Finding Rhythm in Speech: A Response to Cummins

ROGER K. MOORE
*Speech and Hearing Research Group, Department of Computer Science,
University of Sheffield*

ABSTRACT: This paper attempts to address three critical questions left unanswered by Cummins' review: are rhythm and entrainment physical, perceptual or social phenomena, what are the underlying mechanisms, and what is their role in behaviour such as speech and music? These issues are addressed from the perspective of an engineer/computer-scientist/ roboticist for whom modelling such behaviours within a computational framework not only provides an empirical methodology for validating theoretical claims, but also facilitates the construction of artificial devices that are capable of exhibiting/exploiting those behaviours in the context of human-machine interaction. The paper draws on insights from a range of different perspectives, and attempts to weave them together within a coherent theoretical framework. It is concluded that (i) rhythm and entrainment are phenomena that emerge naturally from the structural coupling within and between even simple systems, (ii) living systems have evolved very effective mechanisms for managing such behaviours for intrinsic and extrinsic gains, and (iii) the fields of energetics and information theory provide the appropriate tools for analysing and characterising such behaviour within a general theoretical framework. It is hoped that these insights will inspire future cross-disciplinary research in these areas, and lead to a deeper understanding of these fundamental behaviours.

ONE of the many pleasures (and difficulties) of working in the field of speech research is that it is a much more diverse area than it might at first appear. It turns out that speech is not a unitary behaviour that is entirely independent of the other physical and mental activities engaged in by the human organism; rather it seems to represent the confluence of almost all other human behaviour into a powerful and sophisticated set of mechanisms for coordinating the activity and behaviour of groups of individuals. It provides the 'glue' that allows communities to survive in circumstances that an individual might not, and it is becoming increasingly clear that an understanding of how it functions (and the role that it plays in human interaction) appears to require the integration of knowledge and insights from a wide range of different disciplines across science, engineering and the humanities. It is probably for these reasons that, after over 100 years of pursuing the classic reductionist approach to studying speech, we seem to be lacking any real understanding of some of its most basic properties. Whilst we have rather detailed models of quite sophisticated phonetic and linguistic structures, more fundamental behaviours such as rhythm, timing and prosody are considerably less well understood.

It is into this arena that Cummins steps (Cummins, this issue), first surveying some of the early empirical work on rhythm, before retracing the "sad history" of the 'great isochrony safari' and then returning to the notion of rhythm and entrainment as a whole-body activity functioning as a social affordance (Cummins, 2009). Cummins' review is an eloquent, entertaining and informative canter through some important historical events, but it leaves unanswered three critical questions pertaining to rhythm and entrainment:

- Are the behaviours physical, perceptual or social phenomena?
- What are the underlying mechanisms?
- What is their role?

This paper attempts to address these issues from the perspective of an engineer/computer-scientist/roboticist for whom modelling such behaviours within a computational framework not only provides an empirical methodology for validating theoretical claims, but also facilitates the construction of artificial devices that are capable of exhibiting/exploiting those behaviours in the context of human-machine interaction. From this perspective, the scientific issues ultimately have to

be grounded in functional mechanisms that are capable of implementation in physical artefacts that interact with real people, either for practical applications or for proof-of-principle for hypothetical models of human behaviour.

TYPE OF PHENOMENA

As Cummins points out, in order to be able to discuss the presence/absence of behaviours such as rhythm and entrainment, it is first necessary to define precisely what phenomena such terms are intended to capture, and what empirical measures can be invoked to quantify such effects in observable data. Dictionary definitions of rhythm refer to *repeated patterns* or recurring sequences, and entrainment is usually defined as a *synchronisation* of behaviours between interacting systems. Such definitions lend themselves to mathematical interpretation, and the engineering field of signal processing has much to say about how such phenomena can/should be measured in physical systems.

### *Rhythm and entrainment as physical phenomena*

Rhythm and entrainment are essentially behaviours that emerge from the mechanical properties of the structures that constitute the physical world and the interactions that arise between them. As such, the emergence of rhythmic and synchronised patterns is not peculiar to living systems (c.f. the regular atomic structure in crystals, wave patterns on a body of water, or the orbital trajectories of planets). Such patterning results from the tendency for physical systems to fall into low-energy/low-entropy states (known as 'attractors'), thereby giving rise to highly organised static configurations or, if there is elasticity present, to oscillatory behaviour around such configurations. Since living systems are also physical systems, they too exhibit such behaviour, resulting either from the biomechanical properties of their bodies or from the neurophysiological properties of their nervous systems.

However, what distinguishes living systems from non-living systems is that the former possess the unique ability to exploit such behaviours for evolutionary survival purposes; hence 'natural' rhythms and synchronicities can become actively managed by a living organism to support intrinsic behaviour within itself (such as pumping blood around the circulatory system or sending impulses around the nervous system) or extrinsic behaviour between itself and its external environment (such as moving around or ingesting nutrients). The latter (the extrinsic relationship between an organism and its environment) is particularly interesting when the environment contains other living organisms. In this case, interactional behaviour between different individual organisms can be manipulated for the purposes of competition (e.g. for managing predator/prey relationships) or for the purposes of cooperation (e.g. for coordinating with conspecifics) (Noble, 2000).

These principles apply to the behaviour of any living organism whether it is the synchronised movements of the flagella on a single-celled microbe or a Mexican wave performed by a group of multi-cellular human beings. In particular, the notion that 'natural' (i.e. low energy/entropy) behaviours can be modulated for communicative purposes is a powerful predictor of the types of behaviour that are likely to emerge from given biomechanical and neurophysiological systems. For example, in the case of speech, the prime candidate for the source of rhythmic structure is the biomechanical properties of the various components of the vocal apparatus, especially the diaphragm, lungs, larynx, mandible, tongue and vocal tract. As a direct consequence of their particular volumes, masses and elasticities, each of these physical structures has a 'natural' frequency of oscillatory behaviour, each can be modulated for communicative purposes, and each has the potential to be synchronised with the behaviour of another individual or group of individuals. In addition, the high number of degrees-of-freedom (DoF) available from the low-mass components of the system means that it can support a remarkably high information rate with a minimum of articulatory effort (and with the minimum of interference with other bodily functions). Such, I would claim, is the evolutionary pre-eminence of spoken, as opposed to gestural, language (Tomasello, 2008). Likewise, the high DoF means that it is unlikely that there is one globally optimum solution to managing the behaviour; hence the emergence of many related spoken languages (rather than one universal spoken language) where each language represents a low energy/entropy solution in the epigenetic landscape of physical and neural behaviours.

Supporting evidence for this position comes from applying standard general-purpose signal processing techniques to the analysis of speech signals. For example, it is well established that the *modulation spectrum* of a speech signal and the modulation transfer function of a particular environment (such as a reverberant room) can be used to predict the intelligibility of a speech signal in that environment to a high degree of accuracy (Steeneken & Houtgast, 1980). It turns out that most of the information in speech is in the sub-16 Hz frequency region, and that information for speech recognition peaks around 4 Hz - all of which is predictable from the properties of the physical systems involved (as argued above).

As an example, Todd and Brown (1996) published a robust algorithm for recovering rhythmic grouping in auditory signals based on multi-scale analysis (equivalent to computing the modulation spectrum). The output could be visualised in the form of a 'rhythmogram', and Figure 1 shows the outputs for both a speech utterance and for a piece of music. As can be seen, rhythmic structures are obtained that reflect the metrical properties of the corresponding input signals.
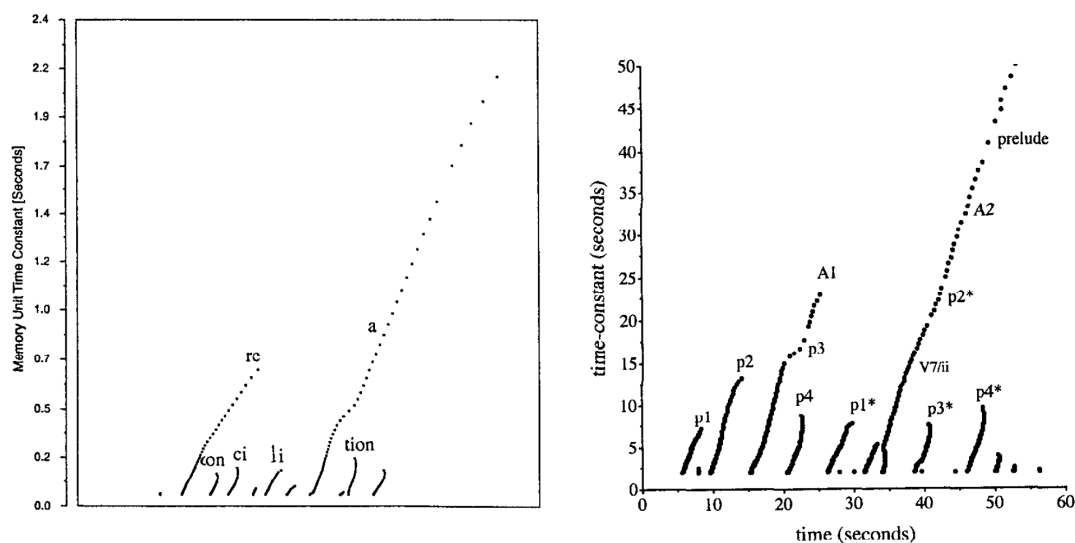
Fig. 1: Rhythmograms for (a) the utterance "reconciliation" and (b) Chopin's Prelude Op. 28 No. 7 (taken from Todd & Brown, 2009).

Given the above, it seems surprising that researchers engaged in Cummins' "sad story" appear to have ignored (or at least have been unaware of) such analysis procedures, and have severely limited themselves by searching for measures based on the precise timing of particular events rather than by characterising the behaviour in a more robust manner using the appropriate general-purpose signal processing transformations.

The existence of particular attractors in the state space of available behaviours is supported by Cummins' own work on rhythm and rhythmic constraints (Cummins & Port, 1998). Again, the hypothesis is that a biomechanical system will inevitably exhibit preferred states and behaviours, and that this can be usefully characterised using dynamical systems theory (Luenberger, 1979). Unfortunately, the dynamical behaviour of most living systems (and especially speech) is far too complex to be easily characterised mathematically.

Rhythm and entrainment as perceptual phenomena

Whilst the arguments presented above for rhythm and entrainment as emergent physical phenomena are compelling, there is a natural tendency to assume that such phenomena are manifest as surface behaviours which are open to straightforward characterisation or measurement. Another possibility is that such phenomena are perceptual constructs (i.e. epiphenomena) taken from the perspective of an 'intelligent' observer. Some of the latest theories of action understanding invoke significant neurophysiological overlap between sensory and motor systems such that an observer's own motor planning system is implicated in the interpretation of observed behaviour in others through a mechanism of emulation and synchronisation (Grush, 2004). Known as 'mirror neuron' (MN) systems (Rizzolatti & Craighero, 2004), the principle is that an organism is able to interpret the intentional behaviour of another organism by hypothesising intentional states using generative structures recruited from its own motor abilities, and such processes are thought to be involved in the evolution of speech as a communicative behaviour (Stamenov & Gallese, 2002; Studdert-Kennedy, 2002; Corballis, 2010).

A consequence of the mirror neuron hypothesis is that entrainment can be interpreted as an inevitable outcome of a recognition/understanding process whereby the observer becomes synchronised with the performer at a deep *intentional* (rather than a surface *realisational*) level. This means that a strong subjective sense of rhythm and entrainment could emerge in an observer despite it being virtually invisible in the associated complex surface behaviour.

Rhythm and entrainment as social phenomena

As has already been pointed out, rhythmic and synchronised behaviours are exhibited by many physical systems - both living and non-living - but those that occur between similar structures are especially interesting since the repetition of structure provides a strong prior for the emergent patterning. From the crystalline structures in a gemstone to the swarm behaviours in insects and birds, the same underlying principles are operating. For living systems it turns out that, not only does coordinated behaviour between groups of individuals deliver significant evolutionary benefits, but conspecifics are likely to have the best internal models of each others' behaviour (e.g. through access to models of their own behaviours achieved by exploiting MN-like mechanisms) (Wilson & Knoblich, 2005). This implies that, especially for highly evolved animals such as human beings, most social interaction is entrained in some sense (as evidenced by behaviours such as the 'Chameleon effect' (Chartrand & Bargh, 1999) in which one individual subconsciously mimics the behaviour of another), and that joint co-active behaviours will inevitably exhibit some level of rhythmic structure (Hagen & Bryant, 2002).

Such effects are well known in conversational interaction (Benuš, 2009; Clark, 2002; Wilson & Wilson, 2005) where participants often accommodate to each other's speaking rate and style, or even complete one another's utterances. However, it is interesting to note that it appears that only vocal learning species can synchronise to a beat (Patel et al, 2009a; Patel et al, 2009b), thus suggesting that *cross-modal entrainment* is a very recent behaviour from an evolutionary perspective – a fact that might explain the special role that music plays in human endeavours (Fitch, 2006; Patel, 2008).

UNDERLYING MECHANISMS

Given the three types of phenomena identified above - physical, perceptual and social - it is natural for an engineer/computer-scientist/roboticist to speculate about the mechanisms that might underlie the consequent emergent behaviours and the associated mathematical/computational models that could be used to investigate or replicate those same behaviours. At the most basic level, intrinsic (periodic) behaviours or extrinsic (synchronised) behaviours arise from *coupling* between the intrinsic or extrinsic components (Assisi et al, 2005; Canolty & Knight, 2010). For example, in a physical system, coupling in the form of a spring results in the transfer of energy from one location to another. In more general terms, coupling represents the transfer of information from one component to another. Hence the behaviour of each component can be said to be correlated with or conditioned upon the behaviour of another, and this can be expressed mathematically using mutual information [1] and/or conditional probabilities. Systems or system components that are not coupled are, by definition, independent of each other (hence any observed periodicities or synchronicities are coincidental rather than causal).

The fact that the world and the organisms within it actually exists, shows that coupling has created a high degree of organisation along many dimensions and at many scales. Both the persistence of matter and the survival of organisms emerge from the myriad of structural dependencies between everything and everything else. Truly random behaviour is rather rare, hence periodicity-within and synchronicity-between systems is, in practice, the norm.

The notion that complex behaviour can arise from simple local coupling is, of course, not new. It is the premise behind the entire 'connectionist' approach to modelling neural behaviour (Rumelhart & McClelland, 1986), and is also the basis for fields such as agent-based modelling.

Coordination

It is in the context of the foregoing that living organisms are able to exploit both intrinsic and extrinsic coupling for their own advantage, e.g. for coordinating group activity. This is true whether we are talking about a collection of cells or an army of soldiers. The complexity of the resulting group behaviour is contingent on the degree of information flow facilitated by the coupling processes and, in the case of spoken language, human beings appear to have evolved a coupling system with a remarkably high information rate - ~100 bits per second. This appears to be well above that exhibited by any other animal communication system (e.g. the pheromone-based chemical communication medium employed by ant communities can probably only support an information rate amounting to fractions of a bit per second).

Interestingly, in most non-living systems, intrinsic and extrinsic coupling is a mutually-dependent two-way process, whereas in living systems extrinsic coupling can be one-way. Organisms literally *manipulate* each other in order to get them to do what they want – a form of entrainment that can be interpreted socially in terms of the establishment and maintenance of dominance relations.

Mimicry

One particularly interesting form of one-way coupling is the copying of one agent's behaviour by another. Whilst imitative behaviour is straightforward using an electronic device such as a tape recorder (where input and output have the same representations), it is an interesting challenge for a living system since it would seem to require a transformation between the representations derived from information arriving at its sensory apparatus and the representations used to drive its motor behaviour. In fact, all an imitator needs to be able to do is to match the sensory consequences of its own output with those of the individual it is attempting to copy (i.e. using a control feedback mechanism). The issue then becomes one of *memory* – the patterns to be copied need to be retained long enough for them to be replayed at a suitable moment – and memory is the first step towards learning. Hence it is possible to construct a link between entrainment and the transfer of skills from one individual to another, either ontogenetically or phylogenetically (Aboitiz et al, 2005).

Synchrony/co-action

The nature of mimicry is that there is some time interval between reception and repetition. Such behaviour is thus coupled, but not synchronised. For behaviours between individuals to be synchronised (i.e. coincident), it is necessary to invoke a control-feedback mechanism that is *predictive*. For simple periodic behaviours, this can be achieved by *phase-locking*. For example, Figure 2 illustrates entrainment between a robot and a speaker, where the robot's goal is to tap in time with the speaker's utterances (Moore, 2007). A simple control loop within the robot begins to phase lock with the peaks in energy in the speaker's voice such that a synchronised rhythm begins to emerge by the third utterance. By the eighth and ninth utterances the synchronisation is well established and, since the ninth utterance is the last, the robot emits one final tap at the predicted time at which the next word would have been expected. This example is very much in line with other work by Cummins (Cummins, 2009a).
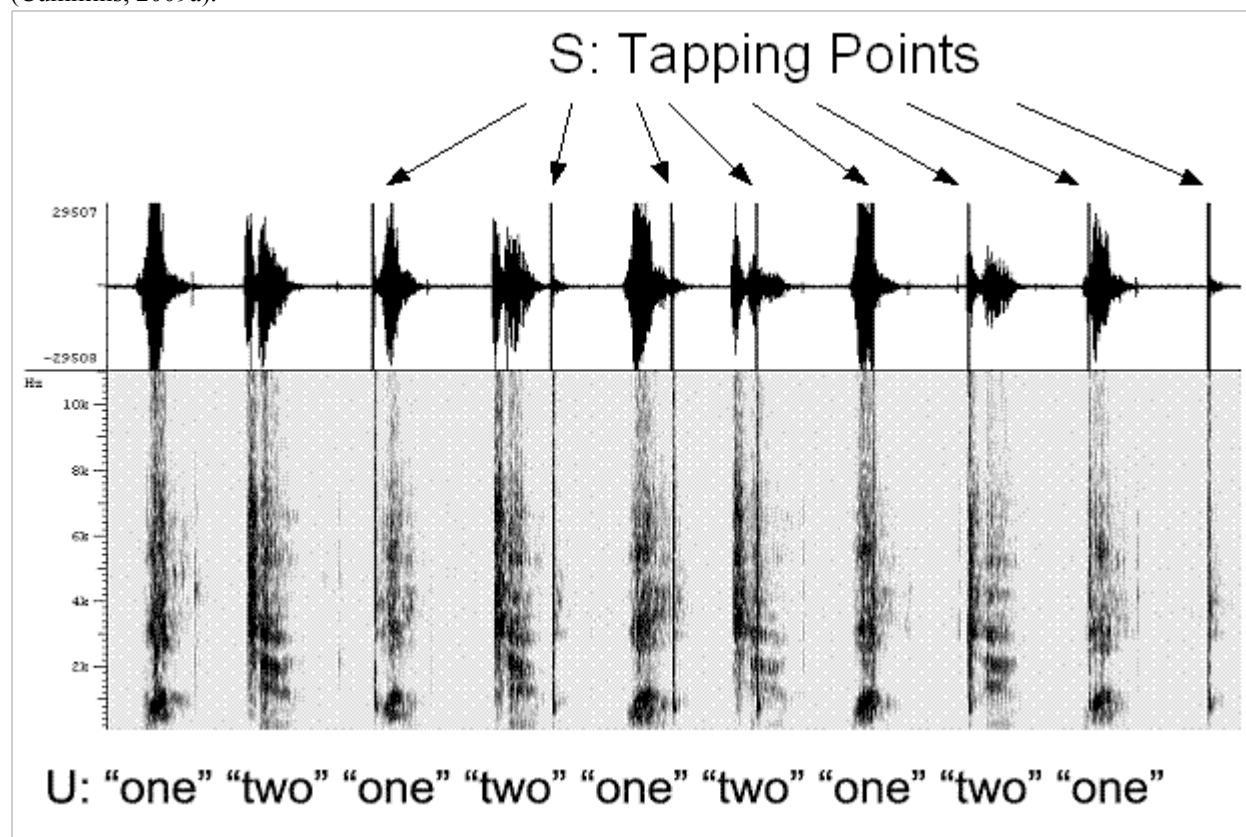

Fig. 2: Illustration of simple entrainment of a tapping robot (S) with spoken input (U).

In more complex behaviours, the most powerful predictor is memory (i.e. a record of similar past events). By *aligning* behaviour played out from memory with ongoing sensory input, it is possible for an organism to synchronise the two. It is this mechanism that is likely to underlie the ability of listeners to shadow a speaker with a surprisingly small delay (Marslen-Wilson, 1973; Cummins,

2009b) – the listener organises their motor planning based on a prediction of what the speaker is most likely to say and when they are going to say it. In this case, the behaviour is entrained, but there is no strong rhythmic component – the periodicities are related to the deeper structural correlations between speech patterns, rather than to the simple repetition of surface behaviour.

THE ROLE OF RHYTHM AND ENTRAINMENT

Most of the foregoing begs the questions as to why motor behaviour in general, and speech and music in particular, exhibit rhythmic properties at all. Yes there are physical, perceptual and social constraints that facilitate such behaviour (as described above), but what is the main driving force behind all of this? Perhaps the clearest view of what is actually going on, and the role that rhythm and entrainment play in organising behaviour within and between living organisms, is to be found by taking an information theoretic approach (hinted at above). Rhythmicity implies predictability, and this means that periodic behaviour contains less information than aperiodic (less predictable) behaviour. This simple fact has deep implications for both performers (for what roboticists call 'action selection') and observers (for 'action understanding').

Planning and organising behaviour has a complexity that is directly related to the number of degrees of freedom in the system – having 100 legs might offer an insect advantages in terms of stability over a rough terrain, but controlling 100 articulators independently requires a very complex, high energy consuming (and high information rate) control system. There are great efficiencies to be gained by coupling articulators, and this is equivalent to the establishment of a control *subspace* with fewer control variables (and hence a lower information rate). Synchrony and repetition emerge as consequences of the lower rate of information flow.

The consequence for an observer of a lower information rate performance is that such observed behaviour would be easier to predict. This would be especially valuable to an observer if it was a predator and the performer is the prey. However, in a more cooperative situation (e.g. between conspecifics), the more predictable the behaviour, the less information is transferred from one individual to another. Hence, for communicative purposes, it is necessary to increase the information rate, and this means deviating from the predictable – exactly as Cummins points out in the last sentence of his paper.

This latter point is crucial; as living systems, performers have active control over the degree of effort they put into increasing the amount of information in their behaviour. Whether it is swimming faster to escape a predator, speaking more clearly to be better understood by a foreigner, adding rhetorical effect to a speech, or giving emphasis and meaning to a musical performance – the same general information theoretic principles are in play. In fact, it turns out that Lindblom's well known 'H&H' (hyper/hypo) theory of speech (Lindblom, 1990) turns out to be a special case of a much more general principle.

These concepts are illustrated in Table 1 using a range of contrasting terms taken from different fields, starting with MacNeilage's concept of 'frame' and 'content' (MacNeilage, 1998).

Periodic	*Aperiodic*
Frame	*Content*
Carrier	*Modulation*
Predictable	*Less predictable*
Non-information bearing	*Information bearing*
Non-communicative	*Communicative*
Meaningless	*Meaningful*

Table 1: Dimensions of contrast between periodic and aperiodic behaviours.

Finally, one interesting consequence of entrainment between different systems is that when there is full entrainment, it is hard for an external observer to determine how many contributing sources are actually present. In other words, a fully-coupled system can be viewed as a single organism (e.g. multi-cellular organism, an ant colony, or a male-voice choir). In these circumstances an observer finds it harder to ascribe individual intentionality, and instead may perceive the behaviour of the collective as an intentional being in its own right. Interestingly, this behaviour can be seen being used for artistic effect in both Ugandan xylophone music and Inuit throat singing, where the listener finds it very hard to perceive that more than one individual is involved.

SUMMARY AND CONCLUSION

Inspired by Cummins (this issue), this paper has attempted to broaden the topics he raises by attempting to link with some of the deeper insights emerging from the engineering/computer-science/robotics fields. The paper has distinguished between intrinsic intra-organism behavioural correlations (rhythm) from extrinsic inter-organism behavioural correlations (entrainment), and has surveyed the physical, perceptual and social phenomena involved. Consideration has been given to putative underlying mechanisms, with special emphasis being given to the emergent consequences of coupling within and between systems. Finally, the role of rhythm and entrainment has been assessed from an information theoretic perspective, and it has been suggested that many key behaviours in living systems may be founded on such principles.

In summary, the analysis suggest that (i) rhythm and entrainment are phenomena that emerge naturally from the structural coupling within and between even simple systems, (ii) living systems have evolved very effective mechanisms for managing such behaviours for intrinsic and extrinsic gains, and (iii) the fields of energetics and information theory provide the appropriate tools for analysing and characterising such behaviour within a general theoretical framework.

Overall, the author has taken the liberty of running free reign over a range of fairly fundamental issues in order to avoid the narrow conclusions that would emerge from a more traditional reductionist approach. It is hoped that some of the ideas in the paper will liberate future researchers to adopt a cross-disciplinary perspective from which will emerge a trans-theoretical framework for modelling behaviour in living systems which is founded on sound mathematical and information theoretic principles, and which is capable of encompassing more mundane behaviours exhibited by simple creatures as well the more sophisticated behaviours (such as speech and music) exhibited by human beings.

END NOTES

[1] Mutual Information is a powerful and yet simple mathematical expression of the dependencies between variables. It is widely used in science and engineering, and is the basis of many learning algorithms.

REFERENCES

Aboitiz, F., Garcia, R., Brunetti, E., & Bosman, C. (2005). Imitation and memory in language origins. *Neural Networks*, Vol.18, pp. 1357.

Assisi, C. G., Jirsa, V. K., & Kelso, J. A. S. (2005). Dynamics of multifrequency coordination using parametric driving: theory and experiment. *Biological Cybernetics*, Vol. 93, pp. 6-21.

Benuš, Š. (2009). Are we 'in sync': Turn-taking in collaborative dialogues, *INTERSPEECH*. Brighton, UK.

Canolty, R. T., & Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, Vol. 14, No. 11, pp. 506-515.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, Vol. 76, pp. 893-910.

Clark, H. H. (2002). Speaking in time. *Speech Communication*, Vol. 36, No. 1-2, pp. 5-13.

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain and Language*, Vol. 112, No. 1, pp. 25-35.

Cummins, F., & Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, Vol. 11, pp. 51-62.

Cummins, F. (2009a). Rhythm as an affordance for the entrainment of movement. *Phonetica*, Vol. 6, No. 1-2, pp. 15-28.

Cummins, F. (2009b). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, Vol. 37, No. 1, pp. 16-28.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, Vol. 100, pp. 173-215.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, Vol. 27, pp. 377-442.

Hagen, E. H., & Bryant, G. A. (2002). Music and dance as a coalition signaling system. *Human Nature*, Vol. 14, No. 1, pp. 21-51.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling*. Dodrecht: Kluwer Academic Publishers, pp. 403-439.

Luenberger, D. G. (1979). *Introduction to Dynamic Systems*. New York: Wiley.

MacNeilage, P. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, Vol. 21, pp. 499-546.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, Vol. 244, No. 5417, pp. 522-523.

Moore, R. K. (2007). PRESENCE: A human-inspired architecture for speech-based human-machine interaction. *IEEE Trans. Computers*, Vol. 56, No. 9, pp. 1176-1188.

Noble, J. (2000). Cooperation, competition and the evolution of prelinguistic communication. In C. Knight & M. Studdert-Kennedy & J. Hurford (Eds.), *The Emergence of Language.* Cambridge: Cambridge University Press, pp. 40-61.

Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford: Oxford University Press.

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009a). Studying synchronization to a musical beat in nonhuman animals. *The Neurosciences and Music III—Disorders and Plasticity: Ann. N.Y. Acad. Sci.*, Vol. 1169, pp. 459-469.

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009b). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, Vol. 19, No. 10, pp. 827-830.

Powers, W. T. (1973). *Behavior: The Control of Perception*. NY: Aldine: Hawthorne.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, Vol. 27, pp. 169-192.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing: Exploration in the microstructure of cognition*. Cambridge, MA: The MIT Press.

Stamenov, M. I., & Gallese, V. (2002). *Mirror Neurons and the Evolution of Brain and Language*. Philadelphia: Benjamins.

Steeneken, H. J. M., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *Journal of the Acoustical Society of America*, Vol. 67, pp. 318-326.

Studdert-Kennedy, M. (2002). Mirror neurons, vocal imitation, and the evolution of particulate speech. In M. I. Stamenov & V. Gallese (Eds.), *Mirror Neurons and the Evolution of Brain and Language*. Philadelphia: Benjamins, pp. 207–227.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Todd, N. P. M., & Brown, G. J. (1996). Visualization of rhythm, time and metre. *Artificial Intelligence Review*, Vol. 10, pp. 253-273.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, Vol. 131, No. 3, pp. 460-473.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review*, Vol. 12, No. 6, pp. 957-968.