1329

TASK DYNAMICS OF THE TONGUE

Khalil Iskarous

Haskins Laboratories

ABSTRACT: The motion of the tongue can be analysed in a variety of ways: curved motion of individual points on the surface, change in the activations of muscles, change of curvature in the dorsum and tip, change in area-function, and change in formant pattern. Each of these descriptions provides a coordinate reference frame with which to describe the motion of the tongue. But how are these different descriptions related to each other? And which of these coordinate frames is used to control tongue movement? In this paper, I propose a task-dynamic framework in which to relate the different descriptions by coordinate transformations between the levels.

INTRODUCTION

The tongue is an essential contributor to the contrasts that distinguish the units of speech, and it stands out in the complexity of its movement. This complexity is attested for in the number of ways that its motion can be described. Several measurements can be made at the surface of the tongue itself. The motion can be quantified in terms of the trajectory of individual points or the change in shape of different portions (Houde 1968, Perkell 1969, Gracco and Lõfqvist 1999, Stone 1990). The motion can also be studied in terms of its cause and effects. Muscle activation is the immediate cause of the motion of the surface of the tongue, therefore it is possible to study the motion in a coordinate system where the variables are muscle activations. The most immediate effect of the motion of the surface is the change in cross-sectional area of the vocal tract, therefore it is also possible to study tongue motion from the perspective of how the area function changes as a function of time, with little attention given to the motion of individual points or the shape of the tongue itself (Wood 1982, Iskarous 2001). A distal effect of tongue motion is the change in formant pattern, and it is possible to analyse the components of tongue motion that are acoustically interpreted in a coordinate system with formant variables. All of these are related descriptions of the motion of what is arguably the most complex speech organ. To truly understand the behaviour of the tongue, however, it is necessary to relate the different levels to each other, and indeed the relation between some levels have already been investigated. The relation between muscle activation and tongue shape change has been studied using factor-analytic techniques (Perrier et al. 1996) and the relation between area function change and formant change has been studied in terms of the Acoustic Theory of Speech Production. In this paper, the focus is on the relation between individual point movement, tongue shape change, and the effect on area function. Task Dynamics (Saltzman and Munhall 1989) can be used in a fruitful manner to relate these levels of analysis. In this framework, different levels are related by coordinate transformations that map the dynamics of one variable to those of another.

LEVELS OF ANALYSIS OF TONGUE MOTION

A common methodology for studying tongue movement is by tracking individual points on its surface, usually in the midsagittal plane. The paths are then analysed in terms of their horizontal and vertical components (Browman and Goldstein 1990, Perkell et al. 1993, Gracco and Lõfqvist 1999, Westbury et al. 1998) or the shape of the path, the loops (Houde 1968). Electromagnetic Articulography (EMA) and X-ray Microbeam (XRMB) are two techniques used to collect motion data from this perspective. This level of analysis is in principle high-dimensional, since there is no limit to the number of points on the tongue that can be tracked. Technical limitation on EMMA and XRMB usually allow for only three to five points to be tracked. Recent advances in dynamic MRI imaging allow for magnetic tagging of small cubes of the tongue and tracking their motion using a stroboscopic technique (Stone et al. 2001a). It is therefore possible as this technology develops to follow a large number of points on the surface and indeed inside the tongue. Much has been learned about the coarticulatory behaviour of the tongue and about the prosodic effects on articulatory movement from analysis at this level (Gracco and Lõfqvist 1999). The high-dimensionality of this level of description, however, makes it unlikely that it is the level controlled by the speech motor control system. The number of degrees of freedom could of course be arbitrarily reduced as is done in EMA and XRMB experiments, but the particular points chosen are in no way special. It is unlikely, therefore, that the system directly controls the trajectory of every point on the tongue, since there are too many points to control.

It is commonly assumed that the variable controlled by the speech motor system is muscle activation. There are about 10 muscles, internal and external, that form the structure of the tongue, and most of them can be differentially controlled, that is different segments of the tongue can be active to different extents (Stone et al. 2003, Sokoloff 2000). The motion of the tongue can therefore be described in terms of the activity of different parts of the different muscles, since it is muscle activation that causes the motion. But due to local innervation of the tongue, there are an estimated 13,000 degrees of freedom at that level (Stone et al. 2003), based on the number of hypoglossal motorneurons innervating the tongue (O'Kusky and Norman, 1995). Of course this is Bernstein's problem of degrees of freedom that has emerged in every branch of motor control—muscle activations are unlikely as the control variable, because there are simply too many degrees of freedom at that level (Bernstein 1996). Perrier et al. (1996) proposed within the Equilibrium Point Hypothesis (EPH) framework that the controlled variable λ is a tongue shape (derived from factor analysis) and that muscle activation is derived by a muscle model from λ. In the Articulatory Phonology framework, the goal for tongue motion is taken to be the size and location of the constriction within the vocal tract (Browman and Goldstein 1990). A third higher control variable that has been proposed is formant values. Perkell et al. (1993) and Guenther et al. (1999) have proposed that the relevant aspect of tongue movement is that which has an effect on the acoustic signal, since that is what is communicated to the listener. Indeed one of the most common methods of studying lingual coarticulation has been through formant patterns (Öhman 1966). The description of the tongue through its shape, effect on the area function, and effect on acoustics are low-dimensional, in contrast to point-motion and muscle activation. The formant description is obviously low dimensional, since only three formants need to be tracked. It is less clear however, why the two other descriptions are low dimensional. The next section describes work that shows that they can indeed be considered low dimensional.

TONGUE SHAPE AND AREA FUNCTION CHANGE

Tongue shape is difficult to analyse and quantify, since the tongue is continuously deformable. Ultrasound and MRI imaging of the vocal tract allow researchers to extract full sections of the vocal tract that reveal the shape of the tongue. The edge is usually detected by using an active method for spline fitting (Stone et al. 2001b); therefore it is possible to use the coefficients of the spline as the representation of the tongue shape. The problem with that approach is that even though the spline coefficients specify the tongue shape, they do not have an easily grasped geometric meaning. To solve this problem, we have proposed a new method for analysing tongue shapes that uses conic curves (Iskarous et al. 2001). The main advantage to these curves is that they are flexible enough to represent a large variety of tongue shapes, while at the same time requiring very few parameters to specify their shape. Moreover the geometric parameters that describe their shape turn out to be ideally suited for describing the shape of the tongue, since they have a clear phonetic interpretation.

There are many sets of parameters that can be used to specify a conic. For the phonetic interpretation to be available, however, it is necessary to use the projective parametrization. The different conics can be generated by intersecting a plane with a conic, as seen in Figure 1. If the angle of the plane is varied continuously, a continuous family of conics running from elliptic to parabolic to hyperbolic is generated. It is possible to parametrize this continuum with three numbers. One parameter specifies the amount of curvature, and its variation can be seen in Figure 2 Left. When the parameter equals zero, the conic degenerates to a line. Between 0 and 1, the conic is an elliptic arc and when it is exactly 1, the conic is a parabolic arc. When the parameter exceeds 1, the conic is a hyperbolic arc. The second parameter specifies the location of curvature, and its variation can be seen in Figure 2 Right. The second parameter is location of curvature. As it varies, the position of the most curved part of the conic varies. Figure 2, right shows two families of arcs that vary in amount of curvature at two different locations. The third parameter is the overall size of the curve.

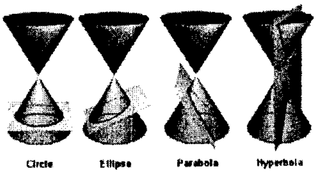

Figure 1. Conics as intersections of a plane and a cone.

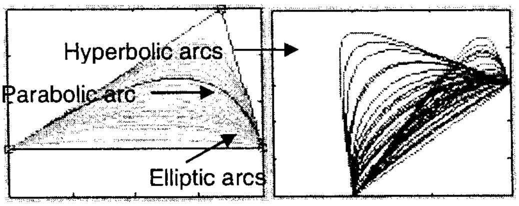

Figure 2. Variation of conic as a function of two parameters. Left: Location of curvature is fixed and amount of curvature is varied. Right: Location and amount of curvature varied.

When conic arcs are fit to the tongue the parameters of the conic take on a phonetic meaning. The location of curvature is correlated with place of constriction and the amount of constriction is correlated to constriction degree. We have written programs to fit conic arcs to data collected from Ultrasound and MRI imaging of the vocal track. The quality of the fits is high, since conic shapes capture the variety of configurations that the tongue assumes. Examples can be seen in Figure 3. The second example shows that when the tongue assumes a shape that occurs for example in American English /r/ or dark /l/, two sets of conic parameters must be used.

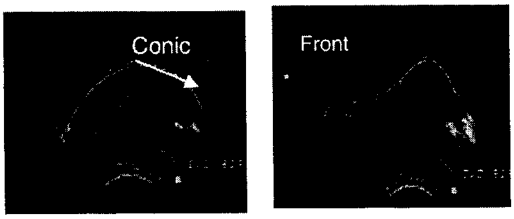

Figure 3. Examples of conic fits to the tongue. Left: American English /e/. Right: American English /r/

The shape of the tongue can therefore be specified in a low-dimensional description, as we have for the acoustics. Moreover, the close relation between the geometric and phonetic parameters allows us to use the geometric parameters to directly quantify the phonetic goal of the tongue in a low-order manner. It is less clear how the effect of tongue movement on area function change could be described with a few parameters, since there are at least as many points in the vocal tract to measure the cross-sectional area of the vocal tract as there are points of the tongue, and the point-movement description has already been argued to be high dimensional.

To understand how the area function contribution of tongue movement can be described low-dimensionally, one has to look at dynamic behaviour rather than single configurations of the tongue. In a study of how the tongue changes its shape from initial to final configuration in 600 two segment sequences of the CV, VC, VV, and CC type, Iskarous (2001, In press), showed that even though every point of the tongue moves from the initial to the final configuration, that the movement occurs in such a way that it is orthogonal to the hard structures at some locations in the vocal tract and parallel to the hard structures at other locations. When the motion is orthogonal to the hard structures, maximal change in the area function occurs there, while parallel motion does not contribute to area function change. In a large number of diverse transitions in both Canadian English and French showed that the locations in the vocal tract where tongue motion incurs maximum area function change are the locations traditionally identified as the places of articulations of the segments in the transition. The location where tongue point movement is parallel to the hard structures is dubbed the "pivot point" after Stone (1990). Examples of transitions of the pivot type are shown in Figure 4. Another pattern that occurs, the arch, is a degenerate case of the pivot that occurs when the places of articulation of the two segments in the transition are too close to each other. 84% of the 600 transitions analysed are accounted for by the two patterns.

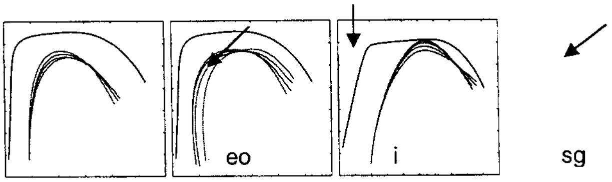

Figure 4. Pivot Point examples. Subjects facing right. Pivots are indicated with an arrow.

In a pivot transition, the vocal tract is divided into 2 regions of maximal area function change divided by a pivot point. This division is fundamentally dynamic, since it is only due to the change of tongue shape over time that the vocal tract is discretized. For any static configuration, the area function can be measured at any point along the longitudinal axis of the vocal tract, a high dimensional description. Pivot motion reduces the degrees of freedom enormously to two locations of maximal area function change and the degrees of constriction at these locations. Therefore the contribution of tongue movement to area function change is also low dimensional, as are formant change and shape change. It must be noted however, that the reduction in dimensionality of degrees of freedom at the area function level is achieved naturally through tongue dynamics, not arbitrarily as is done at the fleshpoint level.

TASK DYNAMICS

The existence of several reference frames within which a particular motion can be described is commonplace in motor control. Based on Bernstein (1996)'s ideas, Task Dynamics (Saltzman & Munhall 1989) was developed to relate different motor reference frames by coordinate transformations. For instance, arm motion can be described in terms of the position of the hand in space or the angles made at the joints. The position of the hand is seen as the task or controlled variable. A coordinate transformation from angle coordinates to the rectangular coordinates of the hand is done with a polar-to-rectangular transformation familiar from geometry and robotics. Once a particular hand position is planned by the system, however, the angles at the joints need to be computed to achieve the task, and that is done by an inverse transformation from hand position to angles accomplished by the Moore-Penrose Pseudoinverse. The application of Task Dynamics to speech (Saltzman and Munhall 1989) relates a gestural level to an articulator level. In this paper, the framework is applied to the various levels of tongue movement. In Figure 5, the different levels discussed are presented with the proposed transformations between them.

$$\begin{tikzcd}
\text{Flesh Point} \arrow[r, shift left, "\text{Normal Equations}"] & \text{Shape} \arrow[l, shift left, "\text{Polynomial Eval.}"] \arrow[r, shift left, "\text{Identity}"] & \text{Area Function} \arrow[l, shift left, "\text{Identity}"] \arrow[r, shift left, "\text{Acous. Theory of Sp. Pr.}"] & \text{Formants} \arrow[l, shift left, "\text{Acoustic Inversion}"]
\end{tikzcd}$$

Figure 5. Task Dynamic levels and transformations between the levels

The transformations are mathematically different from the ones used in the motor theory literature, but that is simply due to the difference between speech and other motor tasks. Now I present the details of the transformations:

1. Fleshpoint ↔ Shape: If the shape of the tongue is specified in terms of the location and degree of the constriction by the conic parameters and the position of the points on the tongue edge are specified in terms of their x and y coordinates, then the coordinate transformations between the levels must transform in a forward and inverse manner between rectangular coordinates of points and conic shape descriptors. The conic is mathematically represented as a rational quadratic B-spline (NURB). The conic parameters describing tongue shape are the coefficients of the polynomial representing the curve (actually it's a set of polynomials since the curve is specified parametrically); and the points on the curve are the fleshpoints. The polynomial is therefore itself the coordinate transformation. Evaluation of the polynomial

yields the rectangular coordinates of as many points on the tongue as desired from a specification of the conic coefficients. Inversion of the process is fitting a shape to a set of points which is done by solving a set of Normal Equations. Programs for both the forward and inverse transformations between these levels are routinely used for the detection of tongue edges from ultrasound images (Iskarous et al. 2001) and for the generation of motion points of the tongue from a specification of the conic shape used for Articulatory Synthesis (Iskarous et al. 2003).

2. Shape ↔ Area Function: The shape of the tongue is described by conic parameters of location and degree of constriction and the area function is described by the location and degree of constriction of the final and initial constrictions in a sequence. Therefore due to the phonetic interpretation of the conic parameters and the reduction of the high dimensional area function to two vocal tract regions by pivot dynamics, a simple identity relates the two levels. If any other shape parameters are used and if the area function were to be specified in continuous form, the maps between the levels would be quite complicated. However, the contribution of the shape of the palate and the rest of the hard structures to the area function is not accounted for here. But since these hard structures are immobile, their shape acts only as a constraint on speech production planning, whereas the constriction location are degree are the active part of the dynamics. Future work would determine how the passive shape of the hard structures constrains transformation between the active levels.

3. Area Function ↔ Formants: The transformation from area function to formants has been known for a long time. If the controlled variable in speech production were formants, the speaker must solve the inverse problem of acoustics to determine the area function necessary to accomplish the desired formant pattern. Interestingly, many who argue against the view that speech perception extracts gestural information citing the impossibility of solving the inverse problem of acoustics espouse the view that the formants are the goal of production, which requires the solution of the same problem.

CONCLUSION

In this paper, several levels of representation of tongue movement have been proposed, along with transformations between the levels. An issue for further research is how to decide which level is the one that is actually controlled by the speech motor control system. A promising method for this purpose used in motor control and recently introduced in speech by Saltzman et al. (In Press) is the Uncontrolled Manifold Method, where potential control variables are quantitatively compared with respect to their stability to perturbation. Future work will attempt to use this method to decide which of the variables describing tongue movement is the control variable, given the representations of the levels proposed here.

ACKNOWLEDGEMENTS

This research was supported by NIH grant DC-02717 to Haskins Laboratories.

REFERENCES

Bernstein, N. (1996) *Dexterity and its development.* Mahwah, NJ : L. Erlbaum Assoc.

Browman, C. P. and Goldstein, L. (1990) "Tiers in articulatory phonology, with some implications for casual speech" *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* J. Kingston and M. E. Beckman (Eds.). Cambridge: Cambridge University Press. 341-376.

Gracco, V. and Löfqvist, A. (1999) "Coarticulation of tongue and jaw movements in speech" *The Journal of the Acoustical Society of America*, 105, 1354-1360.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M. and Perkell, J. S. (1999) "Articulatory tradeoffs reduce acoustic variability during American English /r/ production" *Journal of the Acoustical Society of America*, **105(5)**: 2854-2865.

Houde, R. (1968), A study of tongue body motion during selected speech sounds, Technical report, Speech Communications Research Laboratories, Inc., Santa Barbara.

Iskarous, K. (2001) *Dynamic Acoustic-Articulatory Relations*, University of Illinois at Urbana-Champaign Doctoral Dissertation.

Iskarous, K. (In Press) "Patterns of Tongue Movement", *Journal of Phonetics.*

Iskarous, K., Goldstein, L., Whalen, D.H., Tiede, M., and Rubin, P. (2003) "CASY: The Haskins Configurable Articulatory Synthesizer" *Proceedings of ICPhS XV.*

Iskarous, K., Whalen, D.H. and Mattingly, I. "Modeling tongue shapes with conic arcs," *Journal of the Acoustical Society of America,* vol. 110, p. 2760, 2001.

O'Kusky JR & Norman MG (1995). "Sudden infant death syndrome: increased number of synapses in the hypoglossal nucleus" *Journal of Neuropathology and Experimental Neurology, 54*(5), 627-634.

Öhman, S. (1966) "Coarticulation in vcv sequences: spectrographic measurements" *The Journal of The Acoustical Society of America* (39): 151-168.

Perrier, P., Loevenbruck, H. and Payan, Y. (1996) "Control of tongue movements in speech: The equilibrium point hypothesis perspective" *Journal of Phonetics* (24): 53-75.

Perkell, J. (1969), *Physiology of Speech Production*, MIT Press, Cambridge.

Perkell, J. S., Matthies, M. L., Svirsky, M. A. and Jordan, M. I. (1993) "Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot 'motor equivalence' study" *Journal of the Acoustical Society of America* **93**: 2948-2961.

Saltzman, E., Kubo, M., & Tsao, C.-C. (in press) "Controlled variables, the uncontrolled manifold method, and the task-dynamic model of speech production" In P. Divenyi & G. Meyer, (Eds.). *Dynamics of speech production and perception.* Amsterdam: IOP Press.

Saltzman, E. and Munhall, K. (1989) "A dynamical approach to gestural patterning in speech production" *Ecological Psychology* (1), 333:382.

Scholtz, J.P. & Schöner, G. (1999) "The uncontrolled manifold concept: Identifying control variables for a functional task" *Experimental Brain Research* (135) 289-306.

Sokoloff, A.J., 2000, "Localization and contractile properties of intrinsic longitudinal motor units of the rat tongue" *Journal of Neurophysiology,* **84**, 827-835.

Stone, M., 1990, "A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data" *Journal of the Acoustical Society of America*, **87**, 2207-2217.

Stone, M., Davis, E., Douglas, A., Ness Aiver, M., Gullapalli, R., Levine, W., and Lundberg, A., 2001a, "Modeling motion of the internal tongue from tagged cine-MRI images" *Journal of the Acoustical Society of America*, **109**, 2974-2982.

Stone, M., Davis, E., Douglas, A., Ness Aiver, M., Gullapalli, R., Levine, W., and Lundberg, A., 2001b, "Modeling tongue surface contours from cine-MRI images" *Journal of Speech, Language, and Hearing Research*, **44**, 1026-1040.

Stone, M., Parthasarathy, V., Iskarous, K., NessAiver, M., and Prince., J., 2003, "Tissue strains and tongue shapes: Combining tMRI and Ultrasound" *Proceedings of ICPhS XV.*

Westbury, J. R., Hashi, M. and Lindstrom, M. J. (1998) "Differences among speakers in lingual articulation for American English /r/" *Speech Communication* 26, 203-226.

Wood, S. *X-ray and model studies of vowel articulation.* Working Papers 23. Lund, Sweden: Dept. of Linguistics, Lund University, 1982.