Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Towards a Musical Gesture in the Perspective of Music as a Dynamical System

Beatriz Raposo de Medeiros*[1]

Department of Linguistics, University of São Paulo, Brazil

[1]biarm@usp.br

ABSTRACT

Assuming a perspective of music as a dynamical system in the domain of cognition implies adopting the notion that the cognitive structures (nervous system, body and environment) are integrated. In other words, in each behavior that involves acting and knowing – e.g., a football player kicking a corner ball– cognitive structures act as an entire system. The dynamical view provides the necessary tools and the language required to deal with time, movement and change over time. We present a locus of convergence among studies with different views on music as a dynamical system, whereafter we propose a musical gesture based on the same dynamical principles which in the domain of Linguistics led to a phonological unit called articulatory gesture. The singing voice is presented as a plausible musical gesture as it produces tones and durations combined in order to provide the musical information. This information can be understood as specific tones in a given scale system and rhythmic structure and is part of the musical unit proposed here. The articulatory movements of the singing voice produced by the larynx characterize this unit as a unit of action. Thus we suggest a larynx modeling for music production in an initial attempt to view the singing voice as a basic realization of music, organized and coordinated as a musical gesture.

I. INTRODUCTION

We propose a musical gesture based on a locus of convergence of some approaches on musical aspects related to dynamical systems. Albeit somewhat ambitious, we choose to undertake what is still a tentative proposition of a musical gesture, representing an initial state of a research project.

In the pursuit for a notion of musical gesture, we propose to advance in three stages:

- Comprehension of a Dynamical System, especially in the domain of Cognition
- A locus of convergence of certain dynamical views on music.
- A possible musical gesture: the singing voice and its articulatory aspect.

We argue in favor of understanding music as movement, not only metaphorically, and view the singing voice as a basic realization of music. A possible larynx model is conjectured and proposed as a contribution to the understanding of the musical gesture, since it is produced by the articulatory movements involved in singing.

Viewing musical gesture in the perspective of music as a dynamical system is a form of dealing with issues that are still somewhat scantily discussed in the fields of music theory and music cognition as, for instance, what is the nature of a musical primitive. Moreover, it assists in clarifying the idea of movement in music which seems to be most often used as related to the acoustic phenomenon, but is only rarely related to the movement (or gesture) that causes this music, or, for that matter, to its cognitive structure.

II. DYNAMICAL SYSTEMS AND COGNITION

To begin with, we will present a fairly wide definition of a dynamic system. Subsequently, we establish its relation with cognition, since a cognitive system assumed as dynamic may derive relevant benefits from the tools and language involved in a dynamic approach (Port & van Gelder, 1995).

A dynamical approach is no novelty, and has been applied by several domains of scientific endeavor over the last sixty years or so. A definition such as the following helps us understand why a dynamic system is *par excellence* applied in mathematics: “The dynamics is a rule that transforms one point in the phase space (that is, a world state), representing the state of the system ‘now’, into another point (= world state), representing the state of the system one time unit "later". In mathematical language, the dynamics is a function mapping world states into world states.” (Hochman).

The simplest definition of a dynamical system states that in such a system changes take place over time. It is possible to observe a general state of the system, as well as a given state at some moment in time. Examples of dynamical systems can be the weather, hand movements, tennis or football ball motion, speech, and so many others. A dynamic system may have its movement or change of state explained by a differential equation ($x' = F(x)$, the introductory concepts of which can be verified in detail in Norton (1995).

This aspect and others which typify a dynamical system have been adequately summarized by Marin & Peltzer-Karpf (2009), and involve the *time evolution law, self-organization* and *emergence*. It is not the purpose of this paper to deal with these concepts, although they are of the utmost importance when one chooses to follow a theoretical framework with a dynamic viewpoint. In this sense, their explanations must be pursued in dynamical cognitive approaches, so as to constantly shed light on the non-static nature of cognition and on notions such as coordination among systems, and their opposition to the classical cognitive notion of central command or control.

It would seem inevitable to deal with cognitive systems by comparing the computational approach (that which considers the cognitive system to be a computational system) to the dynamic approach (for which the cognitive system is dynamic). We will not dwell on this comparison, although we admit it might be very useful for a better understanding of the dynamic proposal. In very general terms, one may summarize the comparison as follows: the computational system, on one hand, would encompass separate modules or structures, which function as a sequence; the dynamic cognitive system, on the other hand, would comprise systems which interact among each other, integrated systems which function simultaneously. Under the computational system, examples of modules would be perception and production, which, in turn, are interactive processes in the dynamic cognitive system. The structures, in the computational system, would be the brain, and, in the dynamic system, the nervous system, the body and the environment. From the computational viewpoint, the body and the environment are subject to the control of the brain, which acts on them through its commands. From the dynamic viewpoint, the three structures are mutually integrated.

A notion which might assist us in understanding the relationship which joints the three integrated structures of the dynamic cognitive system is the following: perception, cognition and action are not separated. This is a triad which is useless to explain a given behavior, e.g. speech (Cummins, 2010). Thus, the performance of a vocalic gesture, a speech gesture, is not the product of different cognitive stages which excludes the motor action of the articulators. The vocalic gesture should rather be understood as an orchestration between higher and lower levels. In other words, the higher levels concern the vowel as an information unit which is linguistic in nature; whilst the lower levels are related to the performance of the action unit (movement of the speech articulators) over time.

To choose a cognitive approach based on the theory of dynamic systems as it applies to music implies having to uncover what musical performance can tell us about musical knowledge. Music can be performed in a number of ways, e.g. by playing an instrument, by producing percussion with one's body, by creating electronic sounds, and so on. In this paper, we shall focus on the notion of singing voice.

It is our claim that, in the case of music, one will be able to determine the existence of an action unit which is concurrently an information unit, such as the phonological gesture in Articulatory Phonology (Browman & Goldstein, 1992), which we present in an abridged form under Section V. The sung musical note would be this gesture, performed by the articulators of the larynx, and which informs specific tones within a given scaled system and within a given rhythmical structure.

III. DIFFERENT VIEWS ON MUSIC AS A DYNAMICAL SYSTEM

The works that will be summarized below focus on different aspects of music: emergence of meaning (from a processual analysis), synchronization in musical performance, rhythm perception, and musical development. Bringing these studies together considering that they posses a common idea about dynamical systems related to music has been challenging, since ideas on this matter – from a research program point of view – seems to be sparse and these studies do not refer to each other.

The idea of musical performance and meaning as dynamical systems is that the latter emerges while music is being performed and is not predetermined (Burrows, 1997). In the case of entrainment, Clayton and colleagues (2005) seek to explain the synchronization between two individuals, while singing and snapping fingers, as evidence of entrainment, which indicates an adaptive capacity of living beings, with important implications for ethnomusicology. The issue of rhythm perception related to temporal fluctuating inputs offers a dynamic model of oscillators that uncover our capability to follow a fluctuating structure as a variation of the same temporal organization and not as an interruption of a rhythm structure (Large and Palmer, 2002). Marin and Peltzer-Karpf (2009) proposes that syntax emerges in an initial phase of language development and changes should be explained in longitudinal studies, and the same could be done with musical development, including singing that arises in very early infant ages. They claim that such changes can benefit from a Dynamical System Theory perspective, since it takes the dynamic characteristics of brain and behavior.

The works of Clayton and colleagues and of Large and Palmer involve the issue of a subject adaptation to a musical temporal organization in real time. One must remember that the perceptual results obtained by Large and Palmer were originated by the listening to the performance of pianists that either deviate from the initial structure via *rubatos*, or carried out such deviation by stressing one melodic element or another. Both of these studies can be encompassed by the more general idea of rhythm.

Yet, the study of Burrows and that of Marin and Peltzer-Karpf do not present empirical data, but they assume important positions in order to think a dynamical approach for music. Burrows insists that music is performance, opposing this idea to the traditional view of music as static. He advocates in favor of an analysis that take into account that music unfolds in time and is not a "piece" that has a life on its own. Essentially, it is an encounter between acoustical events and the listener, although a mental performance is obviously also possible.

Burrows' non-static music is in tune with Marin and Peltzer-Karpf's ideas on *change* in development that can be better explained by Dynamical System Theory tenets than by those of symbolist or computational theory, in which the discrete symbols as representations of knowledge are deprived of time and movement.

IV. LOCUS OF CONVERGENCE

Is there any locus of convergence among these approaches?

It does not seem an easy convergence to achieve, since the synchronization task (Clayton and al., 2005) is apparently a low level task and Burrows (1997) adopts a view on performance which seems to assume *a priori* knowledge of music and of expectances one must have from past moments in a given score. However, both approaches refer to musical performance to address changing states and this is related, as well, to the *rubatos* and melodic saliences present in the perceptual tests conducted by Large and Palmer (2002). In respect to changing states in musical development, Marin and Peltzer-Karpf (2009) trace parallels to phases and states in language development.

One common aspect that would permeate these views could be that everything points towards a capacity of adjusting music perception and/or production while music is being performed or acquired. In this sense, we could then deal with a fundamental aspect of a dynamical system: change over time.

This notion (change over time) has been adopted by Articulatory Phonology (AP) in relation to speech production. AP relates the physical structure to the cognitive structure of speech viewing both as one and same system (Browman and Golsdstein, 1992). The dynamical unit proposed by AP is the articulatory gesture, a speech unity of stable pattern (the atom), combined to other gestures as a contrast unity, thus a phonological unit. This combination is achieved through the movement of speech articulators present in the vocal tract, as lips, jaw, tongue, velum and glottis. Articulators correspond to tract variables, as the TTCL (tongue tip constriction location). An articulatory gesture produced with tongue tip constriction coordinated to a tongue body gesture (TBCD, tongue body constriction degree), in which there is a wide space in the pharyngeal region gives us the lexical item /ti/ (form of the second person in Portuguese). The other tract variables belonging to the AP model are LP (lip protrusion), LA, (lip aperture) TTCD (tongue tip constriction degree), TBCL (tongue body constriction location, VEL (velic aperture) and GLO (glottis aperture). Constriction location indicates which vocal tract region is affected by a constriction. The constriction degree indicates if the air passage is closed or critical.

The tract variables activation for gestures coordination happens in a simultaneous way and by partial or total overlapping. The activation is represented by a gestural score, in which tract variables are presented as tiers and their activation times in boxes. One possible adaptation to a gestural score for /ti/ could be the following:

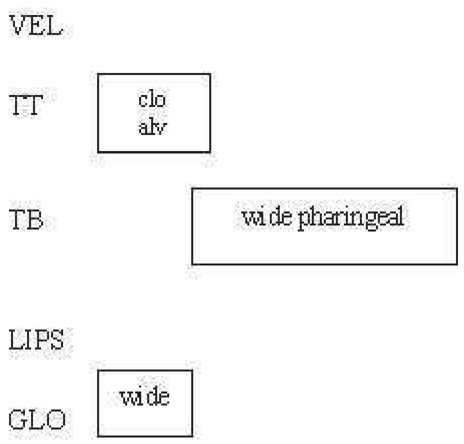

Figure 1. A hypothetical gestural score for /ti/, adapted from Browman and Goldstein (1995). Tongue body constriction location affects the pharyngeal region and its constriction degree is wide. Tongue tip constriction is maximal, closed at the alveolar region. Simultaneously; glottis is open, so it is wide.

The advantage of AP in relation to the dualist phonological theories to which speech realization is merely physical, thus with no link to a cognitive level, was to propose a unity that is abstract and, at the same time, is characterized by time and movement. It goes without saying that the preceding presentation of AP is highly simplified. A better comprehension of the theory must be sought in the seminal texts.

In order to advocate on behalf of the musical gesture as a dynamical system, it is appropriate to highlight both the way AP states its language primitive and the way it thinks the task dynamic model. In this model, the movement of the articulatory gesture is modeled by a mass-spring differential equation. The articulatory gesture as primitive leads us to understand the relationship between the physical and cognitive, while the task dynamic model provides insight on how articulatory patterns combine to create the phonological utterance.

If on one side the gestural score is part of a dynamically defined model that accounts for the phonology of a language, the musical score has no analogous role in relation to music. The musical score registers and represents musical tones and rhythm, and its purpose, as we well know, is to mediate the musician and music execution through his reading, to come to an acoustic result. Nevertheless, except in more contemporary musical notation, it says nothing about the necessary movements which lead to the actual performance of music.

V. WHAT KIND OF MOVEMENT COULD CHARACTERIZE MUSIC?

It is a common place to state that music is movement. If this notion is assumed to be right, what kind of movement is it? Instrument players would quickly respond that they have to move some part of their body to produce sound: e.g. they have to move arms and fingers in order to play the violin. Composers engaged in the production of electroacoustic music would say that a sound wave is an invisible movement through the air. Both answers are good answers, nevertheless they do not contribute to an approach which is in search of a primitive; in other words, we must know what organizes or coordinates this movement. We would like to conceive of a musical unity and its aspects related to action and information. If, as a starting point, we agree about a simple point, that a musical primitive presents specific tones and a given temporal organization, it would seem that the singing voice accounts for this musical primitive (or unity) realization/achievement.

The singing voice seems to be an interesting candidate for our musical movement and for this reason we point it out as a gesture in a broad sense. More intuitively, we would presume that the human voice can offer a good range of expressive possibilities over and above speech, e.g., crying, laughing screaming and singing. Singing, with or without words, allows one to choose specific tones, attain more volume than commonly used for speech (Sundberg, 1987), give longer duration for chosen tones, if compared to speech.

Combination of these heights in a given system (scales) would give us a singing voice. It is assumed that one will produce tones from the acquired musical scale system (e.g. the seven-tone scale, the Chinese scale, the Indian scale, etc.) in a given temporal structure. The ability to tune long or short notes in fast or slow rates, obeying a given temporal structure, gives us rhythm. Our approach on rhythm as part of a musical gesture is limited to what is already known about it in music. A periodic and basic structure underlies rhythmic phrases allowing organization and memorization of variable temporal events (Krumhansl, 2006). We propose that rhythm is coordinated along with the larynx movements and this is an issue which must be further developed in future studies.

Coordinated movements of the larynx in the production of the singing voice could be modeled by following the steps of AP in the vocal tract modeling. A fine understanding of how laryngeal muscles work is necessary, involving a knowledge of the main larynx muscles (thyroid, cricoid and arytenoids) coordination, as well as how it relates to the hyoid bone movements. Obviously, this is a time-demanding study, since the vocal folds articulatory activity is not just a muscular movement, but involves sub-glottic air pressure and vocal folds tension that must be accounted by the myolastic and aerodynamics of phonation (see Titze, 1980). In this sense, larynx modeling is quite different from vocal tract modeling which is *prima facie* based on coordinative structures or muscular assembly unrelated to aerodynamic phenomena.

In order to initially think about a possible larynx modeling for music production, we have attempted to understand the functioning of the larynx as related to vocal fold vibration. Therefore, in a non-exhaustive search, two models appeared to be fairly plausible for an initial hypothesis on the articulatory musical gesture: the Fujisaki model and the two-mass vocal-fold model, as we will see in the following section.

VI. WHAT WE CAN LEARN FROM LARYNX MODELS

Since the larynx has been determined to be a relevant part of the articulation of the singing voice, in the following we present two fold vibration models – the one by Fujisaki (1983) and the one by Lucero & Koenig (2005) – so as to provide an initial understanding of how laryngeal movements can be part of a musical system. The functioning of the vocal fold oscillation as described by Titze (1988 and 1989) must also be taken into account.

The functional model for the process of generating the F0 contour (Fujisaki 1983) explains the fundamental difference in the behavior of the vocal folds in speech and in singing. In speech, there is no specific tone which must be reached, but there is a need for a contour composed of minimum and maximum F0's (baseline) and peaks at any given moment of this contour (accent). Therefore, this baseline of the phrase, always with a falling trend, and the accent are two components of the functioning of F0 and control the vocal folds oscillation in speech. In singing, a specific tone is aimed at, to wit: the musical note. In this sense, the vibration of the vocal folds must quickly reach its goal. The drop in F0, which is normal in speech, cannot be presumed for the sung note; on the other hand, the control of the F0 transition, which must be speedy and precise in singing, need not be specifically controlled in speech. This model conceives of an abstract command (possibly the nervous system) of speech and a motor control over the muscles which generate speech, and, as such, moves away from the notion of an integration of systems. In the specific case of singing, although the model does not explicitly state which component controls the production of tone, one may presume that this component is the musical note. Its relevance relates to the difference in behavior in the oscillation of the folds in singing as compared to speech.

The two-mass model of the vocal folds (Lucero and Koenig 2005a and 2005b) is a model in which each vocal fold is considered a mass-spring system, based on a widespread two-mass model (Ishizaka and Flanagan, 1972). The Lucero & Koenig (2005a and 2005b) model, which was developed based on air flow data from the speech of children, women and men, explains the vibratory movement of the vocal folds by means of movement equations which describe two mass-spring systems (each one of the vocal folds) coupled by a spring. This system relates to the aerodynamics of the glottis and of the vocal tract. Generally speaking, the larger the larynx the easier the vibratory movement.

A detailed description of the actual system for the functioning of the larynx can be found in the body-cover hypothesis, which is based on the physical principles of the vibration of the vocal fold, including the relationship between the displacement of air and the movement of the laryngeal tissue (Titze, 1988). The movements of the mucosa (cover) and of the muscles (body) involved in the vibration of the glottis are basis/c notions in order to understand this vibration. Note, however, that phonation modes which differ greatly from the speech modes involve specific articulations in muscles such as the thyroid and the arytenoid, the movements of which are linked to a rise or a drop in F0. Probably, in singing with an acute voice, the activity of the thyroarytenoid muscle is not the most important component, since its activity is significantly reduced with a high F0 (Titze et al., 2005).

We may learn from the models that a simplification must take into account, first of all, the specific behavior of the larynx when producing sustained musical notes. This behavior differs from normal speech, which, whether in longer utterances, in phrases or in words, has a falling F0 component (baseline). To take into account this basic difference between speech and singing provides a good foundation for the argument that an adjustment for the vibration of the vocal folds is a integrate part of the musical gesture, since the attainment of the intended tone is part of the musical information.

Secondly, as to the mechanical functioning of the larynx, one must take into account the size and the mass of the vocal folds; the relation between size and mass, on one hand, and subglottic pressure; and, perhaps differently for voice in singing, the tension of the folds, as well as the determination of which muscles are involved when very high F0 is reached.

Here, the configuration of the vocal tract is also highly important, so as to ensure a form of continuity of the laryngeal tube (Raposo de Medeiros, 2002; Lucero et al, 2011), and, consequently, the ongoing sustain of the phonation frequency. But this is an issue we will not be dealing with at this point.

VII. CONCLUSION

The course of this work, however brief, has been committed to the idea of a musical gesture, raising the common link among different views on music as a dynamical system. The four views presented here converge on the idea of music as a dynamical system because – either in performance, perception or acquisition – they all deal with the states system relation. One may plausibly state that such an approach is feasible, provided one starts from the premise that cognitive musical units are not static units. In order to approximate this aspect to an initial notion of a musical gesture, we have taken Articulatory Phonology as starting point (AP). Thus, we have proposed that, as the articulatory gesture is at the same time a unit of action and of information, so could the musical gesture be.

Therefore we assume that the musical gesture can be defined as unit of action since it contains movement generated by the larynx functioning in the production of the singing voice. At the same time, this gesture is a unit of information because the coordination of movements and the combination of gestures will concurrently happen within a scale system and a within a duration system, coordinated according to a rhythmic structure.

Admittedly, our proposal might seem somewhat speculative in this initial state. However, as already pointed out, studies adopting a dynamical approach to music are sparse and do not posit principles that set music as a dynamical system in stricter terms. This weakness must be dealt with by means of a deeper discussion about the musical gesture and its suitability to all dynamical views on music. If the musical gesture can be accepted as a plausible contribution within a dynamic theoretical framework, it will be worthwhile to undertake studies aimed at modeling the functioning of the larynx, based on existing models. This would contribute to the understanding of movement required to reach the acoustic output of music, naturally based on the principle that all primitive music can be expressed by singing.

With regard to temporal organization of different tones, this does not seem particularly difficult to represent: it would suffice to consider them as part of a hierarchical metrical structure. Such a structure is inherent to music, i.e., there will always be a metric structure known or established a priori.

What has impelled us to sketch this initial proposal involving a musical gesture is the fact that all music which floats through the air, as an acoustic phenomenon, originates from an organized and coordinated initial movement.

ACKNOWLEDGMENT

My thanks to Francis Aubert for textual revision. This work has been supported by a FAPESP grant 2011/14435-3.

REFERENCES

Browman. C. and Goldstein, L. (1992) Articulatory Phonology: an Overview. *Phonetica, 49*, 155-180.

Browman. C. and Goldstein, L. (1995) Dynamics and Articulatory Phonology. In R. Port, and T. Van Gelder, *Mind as Motion* (pp 175-193). Cambridge: The MIT Press.

Cummins, F. (2010). Coordination, not control, is central to movement. In Esposito, A. M., Martone, R., Müller, V. C., and Scarpetta, G. (Eds.), *Towards Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues*, volume 6456 of *Lecture Notes in Computer Science (LNCS)* (pp 252-264). Springer.

Burrow, D. (1997). A dynamical perspective on music. *The Journal of M usicology*, 15 (4), 529-545.

Clayton, M.; Sager, R. and Will, U. (2005). In time with the music: the concept of entrainment and its significance for ethnomusicology. *ESEM CounterPoint*, 11.

Fujisaki, H. (1983) Dynamic characteristics of voice fundamental frequency in speech and singing. In P. F. MacNeilage (org.) The production of speech. NY: Springer, p.39-55.

Hochman, M. Dynamical system theory. What in the world is it? Retrieved from http://math.huji.ac.il/~mhochman/research-expo.html.

Ishizaka, K. and Flanagan, J.L. (1972) Synthesis of voiced sounds from a two-mass model of the vocal folds. Bell Syst. Tech J., 51, 1233-1268.

Krumhansl, C. L. (2006) Ritmo e altura na cognição musical. In Ilari, B. S. (org.) Em busca da mente musical (pp 45 a 109). Curitiba. Editora da UFPR

Large, E. W. and Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26, 1-37.

Lucero, J. and Koenig, L. (2005a) Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control. J.Acoust. Soc. Am. 117 (3). 1362-1372.

Lucero, J. and Koenig, L. (2005b) Phonation thresholds as a function of laryngeal size in a two-mass model of the vocal folds (L). J.Acoust. Soc. Am., 118 (5), 2798-2801.

Lucero, J., Lourenço, K. G., Hermant, N., Van Hirtum, A., Pelorson, X. (2011) Vocal folds vibrations during phonation: effect of acoustical coupling. *Proceedings of the 18th International Congress on Sound and Vibration*, 10–14 July 2011, Rio de Janeiro, Brazil. 1-8.

Marin, M. M. and Peltzer-Karpf, A. (2009). Towards a Dynamic Systems Approach to the Development of Language and Music: Theoretical Foundations and Methodological Issues. *Proceedings of the 7th Triennal Conference of European Society for the Cognitive Sciences of Music* (ESCOM 2009), Finland. Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors). 284-292.

Norton, A. (1995) Dynamics: an introduction.In R. Port, and T. Van Gelder, *Mind as Motion* (pp 45-68). Cambridge: The MIT Press.

Port, R. and Van Gelder, T. (1995) *Mind as Motion*. The MIT Press. Cambridge.

Raposo de Medeiros, B. (2002) Descrição comparativa de aspectos fonético-acústicos selecionados da fala e do canto em português brasileiro. PhD dissertation. University of Campinas, Campinas, Brazil.

Sundberg, J. (1987) *The science of the singing voice*. Northern Illinois University Press, Dekalb.

Titze I. R. (1980) Comments on the Myoelastic - Aerodynamic Theory of Phonation. *Journal of Speech and Hearing Research*, 23, 495-510.

Titze I. R., Luschei E.S., Hirano, M.(1989) Role of the thyroarytenoid muscle in regulation of fundamental frequency. *Journal of Voice* (3, 3). 213-224. http://dx.doi.org/10.1016/S0892-1997(89)80003-7