frontiers in
HUMAN NEUROSCIENCE

REVIEW ARTICLE
published: 13 October 2014
doi: 10.3389/fnhum.2014.00777

The role of rhythm in speech and language rehabilitation: the SEP hypothesis

***Shinya Fujii*[1]* and *Catherine Y. Wan*[2]**

[1] *Heart and Stroke Foundation Canadian Partnership for Stroke Recovery, Sunnybrook Research Institute, Toronto, ON, Canada*
[2] *Department of Radiology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA*

Edited by:
Eckart Altenmüller, University of Music and Drama Hannover, Germany

Reviewed by:
Cyril R. Pernet, The University of Edinburgh, UK
Sonja A. E. Kotz, Max Planck Institute Leipzig, Germany

***Correspondence:**
Shinya Fujii, Heart and Stroke Foundation Canadian Partnership for Stroke Recovery, Sunnybrook Research Institute, 2075 Bayview Avenue, Toronto, ON M4N 3M5, Canada
e-mail: sfujii@sri.utoronto.ca

For thousands of years, human beings have engaged in rhythmic activities such as drumming, dancing, and singing. Rhythm can be a powerful medium to stimulate communication and social interactions, due to the strong sensorimotor coupling. For example, the mere presence of an underlying beat or pulse can result in spontaneous motor responses such as hand clapping, foot stepping, and rhythmic vocalizations. Examining the relationship between rhythm and speech is fundamental not only to our understanding of the origins of human communication but also in the treatment of neurological disorders. In this paper, we explore whether rhythm has therapeutic potential for promoting recovery from speech and language dysfunctions. Although clinical studies are limited to date, existing experimental evidence demonstrates rich rhythmic organization in both music and language, as well as overlapping brain networks that are crucial in the design of rehabilitation approaches. Here, we propose the "SEP" hypothesis, which postulates that (1) "sound envelope processing" and (2) "synchronization and entrainment to pulse" may help stimulate brain networks that underlie human communication. Ultimately, we hope that the SEP hypothesis will provide a useful framework for facilitating rhythm-based research in various patient populations.

Keywords: rhythm, speech, language, rehabilitation, the SEP hypothesis

INTRODUCTION

Human beings have universally engaged in rhythmic musical activities such as drumming, dancing, singing, and playing musical instruments since ancient times (e.g., Mithen, 2005; Fitch, 2006; Conard et al., 2009). The presence of rhythmic sounds in the environment can result in spontaneous motor responses such as tapping, clapping, stepping, dancing, and singing (e.g., Kirschner and Tomasello, 2009; Sevdalis and Keller, 2010; Fujii et al., 2014). Rhythm serves as a potent catalyst to elicit positive affect (e.g., Zentner and Eerola, 2010), co-operation (e.g., Reddish et al., 2013), and social bonding (e.g., Freeman, 2000). In recent years, researchers have begun to explore the therapeutic potential of rhythm in speech and language rehabilitation. In this paper, we discuss the importance of rhythm as a medium of human communication and social interaction. Specifically, we discuss the important role of rhythm in speech perception and production, and summarize the relevant neuroscience literature. In addition, we present the "SEP" hypothesis, which postulates that (1) sound envelope processing and (2) synchronization and entrainment to a pulse, may help stimulate brain networks that underlie human communication. Finally, we provide examples of speech and language disorders [Parkinson's disease, stuttering, aphasia, and autism] that can potentially benefit from rhythm-based therapy.

RHYTHM AS A MEDIUM OF COMMUNICATION

Rhythm, or the temporal organization of perceived or produced events, mediates communication and social interaction. For example, normal rhythm or rate of syllable production during speech is typically three to eight syllables per second (3–8 Hz) across many languages (Malecot et al., 1972; Crystal and House, 1982; Greenberg et al., 2003; Chandrasekaran et al., 2009). This rate range corresponds to natural movement frequencies of articulators including tongue, palate, cheek, jaw, and lips coupled with voicing (Peelle and Davis, 2012). If the rate is faster than 8 Hz, however, speech intelligibility is significantly reduced (e.g., Ahissar et al., 2001), suggesting that our brain may be "tuned" to the natural rhythm of vocal production.

Primate studies have shown a similar rhythmic tuning to communicative gestures such as lip-smacking (Fitch, 2013; Ghazanfar et al., 2013; Ghazanfar and Takahashi, 2014). Lip-smacking is often directed at another animal during face-to-face interactions, and is characterized by regular cycles of vertical jaw movement, often involving a parting of the lips (e.g., Ghazanfar et al., 2012). For example, Ghazanfar et al. (2013) used three types of video clips as visual stimuli in which monkey avatars were lip-smacking at frequencies of 3, 6, and 10 Hz. Interestingly, the preferential viewing times were significantly longer for the smacking rate of 6 Hz, which corresponds to the natural syllable production rate, compared with those of 3 and 10 Hz. Moreover, the monkeys in the study responded to the avatars in the video with their own rhythmic lip-smacking expressions, as if they were communicating with real monkeys. Based on these observations, Ghazanfar et al. (2013) suggested that monkey lip-smacking and human speech rhythms share a similar sensorimotor mechanism, and that human speech might have evolved from the rhythmic

gestures or motor actions normally produced by our primate ancestors.

The idea that rhythmic motor actions mediate communication is supported by another primate study, which showed that rhythmic non-vocal sounds created by drumming actions served as communicative signals (Remedios et al., 2009). The mean rate of drumming actions by the macaque monkeys were around five beats per second (Remedios et al., 2009), which also corresponded to the natural syllable production rate during speech. Further investigation with functional magnetic resonance imaging (fMRI) showed that neural responses when animals listened to rhythmic drumming sounds overlapped with those when they listened to vocalizations (Remedios et al., 2009).

Human studies have also shown the importance of gestures and rhythmic motor actions for communication and social interaction. For example, lip movements affect the way in which people perceive speech syllables (McGurk and MacDonald, 1976). In congenitally blind individuals, verbal communication is often accompanied by hand gestures, although they have never seen hand gestures, and the listener cannot see the speaker's movements (Iverson and Goldin-Meadow, 1998). Developmental studies have shown that newborn infants imitate adult facial and manual gestures (e.g., Meltzoff and Moore, 1977) and synchronize body movements with the articulated structure of adult speech (e.g., Condon and Sander, 1974). Furthermore, 3- to 4-month-old infants show altered vocalizations and synchronized limb movements in response to rhythmic dance music (Fujii et al., 2014), while 5- to 24-month-old infants engage in more rhythmic limb movements and smile more during music listening (Zentner and Eerola, 2010). Older preschool children spontaneously play the drum in synchrony when a human adult partner plays the drum (Kirschner and Tomasello, 2009). Thus, rhythmic sounds and motor actions are fundamental for communication and social interaction throughout development.

THE ROLE OF RHYTHM IN SPEECH

Rhythm is essential to the understanding of speech. In order to comprehend spoken language, listeners are required to perceive temporal organization of phonemes, syllables, words, and phrases from an ongoing speech stream (Kotz and Schwartze, 2010; Patel, 2011; Peelle and Davis, 2012). An important source of acoustic information that conveys rhythm in speech is the *sound envelope*, which is defined as the acoustic power summed across all frequencies for a given frequency range (Kotz and Schwartze, 2010; Patel, 2011; Peelle and Davis, 2012). As illustrated in **Figure 1**, the phrase "Happy birthday to you" can be broken down into six syllables (i.e., "Ha/ppy/birth/day/to/you"), and these boundaries correspond to the pattern of the sound envelope (denoted by vertical dashed lines). Thus, burst patterns of the sound envelope represent rhythm or temporal organization in vocalization.

A number of studies have demonstrated the importance of rhythm or sound envelope in speech comprehension (e.g.,

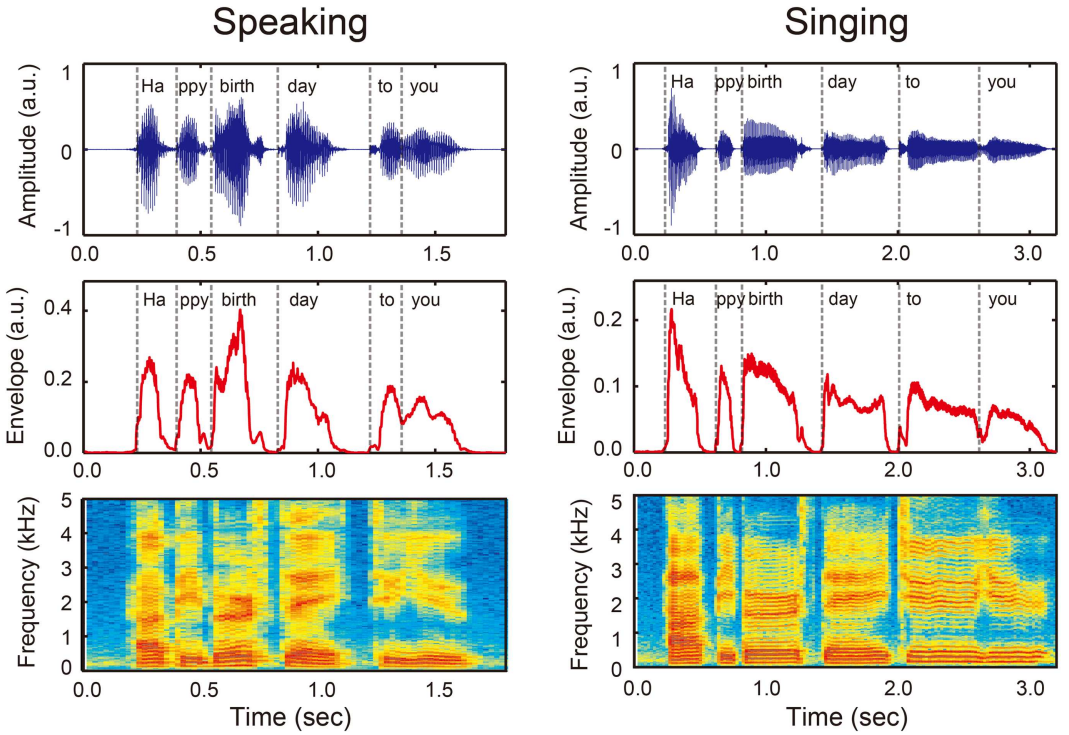

FIGURE 1 | An example of sound wave (upper panels), amplitude envelope (middle panels), and power spectrum (bottom panels) when a person speaks (left) and sings (right) "Happy birthday to you" that can be divided into six syllables (i.e., Ha/ppy/birth/day/to/you, see vertical dashed lines). Sound envelope is an important acoustic information that conveys temporal organization of phonemes and syllables or rhythm in vocalization. Note that rhythm in singing (right) has a more salient pulse- or beat-based timing compared with rhythm in speech (left).

Drullman et al., 1994a,b; Nazzi et al., 1998; Ahissar et al., 2001; Smith et al., 2002; Elliott and Theunissen, 2009; Bertoncini et al., 2011). For example, Shannon et al. (1995) tested the importance of rhythm in speech by minimizing the fine spectral information while preserving the sound envelope. Near perfect speech recognition performance was observed when individuals were presented with these speech stimuli. Consistent with this finding, smearing of the rhythm or sound envelope in speech sounds significantly reduced sentence intelligibility (e.g., Drullman et al., 1994a,b; Elliott and Theunissen, 2009). The reliance on rhythm or sound envelope to discriminate speech sounds has also been reported in infants (Nazzi et al., 1998; Bertoncini et al., 2011). Smith et al. (2002) further investigated the different roles of envelope and fine spectral structure in human auditory perception. They created sound stimuli called "auditory chimeras," which consisted of the envelope of one sound and the fine spectral structure of another (Smith et al., 2002). Interestingly, when the two features (i.e., envelope and fine spectral structure) were in conflict, the pitch and location of sounds were determined by the fine spectral structure, while the words identified were based on the envelope (Smith et al., 2002). Thus, along with fine spectral information, sound envelope or rhythm is essential for speech intelligibility.

NEURAL CORRELATES OF RHYTHMIC SPEECH PERCEPTION

A question arises then, regarding how rhythm or sound envelope is processed in the brain. fMRI studies have shown that the processing of sound envelope or low-frequency temporal feature in the acoustic signal is associated with activities in the inferior colliculus of the brainstem, the medial geniculate body of the thalamus, the Heschl's gyrus (HG), the superior temporal gyrus (STG), and the superior temporal sulcus (STS) (Giraud et al., 2000; Boemio et al., 2005). The "asymmetric sampling in time (AST)" hypothesis postulates that low-frequency temporal features in the acoustic signals are lateralized to the right hemisphere, whereas high-frequency fine spectral features of the acoustic signals are lateralized to the left hemisphere (Poeppel, 2003; McGettigan and Scott, 2012). Consistent with this hypothesis, electroencephalography (EEG) studies have also shown that sound envelope processing is right lateralized (Abrams et al., 2008, 2009). Similarly, phase pattern of theta band (4–8 Hz) responses recorded from the temporal cortex using magnetoencephalography (MEG), especially in the right hemisphere, is correlated with the degree of speech intelligibility (Luo and Poeppel, 2007). Thus, the neural mechanisms underlying rhythm or sound envelope processing are likely to involve the brainstem, the thalamus, and the auditory regions in the temporal cortex, which may be lateralized to right hemisphere (see pink arrows and "R" in **Figure 2B**).

In parallel with the brainstem-thalamo-cortical (temporal auditory regions) pathway, there is another possible pathway for rhythm perception in speech, which involves the brainstem, the cerebellum, the thalamus, the supplementary motor area (SMA), the basal ganglia (BG), and the prefrontal cortex (light blue arrows in **Figure 2B**) [see Kotz and Schwartze (2010)]. Here, early auditory input is transmitted to the cerebellum via the cochlear nuclei of the brainstem (Huang et al., 1982; Wang et al., 1991; Xi et al., 1994; Kotz and Schwartze, 2010). The cerebellum is responsible for the encoding of event-based temporal structure, and relays information to the SMA via the thalamus, which further transmits information to the prefrontal cortex (Kotz and Schwartze, 2010). The SMA and the prefrontal cortex transmit temporal information to the BG, which transmits information back to the cortex via the thalamus forming the BG-thalamo-cortical loop (Kotz et al., 2009; Kotz and Schwartze, 2010, 2011). This closed-loop circuit is assumed to have functions to continuously evaluate temporal relations, extract temporal regularity, and engage in sequencing of temporal events and analysis of hierarchical structure (Kotz et al., 2009; Kotz and Schwartze, 2010, 2011; Schwartze et al., 2011). For example, when we listen to the six syllables of "ha/ppy/birth/day/to/you," they are grouped into four words "happy/birthday/to/you" and perceived as one phrase (**Figure 1**). Thus, rhythm perception in speech can be regarded as finding the hierarchical structure of temporal events (Kotz et al., 2009; Kotz and Schwartze, 2010; Przybylski et al., 2013). The prefrontal cortex integrates information from the temporal cortex with that being processed in the BG-thalamo-SMA loop circuit to optimize speech comprehension (Kotz and Schwartze, 2010). Taken together, the sound envelope or rhythm is processed in the cortical and subcortical auditory-motor systems.

NEURAL CORRELATES OF RHYTHMIC SPEECH PRODUCTION

In the previous section, we described the neural correlates of sound envelope or rhythm processing in speech perception. We now consider the neural correlates of rhythmic speech production. The directions into velocities of articulators (DIVA) model provide a useful framework to consider the neural mechanisms (see Guenther et al., 2006; Bohland et al., 2010; Tourville and Guenther, 2011). According to the DIVA model, rhythmic speech production is achieved by sequential control of the velocities of articulators including the upper and lower lips, the jaw, the tongue, and the larynx. In this model, the bilateral ventral primary motor cortex (vM1), which corresponds to the cortical homunculus of speech articulators (Penfield and Rasmussen, 1950; Penfield and Roberts, 1959), is responsible for outputting motor commands to the muscles of articulators [see also Kalaska et al. (1989), Ludlow (2005), Brown et al. (2008), and Olthoff et al. (2008)]. The vM1 receives inputs from the SMA connected with the BG and the thalamus (green arrows in **Figure 2C**) [see also Jurgens (1984) and Luppino et al. (1993)]. This is supported by fMRI studies that showed bilateral activities in the BG, thalamus, and SMA during speech production (e.g., Bohland and Guenther, 2006; Tourville et al., 2008). The BG-thalamo-SMA circuit is hypothesized to play a role in sequencing and self-initiation of speech, or rhythmic speech production in the DIVA model. This is based on clinical studies that showed speech production problems including involuntary vocalizations, echolalia, lack of prosody, stuttering-like output, variable rate, and difficulties with complex speech sequences following the impairment of the BG-thalamo-SMA circuit (e.g., Jonas, 1981; Ziegler et al., 1997; Ho et al., 1998; Pickett et al., 1998; Vargha-Khadem et al., 1998; Pai, 1999; Watkins et al., 2002).

In the DIVA model, the bilateral vM1 also receives inputs from left ventral premotor cortex (vPMC) and adjacent posterior inferior frontal gyrus (pIFG) (see red arrow and "L" in **Figure 2C**). These areas (i.e., left vPMC and pIFG) are hypothesized to

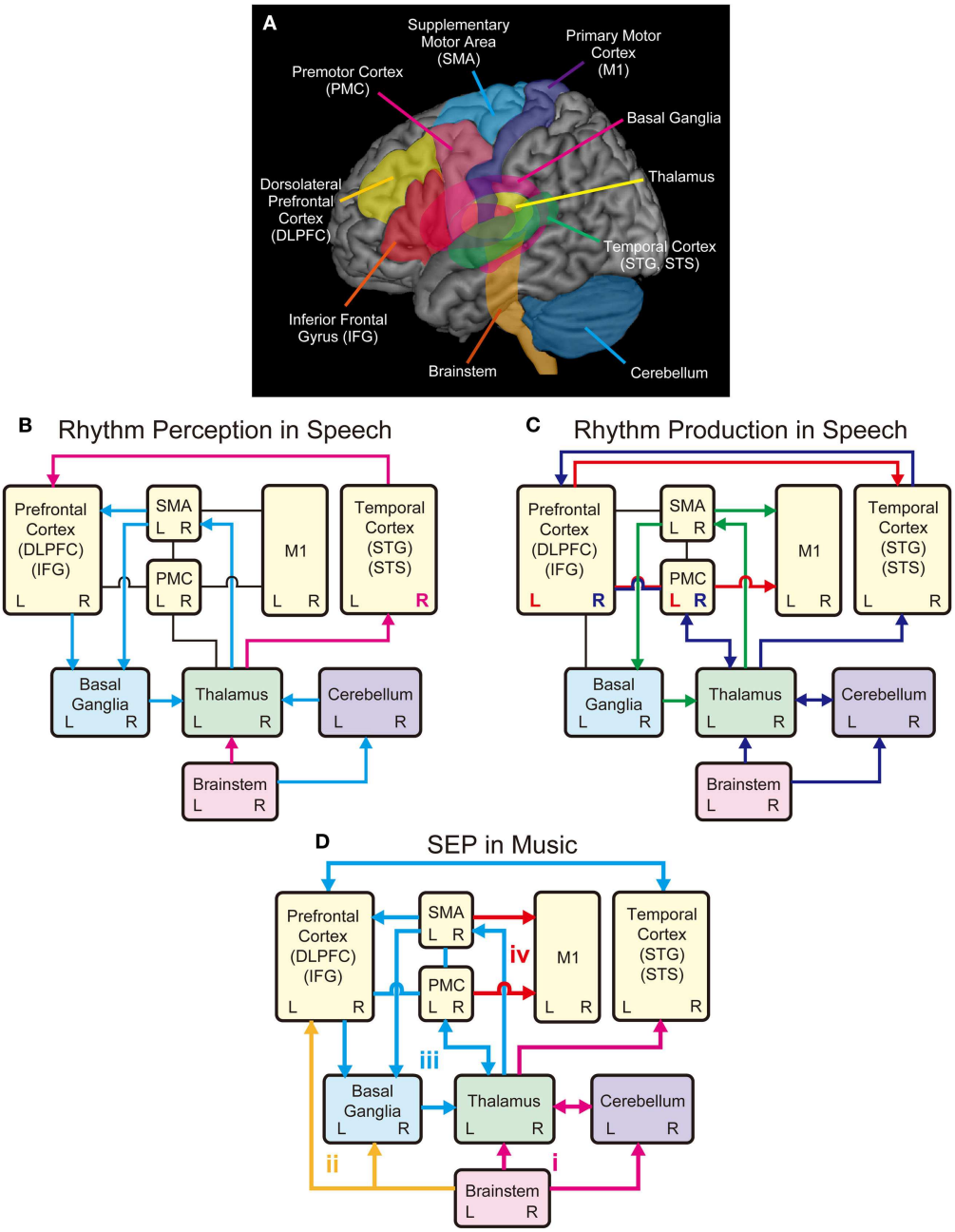

FIGURE 2 | Schematic models of shared brain network for rhythm perception and production in speech and music. **(A)** Possible shared brain regions for rhythm processing in music and speech. **(B)** A model for rhythm perception in speech. The temporal cortex receives auditory inputs from the brainstem via the thalamus, which further transmits information to the prefrontal cortex (pink arrows). Processing of sound envelope or low-frequency temporal feature in the acoustic signals may be lateralized to the right hemisphere in the temporal cortex (see pink R). The cerebellum also receives the auditory inputs from the brainstem and relays information to the SMA via the thalamus, and further transmits information to the prefrontal cortex to process temporal events. The SMA and the prefrontal cortex transmit information to the basal ganglia (BG), which transmits information back to the cortex via the thalamus forming the BG-thalamo-cortical loop (light blue arrows). **(C)** A model for rhythm production in speech. The M1 receives inputs from the SMA, which forms the SMA-BG-thalamo loop, for rhythmic speech production (green arrows). The M1 also receives inputs from the left PMC and IFG, which transform speech sounds into motor commands (see red arrow and L). The left PMC and IFG also transmit information to the temporal cortex, which is associated with sensory predictions. The temporal cortex monitors the sensory predictions and the auditory feedback received from the brainstem via the thalamus. The feedback errors from the temporal cortex are sent to right PMC and IFG, which is interconnected with the thalamus and the cerebellum (see blue arrows and R). **(D)** A model for Sound Envelope Processing (SEP) and Synchronization and Entrainment to a Pulse (SEP) in music. Rhythm-based therapy may help stimulate brain networks involving (i) the auditory-afferent circuit consisting of brainstem, thalamus, cerebellum, and temporal cortex (pink arrows) for precise encoding of sound envelope and temporal events; (ii) the subcortical-prefrontal circuit for emotional and reward-related processing (yellow arrows); (iii) the BG-thalamo-cortical circuit for processing beat-based timing (light blue arrows); and (iv) the cortical-motor efferent circuit for motor output (red arrows).

form the "speech sound map," which transform speech sounds (e.g., phonemes and syllables) into motor commands (Guenther et al., 2006; Bohland et al., 2010; Tourville and Guenther, 2011). In other words, the speech sound map is "mental syllabary" or repository of learned speech motor program [see also Levelt and Wheeldon (1994) and Levelt et al. (1999)]. The speech sound map has anatomical correspondence with Broca's area (e.g., Dronkers et al., 2007) and has functional correspondence with the "mirror-neuron system" (Rizzolatti et al., 1996; Kohler et al., 2002). Language function is generally regarded as left lateralized because impairments of these brain regions (i.e., left vPMC and pIFG) lead to significant deficit of speech production (e.g., Dronkers, 1996; Kent and Tjaden, 1997; Hillis et al., 2004; Duffy, 2005).

The speech sound map (i.e., left vPMC and pIFG) is hypothesized to project not only to the vM1 to form the motor commands but also to the bilateral auditory areas in the temporal cortex (Guenther et al., 2006; Tourville and Guenther, 2011) (see red arrow in **Figure 2C**). The auditory areas include two locations along the posterior superior temporal gyrus (pSTG): the lateral one near the STS, and the medial one at the junction of the temporal and parietal lobes deep in the sylvian fissure (Guenther et al., 2006; Tourville and Guenther, 2011). These auditory areas are activated not only during speech perception but also during speech production (e.g., Buchsbaum et al., 2001; Hickok and Poeppel, 2004). The projections from the speech sound map to these auditory areas are responsible for predicting the sound being produced, which is compared with the actual auditory feedback being processed in the HG and the adjacent anterior planum temporale (PT) (Guenther et al., 2006; Tourville and Guenther, 2011). If the auditory feedback does not fall within the predicted range, the error signals are sent back to the right vPMC and pIFG according to the DIVA model (Guenther et al., 2006; Tourville and Guenther, 2011) (see blue arrows and "R" in **Figure 2C**). The right vPMC and pIFG are interconnected with the cerebellum via the thalamus, forming the "feedback control map," which transforms the sensory error signals into corrective motor commands (Tourville and Guenther, 2011). This is based on fMRI studies that showed increased hemodynamic responses in the right PMC, pIFG, and the cerebellum during speech production under perturbed auditory feedback conditions (Tourville et al., 2008). Taken together, it is assumed that the BG-thalamo-cortical (vM1 and SMA) circuit is essential for rhythmic speech production, and the other areas (vPMC, pIFG, temporal cortex, and cerebellum) also play important roles for sensorimotor transformation and integration processes.

THEORETICAL RATIONALE UNDERLYING THE ROLE OF RHYTHM FOR SPEECH AND LANGUAGE REHABILITATION: THE "OPERA" AND "SEP" HYPOTHESES

The aim of this section is to provide a rationale underlying the role of rhythm for speech and language rehabilitation considering the above mentioned neural correlates of rhythm perception and production in speech. Here, we present the "SEP" hypothesis, which is a rhythm-specific extension of the "OPERA" hypothesis (Patel, 2011, 2012, 2014), to explain how and why musical rhythm can benefit speech and language rehabilitation.

The OPERA hypothesis is a conceptual framework that postulates how general musical activities can facilitate speech and language processing (Patel, 2011, 2012, 2014). The OPERA hypothesis assumes that five conditions are needed to drive the benefit of musical activities for speech and language processing: (1) overlap: there is anatomical overlap in the brain networks during the processing music and speech, (2) precision: music places higher demands on these shared networks than does speech, in terms of the precision of processing, (3) emotion: musical activities that engage this network elicit strong positive emotion, (4) repetition: musical activities that engage this network are frequently repeated, and (5) attention: musical activities that engage this network are associated with focused attention. In other words, condition (1) describes shared neural underpinnings for both music and speech activities while (2)–(5) are distinctions of musical activities that may drive the neural plasticity enhancing the abilities of speech and language processing.

While the OPERA hypothesis provides a useful conceptual framework for general musical processing, it has a number of limitations that preclude its application to rhythm-based therapy. First, the OPERA hypothesis describes the possible overlap in brain networks during music and speech processing, but the explanation is restricted to perception (i.e., afferent process from cochlea to auditory cortex level). Within the context of rhythm, however, it is important to clarify the overlapping brain networks during sensorimotor coupling and production (i.e., cortical auditory-motor interaction and efferent process from the motor cortex to the spinal cord). Second, the OPERA hypothesis covers many aspects of general musical processing that include pitch and timbre. Here, we specifically discuss how rhythm itself meets with the five conditions of the OPERA hypothesis (i.e., overlap, precision, emotion, repetition, and attention). Third, under the OPERA hypothesis, it is not clear whether rhythm *per se* would have therapeutic potential in patient populations.

In order to address the above limitations, we propose the "SEP" hypothesis, which postulates two additional components to describe how and why rhythm in particular can be beneficial for speech and language rehabilitation: (1) sound envelope processing and (2) synchronization and entrainment to a pulse. The first key component of the SEP hypothesis, sound envelope processing, is adapted from the OPERA hypothesis (Patel, 2011, 2012, 2014), which postulates the major sources of overlap in the brain networks during rhythm perception in music and speech [see also Peretz and Coltheart (2003), Corriveau and Goswami (2009), Kotz and Schwartze (2010), and Goswami (2011)]. The second key component in the SEP hypothesis, synchronization and entrainment to a pulse, postulates the major sources of overlap in the brain networks not only for rhythm perception but also for rhythm production and sensorimotor coupling in music and speech (Guenther et al., 2006; Kotz et al., 2009; Bohland et al., 2010; Kotz and Schwartze, 2010, 2011; Tourville and Guenther, 2011).

In the SEP hypothesis, we assume that the overlap in brain networks for rhythm processing between speech and music involve the brainstem, cerebellum, thalamus, BG, M1, SMA, PMC, prefrontal cortex (DLPFC and IFG), and temporal cortex (STG and STS). As illustrated in **Figure 2D**, we propose four key circuits in the shared brain network to explain the potential benefit of rhythm-based

therapy: (a) the auditory afferent circuit consisted of brainstem, thalamus, cerebellum, and temporal cortex (pink arrows) for precise encoding of sound envelope and temporal events; (b) the subcortical–prefrontal circuit for emotional and reward-related processing (yellow arrows); (c) the BG-thalamo-cortical circuit for processing beat-based timing (light blue arrows); and (d) the cortical motor efferent circuit for motor output (red arrows).

AUDITORY AFFERENT CIRCUIT

According to the OPERA hypothesis, musical activities place high demands on precise encoding of acoustic features including the sound envelope (Patel, 2011, 2012, 2014). This notion is supported by neuroimaging studies that showed more precise encoding of sounds in the brainstem of musicians compared with non-musicians (e.g., Musacchia et al., 2007, 2008; Parbery-Clark et al., 2011, 2012; Strait and Kraus, 2014; Strait et al., 2014). Interestingly, individuals with more musical training exhibit better encoding of speech sounds in the brainstem, larger cortical responses, and better speech sound perception (Strait and Kraus, 2014; Strait et al., 2014). Although the role of the cerebellum for speech and music perception is not mentioned in the OPERA hypothesis, we also include the cerebellum as a part of the auditory afferent circuit because it receives input from the cochlear nuclei (Huang et al., 1982; Wang et al., 1991; Xi et al., 1994; Kotz and Schwartze, 2010) and plays an important role in encoding the absolute duration of time intervals in successive acoustic events (Kotz and Schwartze, 2010; Teki et al., 2011a,b). For example, a recent neuroimaging study showed that perception of changes in musical rhythm (so-called "groove") with drum sounds is associated with activity in the cerebellum and the STG (Danielsen et al., 2014). In addition, musicians showed enhanced activity in the cerebellum compared to non-musicians during temporal perception (Lu et al., 2014) and have larger volumes of cerebellum than non-musicians (Hutchinson et al., 2003). Taken together, we postulate that musical rhythm perception or sound envelope processing in music places high demand on the auditory afferent circuit consisted of the brainstem, thalamus, cerebellum, and temporal cortex (pink arrows in **Figure 2D**). This corresponds to the second condition of the OPERA hypothesis ("precision").

SUBCORTICAL–PREFRONTAL CIRCUIT

Rhythm perception or sound envelope processing in music may engage neural activities in the subcortical–prefrontal circuit relevant for emotional processing. A primate study has shown that rhythmic drumming sounds serve as communicative signals and engage the emotional network in the subcortical areas including the amygdala and the putamen (Remedios et al., 2009). Human neuroimaging studies have shown that listening to music elicit pleasant emotion by engaging the reward system in the subcortical and cortical areas including the midbrain (e.g., ventral tegmental area, periaqueductal gray, and pedunculopontine nucleus), the nucleus accumbens, the striatum, the amygdala, the orbitofrontal cortex, and the ventral medial prefrontal cortex (Blood and Zatorre, 2001; Menon and Levitin, 2005; Salimpoor et al., 2011, 2013; Koelsch, 2014). Recent behavioral studies have also shown that listening to musical rhythm elicits positive affect and a desire to move (Zentner and Eerola, 2010; Witek et al., 2014). Similarly, perception of poetry in the presence of rhyme and regular meter lead to enhanced positive emotions, suggesting that perceiving rhythmic vocalizations may result in positive emotions (Obermeier et al., 2013). Thus, we assume that musical rhythm engages the subcortical–prefrontal circuit for emotional and reward-related processing to elicit positive affect, leading to repetition of actions to reinforce the pleasure actions (yellow arrows in **Figure 2D**). This meets the conditions of (3) emotion and (4) repetition in the OPERA hypothesis.

BG-THALAMO-CORTICAL CIRCUIT

Synchronization and entrainment to a pulse in music may place high demands on information process in the BG-thalamo-cortical circuit. This notion is based on the fact that musical rhythm is more periodic while speech rhythm is quasi-periodic (Peelle and Davis, 2012). Compared with speech rhythm, musical rhythm has a more salient pulse- or beat-based timing. For example, for the phrase "happy birthday to you," the onsets of syllable in the sound envelope are more equally time spaced in singing compared to speaking (**Figure 1**). Neuroimaging studies suggest that perception of beat-based timing (i.e., perception of time intervals with respect to a regular pulse) involve brain networks in the BG, the SMA, the PMC, and the DLPFC (Teki et al., 2011a,b). In fact, beat perception and synchronization increase activities in the BG, SMA, PMC, DLPFC, and STG, and enhance connectivity between the BG with the SMA, PMC, and STG (Chen et al., 2006, 2008a,b; Grahn and Brett, 2007; Grahn and Rowe, 2009; Hove et al., 2013; Kung et al., 2013). Animal studies also suggest the importance of brain networks involving the BG and cortical auditory-motor areas for beat perception and synchronization capabilities (Patel et al., 2009; Schachner et al., 2009; Hasegawa et al., 2011). In addition, tapping to a beat is associated with increased cortical responses in the DLPFC and the inferior parietal lobule (Chen et al., 2008b), which are assumed to be responsible for auditory and temporal attention (Zatorre et al., 1999; Lewis and Miall, 2003; Singh-Curry and Husain, 2009). This attention-related brain network has been shown to be more engaged in precise synchronization performance with the musical beat (Chen et al., 2008b). Taken together, synchronization and entrainment to a pulse in music engages enhanced BG-thalamo-cortical activity (light blue arrows in **Figure 2D**), and this fulfills the fourth and fifth conditions of the OPERA hypothesis ("repetition" and "attention").

CORTICAL MOTOR EFFERENT CIRCUIT

Synchronization and entrainment to a pulse in music can modulate the neural pathway for cortical motor output (red arrows in **Figure 2D**). Not only vocalizations but also other body movements can be synchronized and entrained to the pulse of music, such as tapping, clapping, stepping, dancing, and singing. In terms of the motor-output process in the brain, involvements of both the dorsal and ventral portions of the M1, PMC, and PFC are likely. Neuroimaging studies have shown that cortical hand motor areas are involved not only in hand motor control but also in language processing (e.g., Meister et al., 2003, 2009a,b), suggesting the importance of cortical hand motor areas for human communication. In addition, a recent transcranial magnetic stimulation (TMS) study has shown that listening to groove music modulates

cortico-spinal excitability (Stupacher et al., 2013), suggesting that musical rhythm perception itself may also stimulate the motor-output pathway from the M1 to spinal cord. In sum, there are four circuits of interest in the SEP hypothesis, which may help to stimulate the brain networks underlying human communication.

EXAMPLES OF SPEECH AND LANGUAGE DISORDERS THAT CAN BENEFIT FROM RHYTHM-BASED THERAPY: APPLICATION OF THE "SEP" HYPOTHESIS FOR REHABILITATION

The SEP hypothesis postulates that rhythm-based therapy elicits functional and structural reorganization in the neural networks for human communication in various patient populations via sound envelope processing and synchronization and entrainment to a pulse. In this section, we present examples of speech and language disorders and consider the role of rhythm in speech and language rehabilitation under the framework of the SEP hypothesis. We note, however, that the number of rhythm-based techniques currently available is very limited.

PARKINSON'S DISEASE

Parkinson's disease is a neurodegenerative disorder characterized by progressive deterioration of motor function due to a loss of dopaminergic neurons in the substantia nigra (DeLong, 1990; Wichmann and DeLong, 1998; Blandini et al., 2000). In addition to the more commonly known symptoms such as muscular rigidity, tremor, and postural instability, abnormalities of voice and speech (beyond those associated with aging) are highly prevalent. Indeed, it has been estimated that over 80% of patients with PD develop voice and speech problems at some point (Ramig et al., 2008). Examples of deficits reported by clinicians include monopitch, monoloudness, hypokinetic articulation, and altered speech rate and rhythm (Darley et al., 1969). Analysis of the speech rate of patients with PD showed impaired rhythm and timing organization, such as an accelerated rate of articulation during speaking, as well as a reduction in the total number of pauses (Skodda and Schlegel, 2008; Skodda et al., 2010). When combined with the debilitating motor limb deficits, the loss of speech intelligibility and communication skills can significantly impair the quality of life of patients with PD (Streifler and Hofman, 1984).

To date, only a handful of studies have examined the speech deficits in Parkinson's disease using neuroimaging technique (e.g., Liotti et al., 2003; Pinto et al., 2004; Rektorova et al., 2007, 2012). These studies must be interpreted with caution because the results vary depending on treatment status of patients with PD. For example, patients with PD with no medication and no deep brain stimulation showed significant dysarthria accompanying with a lack of activity in the orfacial motor cortex (M1) and cerebellum while increased activities in the PMC, SMA, and DLPFC compared with the healthy controls (Pinto et al., 2004). These abnormal cortical activities disappeared and the dysarthria symptoms improved after the deep brain stimulation of subthalamic nucleus (STN) in these patients (Pinto et al., 2004). The other fMRI studies investigated mild to moderate patients with PD with levodopa medication (Rektorova et al., 2007, 2012). Compared to healthy controls, patients with PD with the levodopa medication had increased activity in the orofacial sensorimotor cortex (Rektorova et al., 2007) and enhanced functional connectivity in the networks seeded from the periaqueductal gray matter of the midbrain (a core subcortical structure involved in human vocalization) (Rektorova et al., 2012). However, speech productions in these patients with PD with levodopa medication was comparable with that in the controls except for speech loudness, suggesting that the increased brain activity and connectivity might reflect the effects of pharmacological treatment or successful compensatory mechanisms (Rektorova et al., 2007, 2012).

Besides the deep brain stimulation and pharmacological treatments, Lee Silverman Voice Treatment (LSVT) technique has received research attention as a rehabilitation method (Ramig et al., 2001, 2004, 2007; Liotti et al., 2003; Sapir et al., 2011; Sackley et al., 2014). LSVT is designed to improve vocal function in patients with PD by enhancing loudness, intonation range, and articulatory functions. LSVT emphasizes use of loud phonation and high intensity vocal exercises to improve respiratory, laryngeal, and articulatory during speech. Compared with placebo therapy, LSVT has resulted in improvements in speech production parameters such as increases in sound pressure level (i.e., loudness) and semitone standard deviation of fundamental frequency (i.e., prosody), and these improvements were sustained even 12 months after cessation of treatment (e.g., Ramig et al., 2001). The neural correlates of LSVT have been studied by administrating levodopa medication for 4 weeks to mild and moderate patients with PD (Liotti et al., 2003). The results showed that the improvement of speech loudness following the LVST accompanied by neural activities in the striatum, insula, and DLPFC (Liotti et al., 2003).

Under the SEP hypothesis, patients with PD can benefit from synchronization and entrainment to a pulse in music to stimulate the subcortical–prefrontal network and the BG-thalamo-cortical network. As illustrated in **Figure 3**, the BG-thalamo-cortical network functions normally in healthy individuals (top left panel), whereas the network becomes abnormal in PD because of the degeneration of the dopamine-producing neurons in the substantia nigra pars compacta (SNc) (top right panel). The projections from the SNc to the striatum regulates the cortico-strital projections, and if the dopamigeneric neurons in the SNc are depleted, it leads to reduced inhibition in the "direct" pathway to the BG output nuclei (i.e., GPi: internal segment of globus pallidus and SNr: substantia nigra pars reticulata), which carries dopamine "D1" receptors. The degeneration of dopamigeneric neurons in the SNc also leads to increased inhibition in the "indirect" pathway to the BG output nuclei (GPi/SNr) via external segment of globus pallidus (GPe) and STN carrying dopamine "D2" receptors. Net action of the degeneration of dopanigeneric neurons in the SNc leads to the hyper-activation of the BG output nuclei (GPi/SNr) inhibiting activities of thalamocortical projection neurons, which in turn negatively affects motor output [for more detail, see DeLong (1990), Wichmann and DeLong (1996), Blandini et al. (2000), Galvan and Wichmann (2008), and Smith et al. (2012)].

In this model, we assume that pleasurable musical rhythm induces increased endogenous dopamine release in the striatum (bottom left panel in **Figure 3**). This is based on a previous study, which showed an increase in dopamine release and hemodynamic response in the striatum during listening pleasurable music

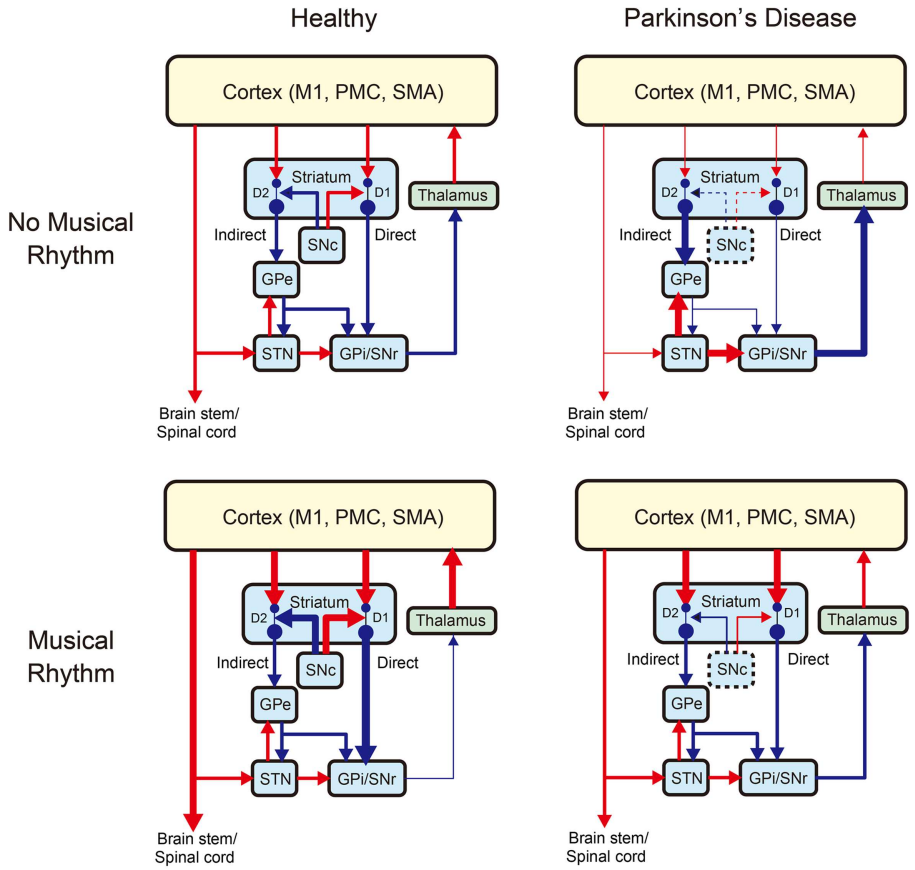

FIGURE 3 | A schematic model of changes in the basal ganglia-thalamo-cortical motor network with Parkinson's disease and/or musical rhythm [modified from Galvan and Wichmann (2008)]. The left and right panels show the network without and with Parkinson's disease, while the upper and lower panels show the network without and with musical rhythm, respectively. Blue arrows indicate inhibitory projections, while red arrows indicate excitatory projections. Changes in the thickness of the arrows indicate increase (thicker arrow) or decrease (thinner arrow) of the projections relative to the normal situation. The dashed lines around the SNc (substantia nigra pars compacta) indicate degeneration of dopaminergic neurons caused by Parkinson's disease. GPe, external segment of globus pallidus; GPi, internal segment of globus pallidus; SNr, substantia nigra pars reticulata; STN, subthalamic nucleus; M1, primary motor cortex; PMC, premotor cortex; and SMA, supplementary motor area. Striatum has "direct" and "indirect" pathways to the basal ganglia (BG) output nuclei (GPi/SNr). Light blue, green, and yellow colors denote BG, thalamus, and cortex, respectively. D1 and D2 indicate subtypes of dopamine receptor. Parkinson's disease induces hyper-activation of BG output nuclei (GPi/SNr) inhibiting activities of the thalamocortical projection neurons, which in turn decreases motor output, while musical rhythm induces hypo-activation BG output nuclei and thereby facilitates thalamo-cortical motor output.

(Salimpoor et al., 2011). In our model, the increased dopamine release leads to increased inhibition in the BG output neuclei (GPi/SNr) and thereby facilitates thalamo-cortical motor output and a desire to move. This idea is also supported by the other studies that showed modulation in the cortico-spinal excitability (Stupacher et al., 2013) and increased activity and connectivity in the brain network including the striatum, PMC, and SMA during perceiving and synchronizing with the musical beats (Chen et al., 2006, 2008a,b; Grahn and Brett, 2007; Grahn and Rowe, 2009; Kung et al., 2013). We also postulate that pleasurable musical rhythm may increase dopamine release even in patients with PD considering that some of the GNc neurons (approximately 30–50%) remain intact even at the time of death (e.g., Davie, 2008). Indeed, a number of studies have shown improvements of motor function in patients with PD during rhythmic auditory stimulation (RAS) (e.g., Thaut et al., 1996; McIntosh et al., 1997; de Bruin et al., 2010; Nombela et al., 2013), suggesting that musical rhythm may facilitate thalamo-cortical motor output in patients with PD (bottom right panel in **Figure 3**).

If our model is correct, then musical rhythm may reduce reliance on levodopa medication and/or deep brain stimulation. However, there remain a few untested assumptions. For example, patients with PD show impaired emotional recognition in music (e.g., van Tricht et al., 2010), suggesting that the ventral portion of the striatum is also affected in patients with PD. Therefore, dopamine release in the striatum may not be increased by music in patients with PD as seen in healthy individuals. Positron emission tomography (PET) can be used to test this hypothesis (Laruelle, 2000; Salimpoor et al., 2011), and whether this may be affected by intervention. A recent study has shown improved perceptual and motor timing in patients with PD after a 4-week music training program with rhythmic auditory cueing (Benoit et al.,

2014), suggesting a possible benefit of RAS on the treatment of PD. However, the participants of that study consisted only of mild to moderate patients with PD (Benoit et al., 2014). Future studies will need to clarify whether the RAS is also beneficial for severe patients with PD, and how the therapeutic effect is different. It may also be important to test whether RAS (simple metronome stimulation as well as rhythmic musical stimulation) improves speech function in patients with PD, given gait functions have been a topic of research interest (e.g., Thaut et al., 1996; McIntosh et al., 1997; de Bruin et al., 2010; Nombela et al., 2013). Additional neuroimaging studies are required to examine the brain networks underlying speech and hand/foot motor processes in patients with PD.

STUTTERING

Stuttering is a developmental condition that affects fluency of speech. It begins during the first few years of life, and affects approximately 5% of preschool-aged children (Bloodstein, 1995). Symptoms include repetition of words or parts of words, as well as prolongations of speech sounds, resulting in disruptions in the normal flow of speech.

As illustrated in **Figure 2C**, during normal speech production, the left IFG and PMC projects to the M1 for vocal-motor output and the right IFG and PMC are monitoring sensory feedback together with the temporal cortex and the cerebellum (blue arrows). However, individuals who stutter show abnormalities in the left IFG and PMC and compensatory hyperactivity in the right IFG and the cerebellum (Fox et al., 1996; Brown et al., 2005; Kell et al., 2009), suggesting that stuttering is associated with poor feed-forward motor command and excessive reliance on auditory feedback control (Max et al., 2004; Civier et al., 2010, 2013). In addition, stuttering is associated with reduced activity and connectivity in the brain network including the BG and the SMA (e.g., Toyomura et al., 2011; Chang and Zhu, 2013), suggesting that stuttering may be due to dysfunction in the BG-thalamo-cortical circuit to produce timing cues for the initiation of the next motor segment in speech (Alm, 2004).

To date, examples of fluency shaping methods for stuttering include altered auditory feedback (e.g., Hargrave et al., 1994; Ryan and Van Kirk Ryan, 1995; Stuart et al., 1996; Armson and Kiefte, 2008), prolonged speech that uses slow and exaggerated speech production (e.g., O'Brian et al., 2010), training of oral-motor co-ordination (e.g., Riley and Ingham, 2000), and the Lidcombe technique, a response-contingent program that involves parents to shape the child's utterances (e.g., Lattermann et al., 2008). The altered auditory feedback methods are considered to be effective to change the excessive reliance on auditory feedback control, while the other speech production trainings would help to reform feed-forward speech commands. Yet, a therapy that focuses specifically on stimulating the BG-thalamo-cortical circuit to enhance rhythmic speech production would also be warranted (Alm, 2004).

The SEP hypothesis assumes that individuals who stutter may benefit from rhythm-based therapy using synchronization and entrainment to a pulse for stimulating the BG-thalamo-cortical circuit. Behavioral studies support this notion by showing that the presence of rhythmic auditory signals such as metronome beats, when synchronized with speech production, induces strong fluency-enhancing effects in individuals who stutter (e.g., Brady, 1969, 1971). A recent fMRI study also supports this notion by showing that BG activities of stuttering speakers increased to the level of normal speech controls when speaking with the metronome beats (Toyomura et al., 2011).

Nevertheless, given that stuttering is a relapse-prone disorder (Craig, 2002), long-term management strategies are likely to be useful when dealing with this disorder over a lifetime. Accordingly, future studies need to test how long the metronome-induced fluency sustains after removing the rhythmic sounds. It has been suggested that the BG-thalamo-SMA circuit is dominant for self-initiation of speech, while the PMC-thalamo-cerebellar circuit is dominant for externally cued speech (Alm, 2004). Therefore, one of the challenges for future studies is to transition from the externally cued (PMC-centered) speech to self-initiated (SMA-centered) speech in the treatments using metronome-guided cues. To engage the BG-thalamo-SMA circuit, it may be useful to use non-isochronous metronome stimuli to promote the patients to find a pulse and initiate rhythmic speech by themselves. In addition, future studies need to test whether structural and functional reorganization occur in the BG-thalamo-cortical circuit after the intervention using the rhythmic auditory cues. Concurrently, more basic studies are needed to clarify whether the neural underpinnings of stuttering overlap with those of rhythm processing in music. Investigations of the abilities of musical rhythm processing in individuals who stutter using an amusia battery (e.g., Peretz et al., 2003; Fujii and Schlaug, 2013) may help further our understanding of possible overlapping brain networks. In addition, there is a need to test whether synchronization to a pulse in music have the similar fluency-enhancing effect for stuttering speakers compared with the synchronization to a metronome.

APHASIA

Aphasia is a common and devastating consequence of stroke or other brain injuries that results in language-related dysfunction. When speech production is impaired, the patients are broadly classified into the category of "non-fluent aphasia." In such cases, a lesion in the left posterior frontal region (Broca's area) is often observed. Many patients with large left hemisphere lesions have poor prognosis, despite having received years of intensive speech therapy (Lazar et al., 2010). However, emerging evidence suggests that some techniques have the potential to improve the verbal communication skills of these patients, as well as to reorganize the underlying neural processes related to language. For example, inspired by the clinical observation that patients with non-fluent aphasia can sing words even though they are unable to speak (Gerstmann, 1964; Yamadori et al., 1977), melodic intonation therapy (MIT) has received much research attention over the past few years. The main components of this speech therapy technique are (1) melodic intonation, (2) the use of formulaic phrases and sentences, and (3) slow and periodic verbalization with left-hand tapping (Schlaug et al., 2008, 2010). Within a therapy session, the therapist instructs the patient to intone (or "sing") simple phrases while slowly tapping their left hand with each syllable.

Emerging evidence involving open-label studies has revealed some positive treatment effects (Wilson et al., 2006; Schlaug et al., 2008; Wan et al., 2014). However, the question of "why" MIT works remains the subject of intense debate. The contribution of singing

is supported by the neuroimaging findings of right hemisphere lateralization of singing processing when compared to speaking (e.g., Ozdemir et al., 2006), as well as by studies showing reorganization of right hemisphere structure and function following therapy (e.g., Schlaug et al., 2008). However, it is important to note that the latter studies often included patients with very large lesions that sometimes cover most of the left hemisphere, thus precluding analysis of language-related areas within that hemisphere. In addition, although melodic intonation is usually emphasized as a major difference of singing compared with speaking, the difference between singing rhythm and speaking rhythm has been overlooked (see **Figure 1** as an example).

Recent studies have highlighted the potential role of rhythm in aphasia treatment. For example, aphasia recovery, as denoted by correct syllable production, was examined by comparing singing therapy, rhythmic therapy, and standard speech therapy (Stahl et al., 2011, 2013). The results showed that, when compared to singing therapy, the rhythmic therapy was similarly effective (Stahl et al., 2011, 2013). Moreover, patients with lesions that cover the BG were found to be highly dependent on the external rhythmic cues (Stahl et al., 2011). Taken together, this study highlights the role of rhythm in aphasia recovery.

The SEP hypothesis postulates that the rhythmic components (e.g., singing rhythm, left-hand tapping) of MIT can help to facilitate sound envelope processing and synchronization and entrainment to a pulse. That is, the predictability of formulaic phrases and sentences requires precise encoding of pulse or periodic timing of vocalizations, while left-hand tapping can facilitate synchronization and entrainment to the pulse. Thus, under the SEP framework, MIT may be interpreted as an effective way to engage (a) the auditory afferent circuit to encourage precise encoding of sounds, (b) the subcortical–prefrontal circuit to motivate patients, (c) the BG-thalamo-cortical circuit to facilitate beat-based timing process, and (d) the efferent cortical motor circuit to promote the motor output (**Figure 2D**).

A rationale for MIT is the potential to engage and unmask language-capable regions in the unaffected right hemisphere such as the structural reorganization of arcuate fasciculus, a fiber bundle connecting the posterior superior temporal region and the posterior inferior frontal region (Schlaug et al., 2008, 2010; Wan et al., 2014). The SEP hypothesis assumes that sound envelope processing may be lateralized to the temporal cortex in the right hemisphere (**Figure 2B**). If MIT engages high demands on the right temporal cortex to encode sound envelope precisely, it may also increase the connectivity from the right temporal cortex to the right inferior frontal gyrus (IFG). Importantly, rhythmic therapy in aphasia patients with left basal ganglia lesion resulted in improved production of common formulaic phrases that are known to be supported by right BG-thalamo-cortical network (Stahl et al., 2013), suggesting that rhythm therapy for aphasia might also induce alterations in the right BG-thalamo-cortical network. The left-hand tapping in MIT might be also interpreted as a way to recruit enlarged involvement of contralateral right motor areas (i.e., dorsal and ventral portions of the right M1, PMC, and PFC) and thereby facilitate motor output of the unaffected hemisphere.

AUTISM

One of the core features of autism spectrum disorder (ASD) is impairment in language and communication. For children with ASD, the ability to speak early is associated with improved quality of life. Research has reported the presence of motor and oral-motor impairments in ASD children who have expressive language deficits (Belmonte et al., 2013; McCleery et al., 2013).

To date, very few interventions have specifically targeted the oral-motor aspects in ASD. One is the prompts for restructuring oral muscular phonetic targets (PROMPT) model, which is a play-based technique that involves vocal modeling and physical manipulations of the children's oral-motor system to facilitate the production of a speech target (Chumpelik, 1984). A pilot study reported speech improvements in five non-verbal children with ASD after receiving PROMPT intervention (Rogers et al., 2006). Another therapy technique that incorporates a motor component is auditory-motor mapping training (AMMT), which is an active multisensory therapy designed to facilitate speech output in completely non-verbal children with autism (Wan et al., 2010). This technique aims to promote speech production directly by training the association between speech sounds and articulatory actions using slow and melodic intonating vocalizations with bimanual motor activities (Wan et al., 2010). While some of the components of AMMT overlap with those of MIT in phasia, a unique aspect of AMMT is the use of a set of tuned drums and bimanual motor actions to facilitate sound-motor mapping. An initial proof-of-concept study indicated the therapeutic potential of AMMT in facilitating speech development in autism (Wan et al., 2011).

Similar to MIT in aphasia, we assume that AMMT can be categorized into non-rhythmic and rhythmic components. The former relates to intoned vocalizations, and the latter relates to spoken syllables being linked with the bimanual motor actions on the tuned drums. Under the SEP framework, the rhythmic component in AMMT can be useful in a number of ways. First, perception of rhythmic drumming and vocal sounds may stimulate the auditory afferent circuit for the precise encoding of sound envelope or temporal events. Indeed, it has been shown that ASD is associated with developmental abnormalities in the brainstem and cerebellum *in utero*, which can lead to abnormal timing and sensory perception in ASD (Trevarthen and Delafield-Butt, 2013). Second, synchronization and entrainment of rhythmic vocalizations and bimanual motor actions may be effective to stimulate the speech motor and language networks in ASD. In the DIVA model, the left vPMC and pIFG are involved in both speech production and sensorimotor mapping and have functional correspondence to the mirror-neuron system (Rizzolatti et al., 1996; Kohler et al., 2002; Guenther et al., 2006; Tourville and Guenther, 2011). A number of neuroimaging studies have suggested that ASD is associated with abnormalities in the IFG and posterior superior temporal sulcus (pSTG) (Herbert et al., 2002; De Fosse et al., 2004; Kleinhans et al., 2008; Mengotti et al., 2011; McCleery et al., 2013). Moreover, another neuroimaging study suggests that motor dysfunction in ASD is associated with abnormality in BG-thalamo-SMA circuit (Enticott et al., 2009). Thus, synchronization and entrainment of rhythmic vocalizations and bimanual motor actions in AMMT

may help ameliorate speech production and sensorimotor mapping deficits in ASD by engaging the BG-thalamo-cortical (SMA, PMC, IFG, and STG) circuit.

CONCLUSION

In this paper, we consider the role of rhythm in speech and language rehabilitation. The emerging research field of music and neuroscience led us to propose the SEP hypothesis, which postulates that (1) sound envelope processing and (2) synchronization and entrainment to a pulse, may help to stimulate brain networks for human communication. Within the SEP framework, we present four possible circuits that may help to stimulate the brain networks underlying human communication: (i) the auditory afferent circuit consisted of brainstem, thalamus, cerebellum, and temporal cortex for precise encoding of sound envelope and temporal events; (ii) the subcortical–prefrontal circuit for emotional and reward-related processing; (iii) the BG-thalamo-cortical circuit for processing beat-based timing; and (iv) the cortical motor efferent circuit for motor output. We hope that future studies combining neuroimaging techniques and randomized control designs with the SEP framework will help to evaluate the efficacy of the rhythm-based therapies.

ACKNOWLEDGMENTS

Shinya Fujii was supported by a fellowship from the Japan Society for the Promotion of Science (JSPS). Catherine Y. Wan was supported by a fellowship from the Nancy Lurie Marks Family Foundation.

REFERENCES

Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.* 28, 3958–3965. doi:10.1523/JNEUROSCI.0187-08.2008

Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2009). Abnormal cortical processing of the syllable rate of speech in poor readers. *J. Neurosci.* 29, 7686–7693. doi:10.1523/JNEUROSCI.5242-08.2009

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13367–13372. doi:10.1073/pnas.201400998

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. *J. Commun. Disord.* 37, 325–369. doi:10.1016/j.jcomdis.2004.03.001

Armson, J., and Kiefte, M. (2008). The effect of SpeechEasy on stuttering frequency, speech rate, and speech naturalness. *J. Fluency Disord.* 33, 120–134. doi:10.1016/j.jfludis.2008.04.002

Belmonte, M. K., Saxena-Chandhok, T., Cherian, R., Muneer, R., George, L., and Karanth, P. (2013). Oral motor deficits in speech-impaired children with autism. *Front. Integr. Neurosci.* 7:47. doi:10.3389/fnint.2013.00047

Benoit, C. E., Dalla Bella, S., Farrugia, N., Obrig, H., Mainka, S., and Kotz, S. A. (2014). Musically cued gait-training improves both perceptual and motor timing in Parkinson's disease. *Front. Hum. Neurosci.* 8:494. doi:10.3389/fnhum.2014.00494

Bertoncini, J., Nazzi, T., Cabrera, L., and Lorenzi, C. (2011). Six-month-old infants discriminate voicing on the basis of temporal envelope cues (L). *J. Acoust. Soc. Am.* 129, 2761–2764. doi:10.1121/1.3571424

Blandini, F., Nappi, G., Tassorelli, C., and Martignoni, E. (2000). Functional changes of the basal ganglia circuitry in Parkinson's disease. *Prog. Neurobiol.* 62, 63–88. doi:10.1016/S0301-0082(99)00067-2

Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11818–11823. doi:10.1073/pnas.191355898

Bloodstein, O. (1995). *A Handbook of Stuttering.* San Diego, CA: Singular Publishing Group, Inc.

Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395. doi:10.1038/nn1409

Bohland, J. W., Bullock, D., and Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *J. Cogn. Neurosci.* 22, 1504–1529. doi:10.1162/jocn.2009.21306

Bohland, J. W., and Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage* 32, 821–841. doi:10.1016/j.neuroimage.2006.04.173

Brady, J. P. (1969). Studies on the metronome effect on stuttering. *Behav. Res. Ther.* 7, 197–204. doi:10.1016/0005-7967(69)90033-3

Brady, J. P. (1971). Metronome-conditioned speech retraining for stuttering. *Behav. Ther.* 2, 129–150. doi:10.1016/S0005-7894(71)80001-1

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., and Fox, P. T. (2005). Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. *Hum. Brain Mapp.* 25, 105–117. doi:10.1002/hbm.20140

Brown, S., Ngan, E., and Liotti, M. (2008). A larynx area in the human motor cortex. *Cereb. Cortex* 18, 837–845. doi:10.1093/cercor/bhm131

Buchsbaum, B. R., Hickok, G., and Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.* 25, 663–678. doi:10.1207/s15516709cog2505_2

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5:e1000436. doi:10.1371/journal.pcbi.1000436

Chang, S. E., and Zhu, D. C. (2013). Neural network connectivity differences in children who stutter. *Brain* 136, 3709–3726. doi:10.1093/brain/awt275

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008a). Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex* 18, 2844–2854. doi:10.1093/cercor/bhn042

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008b). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* 20, 226–239. doi:10.1162/jocn.2008.20018

Chen, J. L., Zatorre, R. J., and Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage* 32, 1771–1781. doi:10.1016/j.neuroimage.2006.04.207

Chumpelik, D. (1984). The PROMPT system of therapy: theoretical framework and applications for developmental apraxia of speech. *Semin. Speech Lang.* 5, 139–155. doi:10.1055/s-0028-1085172

Civier, O., Bullock, D., Max, L., and Guenther, F. H. (2013). Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. *Brain Lang.* 126, 263–278. doi:10.1016/j.bandl.2013.05.016

Civier, O., Tasko, S. M., and Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *J. Fluency Disord.* 35, 246–279. doi:10.1016/j.jfludis.2010.05.002

Conard, N. J., Malina, M., and Munzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460, 737–740. doi:10.1038/nature08169

Condon, W. S., and Sander, L. W. (1974). Neonate movement is synchronized with adult speech: interactional participation and language acquisition. *Science* 183, 99–101. doi:10.1126/science.183.4120.99

Corriveau, K. H., and Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *Cortex* 45, 119–130. doi:10.1016/j.cortex.2007.09.008

Craig, A. R. (2002). Fluency outcomes following treatment for those who stutter. *Percept. Mot. Skills* 94, 772–774. doi:10.2466/PMS.94.2.772-774

Crystal, T. H., and House, A. S. (1982). Segmental durations in connected speech signals: preliminary results. *J. Acoust. Soc. Am.* 72, 705–716. doi:10.1121/1.388251

Danielsen, A., Otnaess, M. K., Jensen, J., Williams, S. C., and Ostberg, B. C. (2014). Investigating repetition and change in musical rhythm by functional MRI. *Neuroscience* 275, 469–476. doi:10.1016/j.neuroscience.2014.06.029

Darley, F. L., Aronson, A. E., and Brown, J. R. (1969). Clusters of deviant speech dimensions in the dysarthrias. *J. Speech Hear. Res.* 12, 462–496. doi:10.1044/jshr.1203.462

Davie, C. A. (2008). A review of Parkinson's disease. *Br. Med. Bull.* 86, 109–127. doi:10.1093/bmb/ldn013

de Bruin, N., Doan, J. B., Turnbull, G., Suchowersky, O., Bonfield, S., Hu, B., et al. (2010). Walking with music is a safe and viable tool for gait training in Parkinson's disease: the effect of a 13-week feasibility study on single and dual task walking. *Parkinsons Dis.* 2010, 483530. doi:10.4061/2010/483530

De Fosse, L., Hodge, S. M., Makris, N., Kennedy, D. N., Caviness, V. S. Jr., McGrath, L., et al. (2004). Language-association cortex asymmetry in autism and specific language impairment. *Ann. Neurol.* 56, 757–766. doi:10.1002/ana.20275

DeLong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends Neurosci.* 13, 281–285. doi:10.1016/0166-2236(90)90110-V

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature* 384, 159–161. doi:10.1038/384159a0

Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., and Cabanis, E. A. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain* 130, 1432–1441. doi:10.1093/brain/awm042

Drullman, R., Festen, J. M., and Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.* 95, 2670–2680. doi:10.1121/1.408467

Drullman, R., Festen, J. M., and Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064. doi:10.1121/1.408467

Duffy, J. R. (2005). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. St. Louis, MO: Elsevier Mosby.

Elliott, T. M., and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* 5:e1000302. doi:10.1371/journal.pcbi.1000302

Enticott, P. G., Bradshaw, J. L., Iansek, R., Tonge, B. J., and Rinehart, N. J. (2009). Electrophysiological signs of supplementary-motor-area deficits in high-functioning autism but not Asperger syndrome: an examination of internally cued movement-related potentials. *Dev. Med. Child Neurol.* 51, 787–791. doi:10.1111/j.1469-8749.2009.03270.x

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi:10.1016/j.cognition.2005.11.009

Fitch, W. T. (2013). Speech science: tuned to the rhythm. *Nature* 494, 434–435. doi:10.1038/494434a

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., et al. (1996). A PET study of the neural systems of stuttering. *Nature* 382, 158–161. doi:10.1038/382158a0

Freeman, W. J. (2000). "A neurobiological role of music in social bonding," in *The Origins of Music*, eds N. Wallin, B. Merkur, and S. Brown (Cambridge MA: MIT Press), 411–424.

Fujii, S., and Schlaug, G. (2013). The Harvard beat assessment test (H-BAT): a battery for assessing beat perception and production and their dissociation. *Front. Hum. Neurosci.* 7:771. doi:10.3389/fnhum.2013.00771

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., and Taga, G. (2014). Precursors of dancing and singing to music in three- to four-months-old infants. *PLoS ONE* 9:e97680. doi:10.1371/journal.pone.0097680

Galvan, A., and Wichmann, T. (2008). Pathophysiology of parkinsonism. *Clin. Neurophysiol.* 119, 1459–1474. doi:10.1016/j.clinph.2008.03.017

Gerstmann, H. L. (1964). A case of aphasia. *J. Speech Hear. Disord.* 29, 89–91. doi:10.1044/jshd.2901.89

Ghazanfar, A. A., Morrill, R. J., and Kayser, C. (2013). Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1959–1963. doi:10.1073/pnas.1214956110

Ghazanfar, A. A., and Takahashi, D. Y. (2014). Facial expressions and the evolution of the speech rhythm. *J. Cogn. Neurosci.* 26, 1196–1207. doi:10.1162/jocn_a_00575

Ghazanfar, A. A., Takahashi, D. Y., Mathur, N., and Fitch, W. T. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Curr. Biol.* 22, 1176–1182. doi:10.1016/j.cub.2012.04.055

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., et al. (2000). Representation of the temporal envelope of sounds in the human brain. *J. Neurophysiol.* 84, 1588–1598.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* 15, 3–10. doi:10.1016/j.tics.2010.10.001

Grahn, J. A., and Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906. doi:10.1162/jocn.2007.19.5.893

Grahn, J. A., and Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* 29, 7540–7548. doi:10.1523/JNEUROSCI.2018-08.2009

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech: a syllable-centric perspective. *J. Phon.* 31, 46485. doi:10.1016/j.wocn.2003.09.005

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* 96, 280–301. doi:10.1016/j.bandl.2005.06.001

Hargrave, S., Kalinowski, J., Stuart, A., Armson, J., and Jones, K. (1994). Effect of frequency-altered feedback on stuttering frequency at normal and fast speech rates. *J. Speech Hear. Res.* 37, 1313–1319. doi:10.1044/jshr.3706.1313

Hasegawa, A., Okanoya, K., Hasegawa, T., and Seki, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Sci. Rep.* 1, 120. doi:10.1038/srep00120

Herbert, M. R., Harris, G. J., Adrien, K. T., Ziegler, D. A., Makris, N., Kennedy, D. N., et al. (2002). Abnormal asymmetry in language association cortex in autism. *Ann. Neurol.* 52, 588–596. doi:10.1002/ana.10349

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi:10.1016/j.cognition.2003.10.011

Hillis, A. E., Work, M., Barker, P. B., Jacobs, M. A., Breese, E. L., and Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. *Brain* 127, 1479–1487. doi:10.1093/brain/awh172

Ho, A. K., Bradshaw, J. L., Cunnington, R., Phillips, J. G., and Iansek, R. (1998). Sequence heterogeneity in parkinsonian speech. *Brain Lang.* 64, 122–145. doi:10.1006/brln.1998.1959

Hove, M. J., Fairhurst, M. T., Kotz, S. A., and Keller, P. E. (2013). Synchronizing with auditory and visual rhythms: an fMRI assessment of modality differences and modality appropriateness. *Neuroimage* 67, 313–321. doi:10.1016/j.neuroimage.2012.11.032

Huang, C. M., Liu, G., and Huang, R. (1982). Projections from the cochlear nucleus to the cerebellum. *Brain Res.* 244, 1–8. doi:10.1016/0006-8993(82)90897-6

Hutchinson, S., Lee, L. H., Gaab, N., and Schlaug, G. (2003). Cerebellar volume of musicians. *Cereb. Cortex* 13, 943–949. doi:10.1093/cercor/13.9.943

Iverson, J. M., and Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature* 396, 228. doi:10.1038/24300

Jonas, S. (1981). The supplementary motor region and speech emission. *J. Commun. Disord.* 14, 349–373. doi:10.1016/0021-9924(81)90019-8

Jurgens, U. (1984). The efferent and afferent connections of the supplementary motor area. *Brain Res.* 300, 63–81. doi:10.1016/0006-8993(84)91341-6

Kalaska, J. F., Cohen, D. A., Hyde, M. L., and Prud'homme, M. (1989). A comparison of movement direction-related versus load direction-related activity in primate motor cortex, using a two-dimensional reaching task. *J. Neurosci.* 9, 2080–2102.

Kell, C. A., Neumann, K., Von Kriegstein, K., Posenenske, C., Von Gudenberg, A. W., Euler, H., et al. (2009). How the brain repairs stuttering. *Brain* 132, 2747–2760. doi:10.1093/brain/awp185

Kent, R. D., and Tjaden, K. (1997). "Brain functions underlying speech," in *Handbook of Phonetic Sciences*, eds W. J. Hardcastle and J. Laver (Oxford: Blackwell), 220–255.

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. doi:10.1016/j.jecp.2008.07.005

Kleinhans, N. M., Muller, R. A., Cohen, D. N., and Courchesne, E. (2008). Atypical functional lateralization of language in autism spectrum disorders. *Brain Res.* 1221, 115–125. doi:10.1016/j.brainres.2008.04.080

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi:10.1038/nrn3666

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848. doi:10.1126/science.1070311

Kotz, S. A., and Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn. Sci.* 14, 392–399. doi:10.1016/j.tics.2010.06.005

Kotz, S. A., and Schwartze, M. (2011). Differential input of the supplementary motor area to a dedicated temporal processing network: functional and clinical implications. *Front. Integr. Neurosci.* 5:86. doi:10.3389/fnint.2011.00086

Kotz, S. A., Schwartze, M., and Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex* 45, 982–990. doi:10.1016/j.cortex.2009.02.010

Kung, S. J., Chen, J. L., Zatorre, R. J., and Penhune, V. B. (2013). Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *J. Cogn. Neurosci.* 25, 401–420. doi:10.1162/jocn_a_00325

Laruelle, M. (2000). Imaging synaptic neurotransmission with in vivo binding competition techniques: a critical review. *J. Cereb. Blood Flow Metab.* 20, 423–451. doi:10.1097/00004647-200003000-00001

Lattermann, C., Euler, H. A., and Neumann, K. (2008). A randomized control trial to investigate the impact of the Lidcombe program on early stuttering in German-speaking preschoolers. *J. Fluency Disord.* 33, 52–65. doi:10.1016/j.jfludis.2007.12.002

Lazar, R. M., Minzer, B., Antoniello, D., Festa, J. R., Krakauer, J. W., and Marshall, R. S. (2010). Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke* 41, 1485–1488. doi:10.1161/STROKEAHA.109.577338

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38; discussion 38–75. doi:10.1017/S0140525X99001776

Levelt, W. J., and Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition* 50, 239–269. doi:10.1016/0010-0277(94)90030-2

Lewis, P. A., and Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Curr. Opin. Neurobiol.* 13, 250–255. doi:10.1016/S0959-4388(03)00036-9

Liotti, M., Ramig, L. O., Vogel, D., New, P., Cook, C. I., Ingham, R. J., et al. (2003). Hypophonia in Parkinson's disease: neural correlates of voice treatment revealed by PET. *Neurology* 60, 432–440. doi:10.1212/WNL.60.3.432

Lu, Y., Paraskevopoulos, E., Herholz, S. C., Kuchenbuch, A., and Pantev, C. (2014). Temporal processing of audiovisual stimuli is enhanced in musicians: evidence from magnetoencephalography (MEG). *PLoS ONE* 9:e90686. doi:10.1371/journal.pone.0090686

Ludlow, C. L. (2005). Central nervous system control of the laryngeal muscles in humans. *Respir. Physiol. Neurobiol.* 147, 205–222. doi:10.1016/j.resp.2005.04.015

Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi:10.1016/j.neuron.2007.06.004

Luppino, G., Matelli, M., Camarda, R., and Rizzolatti, G. (1993). Corticocortical connections of area F3 (SMA-proper) and area F6 (pre-SMA) in the macaque monkey. *J. Comp. Neurol.* 338, 114–140. doi:10.1002/cne.903380109

Malecot, A., Johnston, R., and Kizziar, P. A. (1972). Syllabic rate and utterance length in French. *Phonetica* 26, 235–251. doi:10.1159/000259414

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., and Wallace, M. E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: a theoretical model of stuttering. *Contemp. Issues Commun. Sci. Disord.* 31, 105–122.

McCleery, J. P., Elliott, N. A., Sampanis, D. S., and Stefanidou, C. A. (2013). Motor development and motor resonance difficulties in autism: relevance to early intervention for language and communication skills. *Front. Integr. Neurosci.* 7:30. doi:10.3389/fnint.2013.00030

McGettigan, C., and Scott, S. K. (2012). Cortical asymmetries in speech perception: what's wrong, what's right and what's left? *Trends Cogn. Sci.* 16, 269–276. doi:10.1016/j.tics.2012.04.006

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi:10.1038/264746a0

McIntosh, G. C., Brown, S. H., Rice, R. R., and Thaut, M. H. (1997). Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease. *J. Neurol. Neurosurg. Psychiatr.* 62, 22–26. doi:10.1136/jnnp.62.1.22

Meister, I. G., Boroojerdi, B., Foltys, H., Sparing, R., Huber, W., and Topper, R. (2003). Motor cortex hand area and speech: implications for the development of language. *Neuropsychologia* 41, 401–406. doi:10.1016/S0028-3932(02)00179-3

Meister, I. G., Buelte, D., Staedtgen, M., Boroojerdi, B., and Sparing, R. (2009a). The dorsal premotor cortex orchestrates concurrent speech and fingertapping movements. *Eur. J. Neurosci.* 29, 2074–2082. doi:10.1111/j.1460-9568.2009.06729.x

Meister, I. G., Weier, K., Staedtgen, M., Buelte, D., Thirugnanasambandam, N., and Sparing, R. (2009b). Covert word reading induces a late response in the hand motor system of the language dominant hemisphere. *Neuroscience* 161, 67–72. doi:10.1016/j.neuroscience.2009.03.031

Meltzoff, A. N., and Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science* 198, 75–78. doi:10.1126/science.198.4312.75

Mengotti, P., D'Agostini, S., Terlevic, R., De Colle, C., Biasizzo, E., Londero, D., et al. (2011). Altered white matter integrity and development in children with autism: a combined voxel-based morphometry and diffusion imaging study. *Brain Res. Bull.* 84, 189–195. doi:10.1016/j.brainresbull.2010.12.002

Menon, V., and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage* 28, 175–184. doi:10.1016/j.neuroimage.2005.05.053

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. London: Weidenfeld & Nicolson.

Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898. doi:10.1073/pnas.0701498104

Musacchia, G., Strait, D., and Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* 241, 34–42. doi:10.1016/j.heares.2008.04.013

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 756–766. doi:10.1037/0096-1523.24.3.756

Nombela, C., Hughes, L. E., Owen, A. M., and Grahn, J. A. (2013). Into the groove: can rhythm influence Parkinson's disease? *Neurosci. Biobehav. Rev.* 37, 2564–2570. doi:10.1016/j.neubiorev.2013.08.003

Obermeier, C., Menninghaus, W., Von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., et al. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Front. Psychol.* 4:10. doi:10.3389/fpsyg.2013.00010

O'Brian, S., Packman, A., and Onslow, M. (2010). "The camperdown program," in *Treatment of Stuttering: Established and Emerging Interventions*, eds B. Guitar and R. McCauley (Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins), 256–276.

Olthoff, A., Baudewig, J., Kruse, E., and Dechent, P. (2008). Cortical sensorimotor control in vocalization: a functional magnetic resonance imaging study. *Laryngoscope* 118, 2091–2096. doi:10.1097/MLG.0b013e31817fd40f

Ozdemir, E., Norton, A., and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage* 33, 628–635. doi:10.1016/j.neuroimage.2006.07.013

Pai, M. C. (1999). Supplementary motor area aphasia: a case report. *Clin. Neurol. Neurosurg.* 101, 29–32. doi:10.1016/S0303-8467(98)00068-7

Parbery-Clark, A., Strait, D. L., and Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia* 49, 3338–3345. doi:10.1016/j.neuropsychologia.2011.08.007

Parbery-Clark, A., Tierney, A., Strait, D. L., and Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience* 219, 111–119. doi:10.1016/j.neuroscience.2012.05.042

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychol.* 2:142. doi:10.3389/fpsyg.2011.00142

Patel, A. D. (2012). The OPERA hypothesis: assumptions and clarifications. *Ann. N. Y. Acad. Sci.* 1252, 124–128. doi:10.1111/j.1749-6632.2011.06426.x

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* 308, 98–108. doi:10.1016/j.heares.2013.08.011

Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830. doi:10.1016/j.cub.2009.03.038

Peelle, J. E., and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3:320. doi:10.3389/fpsyg.2012.00320

Penfield, W., and Rasmussen, T. (1950). *The Cerebral Cortex of Man: A Clinical Study of Localization of Function*. New York, NY: Macmillan.

Penfield, W., and Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press.

Peretz, I., Champod, A. S., and Hyde, K. (2003). Varieties of musical disorders. The montreal battery of evaluation of amusia. *Ann. N. Y. Acad. Sci.* 999, 58–75. doi:10.1196/annals.1284.006

Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691. doi:10.1038/nn1083

Pickett, E. R., Kuniholm, E., Protopapas, A., Friedman, J., and Lieberman, P. (1998). Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nucleus: a case study. *Neuropsychologia* 36, 173–188. doi:10.1016/S0028-3932(97)00065-1

Pinto, S., Thobois, S., Costes, N., Le Bars, D., Benabid, A. L., Broussolle, E., et al. (2004). Subthalamic nucleus stimulation and dysarthria in Parkinson's disease: a PET study. *Brain* 127, 602–615. doi:10.1093/brain/awh074

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as "symmetric sampling in time". *Speech Commun.* 41, 245–255. doi:10.1016/S0167-6393(02)00107-3

Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Leculier, L., et al. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology* 27, 121–131. doi:10.1037/a0031277

Ramig, L. O., Fox, C., and Sapir, S. (2004). Parkinson's disease: speech and voice disorders and their treatment with the Lee Silverman voice treatment. *Semin. Speech Lang.* 25, 169–180. doi:10.1055/s-2004-825653

Ramig, L. O., Fox, C., and Sapir, S. (2007). Speech disorders in Parkinson's disease and the effects of pharmacological, surgical and speech treatment with emphasis on Lee Silverman voice treatment (LSVT(R)). *Handb. Clin. Neurol.* 83, 385–399. doi:10.1016/S0072-9752(07)83017-X

Ramig, L. O., Fox, C., and Sapir, S. (2008). Speech treatment for Parkinson's disease. *Expert Rev. Neurother.* 8, 297–309. doi:10.1586/14737175.8.2.297

Ramig, L. O., Sapir, S., Countryman, S., Pawlas, A. A., O'Brien, C., Hoehn, M., et al. (2001). Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow up. *J. Neurol. Neurosurg. Psychiatr.* 71, 493–498. doi:10.1136/jnnp.71.4.493

Reddish, P., Fischer, R., and Bulbulia, J. (2013). Let's dance together: synchrony, shared intentionality and cooperation. *PLoS ONE* 8:e71182. doi:10.1371/journal.pone.0071182

Rektorova, I., Barrett, J., Mikl, M., Rektor, I., and Paus, T. (2007). Functional abnormalities in the primary orofacial sensorimotor cortex during speech in Parkinson's disease. *Mov. Disord.* 22, 2043–2051. doi:10.1002/mds.21548

Rektorova, I., Mikl, M., Barrett, J., Marecek, R., Rektor, I., and Paus, T. (2012). Functional neuroanatomy of vocalization in patients with Parkinson's disease. *J. Neurol. Sci.* 313, 7–12. doi:10.1016/j.jns.2011.10.020

Remedios, R., Logothetis, N. K., and Kayser, C. (2009). Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18010–18015. doi:10.1073/pnas.0909756106

Riley, G. D., and Ingham, J. C. (2000). Acoustic duration changes associated with two types of treatment for children who stutter. *J. Speech Lang. Hear. Res.* 43, 965–978. doi:10.1044/jslhr.4304.965

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.* 3, 131–141. doi:10.1016/0926-6410(95)00038-0

Rogers, S. J., Hayden, D., Hepburn, S., Charlifue-Smith, R., Hall, T., and Hayes, A. (2006). Teaching young nonverbal children with autism useful speech: a pilot study of the Denver model and PROMPT interventions. *J. Autism Dev. Disord.* 36, 1007–1024. doi:10.1007/s10803-006-0142-x

Ryan, B. P., and Van Kirk Ryan, B. (1995). Programmed stuttering treatment for children: comparison of two establishment programs through transfer, maintenance, and follow-up. *J. Speech Hear. Res.* 38, 61–75. doi:10.1044/jshr.3801.61

Sackley, C. M., Smith, C. H., Rick, C., Brady, M. C., Ives, N., Patel, R., et al. (2014). Lee Silverman voice treatment versus standard NHS speech and language therapy versus control in Parkinson's disease (PD COMM pilot): study protocol for a randomized controlled trial. *Trials* 15, 213. doi:10.1186/1745-6215-15-213

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262. doi:10.1038/nn.2726

Salimpoor, V. N., Van Den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., and Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219. doi:10.1126/science.1231059

Sapir, S., Ramig, L. O., and Fox, C. M. (2011). Intensive voice treatment in Parkinson's disease: Lee Silverman voice treatment. *Expert Rev. Neurother.* 11, 815–830. doi:10.1586/ern.11.43

Schachner, A., Brady, T. F., Pepperberg, I. M., and Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr. Biol.* 19, 831–836. doi:10.1016/j.cub.2009.03.061

Schlaug, G., Marchina, S., and Norton, A. (2008). From singing to speaking: why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music Percept.* 25, 315–323. doi:10.1525/mp.2008.25.4.315

Schlaug, G., Norton, A., Marchina, S., Zipse, L., and Wan, C. Y. (2010). From singing to speaking: facilitating recovery from nonfluent aphasia. *Future Neurol.* 5, 657–665. doi:10.2217/fnl.10.44

Schwartze, M., Keller, P. E., Patel, A. D., and Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behav. Brain Res.* 216, 685–691. doi:10.1016/j.bbr.2010.09.015

Sevdalis, V., and Keller, P. E. (2010). Cues for self-recognition in point-light displays of actions performed in synchrony with music. *Conscious. Cogn.* 19, 617–626. doi:10.1016/j.concog.2010.03.017

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304. doi:10.1126/science.270.5234.303

Singh-Curry, V., and Husain, M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia* 47, 1434–1448. doi:10.1016/j.neuropsychologia.2008.11.033

Skodda, S., Flasskamp, A., and Schlegel, U. (2010). Instability of syllable repetition as a model for impaired motor processing: is Parkinson's disease a "rhythm disorder"? *J. Neural Transm.* 117, 605–612. doi:10.1007/s00702-010-0390-y

Skodda, S., and Schlegel, U. (2008). Speech rate and rhythm in Parkinson's disease. *Mov. Disord.* 23, 985–992. doi:10.1002/mds.21996

Smith, Y., Wichmann, T., Factor, S. A., and DeLong, M. R. (2012). Parkinson's disease therapeutics: new developments and challenges since the introduction of levodopa. *Neuropsychopharmacology* 37, 213–246. doi:10.1038/npp.2011.212

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416, 87–90. doi:10.1038/416087a

Stahl, B., Henseler, I., Turner, R., Geyer, S., and Kotz, S. A. (2013). How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery. *Front. Hum. Neurosci.* 7:35. doi:10.3389/fnhum.2013.00035

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., and Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain* 134, 3083–3093. doi:10.1093/brain/awr240

Strait, D. L., and Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear. Res.* 308, 109–121. doi:10.1016/j.heares.2013.08.004

Strait, D. L., O'Connell, S., Parbery-Clark, A., and Kraus, N. (2014). Musicians' enhanced neural differentiation of speech sounds arises early in life: developmental evidence from ages 3 to 30. *Cereb. Cortex* 24, 2512–2521. doi:10.1093/cercor/bht103

Streifler, M., and Hofman, S. (1984). "Disorders of verbal expression in parkinsonism," in *Advances in Neurology*, eds R. Hassler and J. Christ (New York: Raven Press), 385–393.

Stuart, A., Kalinowski, J., Armson, J., Stenstrom, R., and Jones, K. (1996). Fluency effect of frequency alterations of plus/minus one-half and one-quarter octave shifts in auditory feedback of people who stutter. *J. Speech Hear. Res.* 39, 396–401. doi:10.1044/jshr.3902.396

Stupacher, J., Hove, M. J., Novembre, G., Schutz-Bosbach, S., and Keller, P. E. (2013). Musical groove modulates motor cortex excitability: a TMS investigation. *Brain Cogn.* 82, 127–136. doi:10.1016/j.bandc.2013.03.003

Teki, S., Grube, M., and Griffiths, T. D. (2011a). A unified model of time perception accounts for duration-based and beat-based timing mechanisms. *Front. Integr. Neurosci.* 5:90. doi:10.3389/fnint.2011.00090

Teki, S., Grube, M., Kumar, S., and Griffiths, T. D. (2011b). Distinct neural substrates of duration-based and beat-based auditory timing. *J. Neurosci.* 31, 3805–3812. doi:10.1523/JNEUROSCI.5561-10.2011

Thaut, M. H., McIntosh, G. C., Rice, R. R., Miller, R. A., Rathbun, J., and Brault, J. M. (1996). Rhythmic auditory stimulation in gait training for Parkinson's disease patients. *Mov. Disord.* 11, 193–200. doi:10.1002/mds.870110213

Tourville, J. A., and Guenther, F. H. (2011). The DIVA model: a neural theory of speech acquisition and production. *Lang. Cogn. Process.* 26, 952–981. doi:10.1080/01690960903498424

Tourville, J. A., Reilly, K. J., and Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39, 1429–1443. doi:10.1016/j.neuroimage.2007.09.054

Toyomura, A., Fujii, T., and Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. *Neuroimage* 57, 1507–1516. doi:10.1016/j.neuroimage.2011.05.039

Trevarthen, C., and Delafield-Butt, J. T. (2013). Autism as a developmental disorder in intentional movement and affective engagement. *Front. Integr. Neurosci.* 7:49. doi:10.3389/fnint.2013.00049

van Tricht, M. J., Smeding, H. M., Speelman, J. D., and Schmand, B. A. (2010). Impaired emotion recognition in music in Parkinson's disease. *Brain Cogn.* 74, 58–65. doi:10.1016/j.bandc.2010.06.005

Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., et al. (1998). Neural basis of an inherited speech and language disorder. *Proc. Natl. Acad. Sci. U.S.A.* 95, 12695–12700. doi:10.1073/pnas.95.21.12695

Wan, C. Y., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J., et al. (2011). Auditory-motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: a proof of concept study. *PLoS ONE* 6:e25505. doi:10.1371/journal.pone.0025505

Wan, C. Y., Demaine, K., Zipse, L., Norton, A., and Schlaug, G. (2010). From music making to speaking: engaging the mirror neuron system in autism. *Brain Res. Bull.* 82, 161–168. doi:10.1016/j.brainresbull.2010.04.010

Wan, C. Y., Zheng, X., Marchina, S., Norton, A., and Schlaug, G. (2014). Intensive therapy induces contralateral white matter changes in chronic stroke patients with Broca's aphasia. *Brain Lang.* 136C, 1–7. doi:10.1016/j.bandl.2014.03.011

Wang, X. F., Woody, C. D., Chizhevsky, V., Gruen, E., and Landeira-Fernandez, J. (1991). The dentate nucleus is a short-latency relay of a primary auditory transmission pathway. *Neuroreport* 2, 361–364. doi:10.1097/00001756-199107000-00001

Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J., et al. (2002). MRI analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain* 125, 465–478. doi:10.1093/brain/awf057

Wichmann, T., and DeLong, M. R. (1996). Functional and pathophysiological models of the basal ganglia. *Curr. Opin. Neurobiol.* 6, 751–758. doi:10.1016/S0959-4388(96)80024-9

Wichmann, T., and DeLong, M. R. (1998). Models of basal ganglia function and pathophysiology of movement disorders. *Neurosurg. Clin. N. Am.* 9, 223–236.

Wilson, S. J., Parsons, K., and Reutens, D. C. (2006). Preserved singing in aphasia: a case study of the efficacy of melodic intonation therapy. *Music Percept.* 24, 23–36. doi:10.1525/mp.2006.24.1.23

Witek, M. A., Clarke, E. F., Wallentin, M., Kringelbach, M. L., and Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLoS ONE* 9:e94446. doi:10.1371/journal.pone.0094446

Xi, M. C., Woody, C. D., and Gruen, E. (1994). Identification of short latency auditory responsive neurons in the cat dentate nucleus. *Neuroreport* 5, 1567–1570. doi:10.1097/00001756-199408150-00006

Yamadori, A., Osumi, Y., Masuhara, S., and Okubo, M. (1977). Preservation of singing in Broca's aphasia. *J. Neurol. Neurosurg. Psychiatr.* 40, 221–224. doi:10.1136/jnnp.40.3.221

Zatorre, R. J., Mondor, T. A., and Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *Neuroimage* 10, 544–554. doi:10.1006/nimg.1999.0491

Zentner, M., and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5768–5773. doi:10.1073/pnas.1000121107

Ziegler, W., Kilian, B., and Deger, K. (1997). The role of the left mesial frontal cortex in fluent speech: evidence from a case of left supplementary motor area hemorrhage. *Neuropsychologia* 35, 1197–1208. doi:10.1016/S0028-3932(97)00040-7

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 July 2014; accepted: 12 September 2014; published online: 13 October 2014.

Citation: Fujii S and Wan CY (2014) The role of rhythm in speech and language rehabilitation: the SEP hypothesis. Front. Hum. Neurosci. ***8****:777. doi: 10.3389/fnhum.2014.00777*

This article was submitted to the journal Frontiers in Human Neuroscience.