

## Entraining the Brain: Applications to Language Research and Links to Musical Entrainment

USHA GOSWAMI [1]

*Centre for Neuroscience in Education, University of Cambridge*

**ABSTRACT:** Clayton's paper provides a clear and accessible summary of the significance of entrainment for music making, and for human behaviour in general. He notes the central role of metrical structure in musical entrainment, the possible role of oscillatory neural activity, and the core notion of phase alignment. Here I show how these same factors are central to speech processing by the human brain. I argue that entrainment to metrical structure is core to linguistic as well as musical human behaviour. I illustrate this view using entrainment data from developmental dyslexia. The core role of entrainment in efficient speech processing suggests that language difficulties in childhood may benefit from music-based remediation that focuses on multi-modal rhythmic entrainment. Alignment of linguistic and musical metrical structure seems likely to be fundamental to successful remediation.

Submitted 2012 January 6; accepted 2012 July 13.

**KEYWORDS:** *entrainment, amplitude envelope, oscillations, rise time*

MANY aspects of human behaviour are rhythmic. Clayton highlights some of them: the sleep/wake cycle, the beating of the heart, tapping a foot to the beat of music, and musicians playing together in groups. Further, human motor behaviour tends to be rhythmic. As noted by Cummins (this volume), any cyclic movement of the limbs entails a rhythm, be it walking, dancing, juggling, breathing or speaking. Although speaking may not at first sight appear to be a rhythmic activity, constraints linked to how we produce language in fact impose metrical constraints. The temporal relationships between the sequencing of syllables, their organisation into larger linguistic units (foot, phrase) and their positions (relative timing) with respect to each other cannot vary freely, because the articulators can only shape speech sounds in certain ways. Here I will argue that the perception of speech also depends on rhythmic and metrical constraints, hence sharing marked similarities with musical perception and musical behaviour as discussed by Clayton.

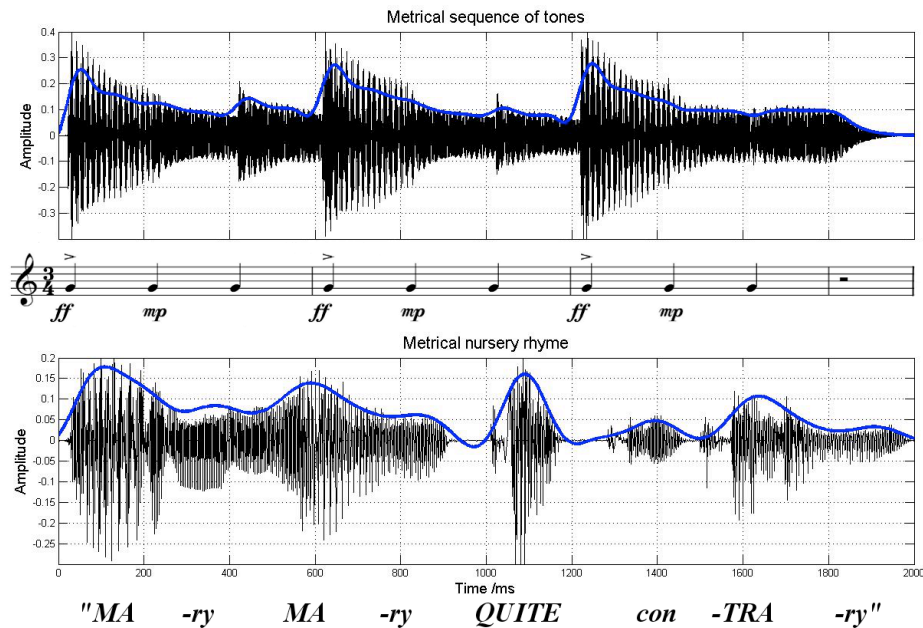
Musical entrainment is multi-modal, and Clayton discusses the role of visual (e.g. body sway) and motor (e.g., hand gesture) cues as well as auditory cues to successful entrainment. The important rhythms for speech processing are also multimodal, for example we perceive both auditory and visual rhythm and we create linguistic rhythms with our articulators, hence entrainment to speech is also multimodal. Configuring the vocal tract to shape the acoustic signal simultaneously deforms the face, and speech acoustics can be estimated reliably from face motion or "visual prosody" (e.g., Munhall et al., 2002; Yehia, Kuratate & Vatikiotis-Bateson, 2002). This is also the case for musical processing, as visual motion cues will be linked systematically to how the music is produced (indeed, Clayton gives the example of a tanpura player unconsciously entraining to her teacher's body motion).

Auditory prosody, which is related to speech rhythm, is critical for speech segmentation and depends on the accurate perception of syllable structure, which is reflected acoustically by the speech envelope. Recent developments in auditory neuroscience demonstrate the role of neural oscillations in entraining to the syllable structure of speech, with visual and auditory low frequency phase alignment playing a critical role in perception (Luo & Poeppel, 2007; Luo, Lu & Poeppel, 2010). Efficient resetting of the phase of oscillations within audiovisual neural networks appears to be of primary importance for efficient speech processing. Hence the themes explored in Clayton's article are of central importance to language as well as to music.

### THE SPEECH ENVELOPE AND AUDITORY RHYTHM

Auditory rhythmic entrainment to syllable structure depends in part on accurate perception of the speech envelope, which is the energy-varying waveform that the ear receives (amplitude variation over

time, Houtgast & Steeneken, 1985). As described clearly in a recent review by Giraud and Poeppel (2012), rhythmic structure is given by regular modulations of signal energy over time, which for speech peak at a rate of 4 – 6 Hz, the “syllable rate” (Drullman, 2006; Elliott & Theunissen, 2009). The onsets of successive modulations in the amplitude envelope and their rates of change (rise times) are critical linguistic perceptual events, as they signal the onset of new syllables. The rise times of stressed syllables are larger than the rise times of unstressed syllables, thereby providing a cue to linguistic metrical structure (the alternation of strong and weak syllables, see Figure 1). Attention to rise times and related syllable-onset cues enables the temporal segmentation of speech into syllables and words. The temporal experience of rhythm in music is also likely to be related to rise time, and the amplitude envelope of a simple metrical rhythm is also shown in Figure 1. Different instruments will produce notes with different rise times (attack times, see Gordon, 1986), hence musicians playing different instruments who entrain together must (unconsciously?) phase-align the rise times of their notes in order to keep accurate time.



**Fig. 1:** Schematic depiction of the waveforms (amplitude envelopes) for one of the metrical musical sequences used by Huss et al. (2011, wav 002) and for the first line of the nursery rhyme “Mary Mary Quite Contrary”. The amplitude envelope is marked in blue in each case, showing that the stressed syllables (MA-ry MA-ry QUITE con-TRA-ry) and the accented notes have larger rise times.

Most of the critical information for speech intelligibility occurs below around 10 Hz (Cogan & Poeppel, 2011). As noted by Clayton, musical entrainment also occurs within a frequency range of 0.5 – 10 Hz. Speech essentially has a rhythmic temporal structure, with syllables occurring between 4 – 7 times per second across languages (mean duration 200 ms, core range 100 – 300 ms, see Ghitza & Greenberg, 2009), and stressed syllables, which anchor the perception of linguistic prosody, occurring on average every 500 ms across languages (see Arvaniti, 2009). By this account, spoken language can be considered a rhythmic stimulus stream analogous to music, to which the brain can entrain.

## NEURAL ENTRAINMENT TO THE SPEECH SIGNAL

Recent developments in both human and animal neuroscience suggest that oscillatory networks in auditory cortex entrain to acoustic input by aligning their firing patterns, so that they are firing in-phase with the rhythmic structure of acoustic signals. This entrainment or “phase-locking” activity is thought to support signal parsing and efficient signal encoding. In animals, electrophysiology has revealed that the natural rhythmic fluctuations in excitability of neuronal networks in auditory cortex can entrain to the temporal structure of rhythmic input (e.g., Lakatos et al., 2008; Schroeder & Lakatos, 2009). When rhythmic predictability is established, neural networks realign the phase of their oscillations so that they are in a high excitability phase when a new event occurs, which enhances (amplifies) the neuronal processing of that event. Auditory events that are “out of phase” with the rhythmic stimulus stream are suppressed, as they arrive during low excitability phases. This neurophysiological work supports a cognitive oscillatory model of human attentional processing proposed by Jones and her colleagues (e.g., Jones, Moynihan, MacKenzie & Puente, 2002). In the human brain, it has been shown that inherent cortical rhythms in auditory cortex phase-reset to align themselves to the syllable structure of

incoming speech, thereby tracking speech dynamics (using magnetoencephalography and functional magnetic resonance imaging, e.g., Luo & Poeppel, 2007; Giraud et al., 2007). Neural phase alignment is driven by onsets or auditory “edges” in the acoustic signal (see Giraud & Poeppel, 2012), which for speech correspond to cues in the amplitude envelope such as rise time (see Goswami, 2011). In fact, on the basis of patient data, it has been argued that stressed syllables, which have particularly marked rise times, provide the critical cues for human oscillatory entrainment (Kotz & Schwartz, 2010).

As noted earlier, the temporal occurrence of stressed syllables is fairly similar across languages, with an approximate rhythmic rate of 2 Hz (Arvaniti, 2009). This approximate regularity may be important in linguistic processing, providing a framework for rhythmic entrainment. When a stressed syllable is produced, the rise time to the vowel is large, hence the rhythmic succession of vowel nuclei in stressed syllables might provide a framework for oscillatory entrainment (Kotz & Schwartz, 2010). Attention to when the next stressed syllable should occur would guide the listener to when important information may be expected to appear in the speech signal, and stressed syllables carry high informational content (Altman & Carter, 1989). The rhythmic framework provided by successive rise times might facilitate neural entrainment to the speech signal, enabling phase alignment and amplifying neuronal processing.

Importantly, this rhythmic framework will be cross-modal. Visual prosodic cues such as head and eyebrow movements, and lip, jaw and cheek movements, will provide complementary visual rhythmic information. Visual prosodic cues are known to be systematically related to speech amplitude and fundamental frequency (e.g., Munhall, Kross & Vatikiotis-Bateson, 2002; Munhall et al., 2004), and most visual speech information is recovered from the lower spatial and temporal frequencies, which favours a perceptual reliance on amplitude modulation. It is well-known that babies attend selectively to the human face, often showing rapt attention during speech processing to the face of the speaker (Kuhl, 2004). One possibility is that infants are using the combination of visual and auditory prosodic cues to segment the speech stream efficiently (it is already known that auditory rhythm is used by neonates for speech segmentation, e.g., Mehler et al., 1988; Nazzi et al., 1998). This would be analogous to the unconscious use of visual cues in the examples of musical entrainment noted by Clayton (e.g., he notes that in ritual processional music, groups from different communities will avoid looking at each other in order to avoid unintentional entrainment).

## BEHAVIOURAL ENTRAINMENT TO THE SPEECH SIGNAL

Recent studies of neural entrainment in infants show that syllable-rate oscillatory networks in auditory cortex are activated by non-speech rhythmic input from birth onwards (Telkemeyer et al., 2009, 2011). There is also growing evidence from children that rhythmic and metrical skills are related to linguistic development. If we take direct measures of behavioural entrainment, such as tapping in time with a beat, then a series of studies suggest that rhythmic entrainment is impaired in children with language impairments. Children with developmental dyslexia show tapping deficits in comparison to non-dyslexic controls (the task was tapping to a metronome beat at 2.5 and 2 Hz, see Thomson & Goswami, 2008), and dyslexic deficits in rhythmic entrainment are still present in adulthood (Thomson et al., 2006). Children with speech and language impairments (SLI) also show rhythmic entrainment deficits (tapping to a beat at 1.5 Hz and 2 Hz, Corriveau & Goswami, 2009).

By hypothesis, this rhythmic entrainment deficit should be linked to individual differences in processing the structure of the auditory signal, that is in processing the phonological structure of speech. Phonological awareness in children is measured by tasks that assess their awareness of syllables in words, of whether words rhyme, and of individual speech sounds in words. For both children with SLI and children with dyslexia, individual differences in active rhythmic entrainment are significantly related to different measures of phonological awareness, and are also related to rise time discrimination (Corriveau et al., 2007; Thomson & Goswami, 2008). In fact, rise time discrimination is impaired in children with developmental dyslexia across languages (studies so far include English, French, Dutch, Hungarian, Spanish, Chinese and Finnish; see Goswami, Gerson, & Astruc, 2010; Goswami et al., 2002; Goswami et al., 2011; Hämäläinen et al., 2009; Hämäläinen et al., 2005; Muneaux et al., 2004; Poelmans et al., 2011; Richardson et al., 2004; Suranyi et al., 2009). A recent study of neural entrainment to rhythmic input in dyslexic adults further found a significant phase locking (phase alignment) deficit for sinusoidal amplitude-modulated noise at 2 Hz (a rhythmic stimulus stream, see Hamalainen et al., 2012). 2 Hz corresponds to the rhythmic rate of stressed syllables in the amplitude envelope of speech, suggesting that the rhythmic entrainment deficits and the rise time deficits are indeed linked to language deficits (Goswami, 2011).

## METRICAL STRUCTURE, PHONOLOGY AND MUSIC

The data showing developmental links between rise time perception, rhythmic entrainment and phonological awareness can therefore be interpreted as showing the core importance of rhythmic entrainment to metrical structure (the succession of stressed syllables) for successful language acquisition by children. Impaired perception of rise time seems to accompany impaired rhythmic entrainment and impaired perception of linguistic phonological structure. This theoretical perspective predicts that children and adults with dyslexia should have difficulty in perceiving speech prosody and syllable stress, and this is indeed the case (e.g., Goswami et al., 2010; Goswami, 2011; Leong et al., 2011). In fact, children with developmental dyslexia also have difficulty in perceiving metrical structure in music (Huss et al., 2011). In contrast to metrical structure in speech, musical meter is perfectly periodic, dependent on the accenting of certain notes in the underlying beat structure. Nevertheless, children with dyslexia are impaired at perceiving metrical musical structure, and individual differences are linked to rise time discrimination and phonological awareness.

As long argued by Trehub, language and music may result from general perceptual mechanisms that are neither music- nor language-specific (e.g., Trehub & Hannon, 2006). If one core mechanism is the perception of rhythm and metrical structure, driven by rise time perception, then it is possible that the more overt rhythms of music may be utilised to enhance children's processing of the less obvious rhythms of language. Many of the language routines of early childhood emphasise metrical structure (nursery rhymes are perfect metrical poems), and if children are singing nursery rhymes then they are learning to align the "beats" of stressed syllables to the beat of the music. The rhythmic structure of music and language are aligned when we sing, and therefore children who are perfectly "in phase" when singing should also be those children who entrain more accurately and who develop better phonological skills. This prediction about music was supported by a recent study by Dege and Schwarzer (2011), who provided musical training to young German children (5-year-olds). They demonstrated a positive effect of training on the children's phonological awareness of syllables and rhymes. We have been running a similar study using bongo drumming and singing (Verney & Goswami, 2011), which also varies the pulse rate of the musical training. We also find that individual differences in entrainment accuracy are related to syllable and rhyme awareness in 5-year-olds. The most consistent links are found at entrainment rates linked to the syllable structure of speech, namely 500 ms (2 Hz). Although intervention studies of children with phonological impairments are yet to be conducted, it seems likely that musical activities that emphasise metrical structure and which align metrical structure in language and music could be very beneficial (Overy, 2003; see also Huss et al., 2011, for discussion). Meanwhile, these ideas extend the range of entrainment phenomena in music beyond those discussed by Clayton. A theoretical framework based on amplitude rise time and metrical entrainment offers a principled cognitive and neural framework whereby multi-modal rhythmic entrainment may underpin not just musical processing, but speech processing as well (Goswami, 2011).

### END NOTES

[1] Corresponding author: Professor Usha Goswami, Centre for Neuroscience in Education, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK; email [ucg10@cam.ac.uk](mailto:ucg10@cam.ac.uk). Thanks are due to Vicky Leong for creating Figure 1. Usha Goswami is supported by a Major Research Fellowship from the Leverhulme Trust.

### REFERENCES

- Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66, 46-63.
- Cummins, F. (2012). Looking for rhythm in speech. *Empirical Musicology Review*, Vol. 7, Nos. 1-2 (forthcoming).
- Corriveau, K., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairment: Tapping to the beat. *Cortex*, Vol. 45, pp. 119-130.
- Dege, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology*, Vol. 2, No. 124.

Drullman, R. (2006). The significance of temporal modulation frequencies for speech intelligibility. In (Eds.) S. Greenberg & W.A. Ainsworth, *Listening to Speech, An Auditory Perspective*. Mahwah, NJ: LEA, pp. 39-47.

Drullman, R., Festen, J.M., & Plomp, R. (1994). Effect of temporal envelope smearing on speech perception. *Journal of the Acoustical Society of America*, Vol. 95, pp. 1053-1064.

Elliott, T.M., & Theunissen, F.E. (2009). The modulation transfer function for speech intelligibility. *Public Library of Science, Computational Biology*, 5, e1000302, doi:10.1371/journal.pcbi.1000302.

Ghitza O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, Vol. 2, No. 130.

Ghitza O. & Greenberg S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, Vol. 66, pp. 113-126.

Giraud, A.L., Kleinschmidt, A., Poeppel, D., Lund, T.E., Frackowiak, R.S.J. & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, Vol. 56, pp. 1127-34.

Giraud, A.L. & Poeppel, D. (2012). Speech perception from a neurophysiological perspective. In Poeppel, D., Overath, T., Popper, A.N., and Fay, R.R. (Eds). *Human Auditory Cortex*. New York: Springer Science+Business Media, LLC, pp. 225-260.

Gordon, J.W. (1986). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America*, Vol. 82, pp. 88-105.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, Vol. 15, No. 1, pp. 3-10.

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading & Writing*, Vol. 23, pp. 995-1019.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S. & Scott, S.K. (2002). Amplitude envelope onsets and developmental dyslexia: a new hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 99, No. 16, pp.10911-10916.

Goswami, U., Wang, H-L., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish and Chinese. *Journal of Cognitive Neuroscience*, Vol. 23, pp. 325-337.

Greenberg, S., Carvey, H., Hitchcock, L. & Chang, S. (2003). Temporal properties of spontaneous speech – a syllable-centric perspective. *Journal of Phonetics*, Vol. 31, pp. 465-485.

Hämäläinen, J., Rupp, A., Soltesz, F., Szucs, D., & Goswami, U. (2012). Reduced phase locking to slow amplitude modulation in adults with dyslexia: An MEG study. *Neuroimage*, Vol. 59, No. 1, pp. 2952-62.

Hämäläinen, J., Leppänen, P.H.T., Torppa, M., Muller, K., & Lyytinen, H. (2005). Detection of sound rise time by adults with dyslexia. *Brain & Language*, Vol. 94, No. 1, pp. 32-42.

Hämäläinen, J., Leppänen, P. H. T., Eklund, K., Thomson, J., Richardson, U., Guttorm, T. K., Witton, C., Poikkeus, A. M., Goswami, U., & Lyytinen, H. (2009). Common variance in amplitude envelope perception tasks and their impact on phoneme duration perception and reading and spelling in Finnish children with reading disabilities. *Applied Psycholinguistics*, Vol. 30, pp. 511-530.

Houtgast, T. & Steeneken, H. J. M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America*, Vol. 77, pp. 66-73.

Huss, M., Verney, J.P., Fosker, T., Fegan, N. & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, Vol. 47, pp. 674-689.

Jones, M.R., Moynihan, H. MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, Vol. 13, No. 4, 313-319.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, Vol. 5, pp. 831-843.

Kotz, S. & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences*, Vol. 14, No. 9, 392-399

Lakatos. P., Karmos, G., Mehta, A. D., Ulbert, I., Schroeder, C. E. (2008). Oscillatory entrainment as a mechanism of attentional selection. *Science*, Vol. 320, pp. 110-113.

Leong, V., Hämäläinen, J., Soltesz, F., & Goswami, U. (2011). Rise Time Perception and Detection of Syllable Stress in Adults with Developmental Dyslexia. *Journal of Memory and Language*, Vol. 64, pp. 59-73.

Luo H, Liu Z, Poeppel D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, Vol. 8, 8: e1000445.

Luo, H., & Poeppel D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, Vol. 54, pp. 1001-10.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, Vol. 29, pp. 143-78.

Muneaux, M., Ziegler, J.C., Truc, C., Thomson, J. & Goswami, U. (2004). Deficits in beat perception and dyslexia : evidence from French. *NeuroReport*, Vol. 15, No. 8, pp. 1255-1259.

Munhall, K.G, Jones, J.A, Callan, D.E, Kuratate, T. & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychological Science*, Vol.15, pp. 133-137.

Munhall, K.G., Kross, C., & Vatikiotis-Bateson, E. (2002). Audiovisual perception of band-pass filtered faces. *Journal of the Acoustical Society of Japan*, Vol. 21, pp. 519-520.

Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination by newborns : Towards an understanding of the role of rhythm. *Journal of Experimental Psychology, Human Perception and Performance*, Vol. 24, pp. 756-766.

Overy, K. (2003). From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*, Vol. 999, pp. 497-505.

Poelmans, H., Luts, H., Vandermosten, M., Boets, B., Ghesquiere, P. & Wouters, J. (2011). Reduced sensitivity to slow-rate dynamic auditory information in children with dyslexia. *Research in Developmental Disabilities*, doi:10.1016/j.ridd.2011.05.025

Richardson, U. Thomson, J., Scott, S.K. & Goswami, U. (2004). Supra-segmental auditory processing skills and phonological representation in dyslexic children. *Dyslexia*, Vol. 10, No. 3, pp. 215-233.

Schroeder CE, Lakatos P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neuroscience*, Vol. 32, pp. 9-18.

Shannon, R.V., Zeng, F-G, Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, Vol. 270, pp. 303-304.

Surányi, Z., Csépe, V., Richardson, U., Thomson, J.M., Honbolygó, F., & Goswami, U. (2009). Sensitivity to rhythmic parameters in dyslexic children: A comparison of Hungarian and English. *Reading & Writing*, Vol. 22, pp. 41-56.

Telkemeyer S, Rossi S, Koch SP, Nierhaus T, Steinbrink J, Poeppel D, Obrig H and Wartenburger I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience*, Vol. 29, pp. 14726-14733.

Telkemeyer, S., Rossi, S., Nierhaus, T., Steinbrink, J., Obrig, H., & Wartenburger, I. (2011). Acoustic processing of temporally modulated sounds in infants: Evidence from a combined near-infrared spectroscopy and EEG study. *Frontiers in Psychology*, Vol. 2, article 62.

Thomson, J.M. & Goswami, U. (2008). Rhythmic Processing in Children with Developmental Dyslexia: Auditory and Motor Rhythms link to Reading and Spelling. *Journal of Physiology-Paris*, Vol. 102, pp. 120-129.

Thomson, J.M., Fryer, B., Maltby, J. & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading*, 29, 3, 334-348.

Trehub, S.E., & Hannon, E.E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition*, Vol. 100, pp. 73-99.

Verney, J.P. & Goswami, U. (2011). Rhythmic entrainment and singing in 5-year-old children: temporal accuracy at four temporal rates and relations with phonological awareness. The Neurosciences and Music IV: Learning and Memory, Edinburgh, 9-12 June 2011, c28 (poster session).

Yehia, H., Kuratate, T. & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, Vol. 30, pp. 555-568.