From singing to speaking: facilitating recovery from nonfluent aphasia

Gottfried Schlaug1, Andrea Norton1, Sarah Marchina1, Lauryn Zipse1 & Catherine Y Wan1

1Department of Neurology, Music, Neuroimaging & Stroke Recovery Laboratories, Beth Israel Deaconess Medical Center & Harvard Medical School, 330 Brookline Avenue, Boston, MA 02215, USA

It has been reported for more than 100 years that patients with severe nonfluent aphasia are better at singing lyrics than they are at speaking the same words. This observation led to the development of melodic intonation therapy (MIT). However, the efficacy of this therapy has yet to be substantiated in a randomized controlled trial. Furthermore, its underlying neural mechanisms remain unclear. The two unique components of MIT are the intonation of words and simple phrases using a melodic contour that follows the prosody of speech and the rhythmic tapping of the left hand that accompanies the production of each syllable and serves as a catalyst for fluency. Research has shown that both components are capable of engaging frontal–temporal regions in the right hemisphere, thereby making MIT particularly well suited for patients with large left hemisphere lesions who also suffer from nonfluent aphasia. Recovery from aphasia can happen in two ways: either through the recruitment of perilesional brain regions in the affected hemisphere, with variable recruitment of right-hemispheric regions if the lesion is small, or through the recruitment of homologous language and speech-motor regions in the unaffected hemisphere if the lesion of the affected hemisphere is extensive. Treatment-associated neural changes in patients undergoing MIT indicate that the unique engagement of right-hemispheric structures (e.g., the superior temporal lobe, primary sensorimotor, premotor and inferior frontal gyrus regions) and changes in the connections across these brain regions may be responsible for its therapeutic effect.

Of the estimated 750,000–800,000 new stroke cases occurring in the USA each year, approximately 25–50% present with some form of aphasia, an estimation that is based on studies performed in countries other than the USA [1,2]. Approximately 40% of these acute patients were available for follow-up at 1 year (attrition was due to death or inability to participate in trial). Approximately two-thirds of this 40% of patients from the original cohort showed abnormal scores on aphasia testing, with approximately a quarter of them being in the nonfluent category [3–4]. In right-handed individuals, nonfluent aphasia generally results from lesions in the left frontal lobe, including the portion of the left frontal lobe known as Broca’s region. The region is named after Paul Broca (1864), who first linked this area of the brain with nonfluent aphasia; this region is thought to consist of the posterior inferior frontal gyrus encompassing Brodmann’s areas 44 and 45. However, subsequent reports have shown that a wider array of lesions in the frontal lobes, the inferior perirolandic regions and in subcortical brain structures can also present as a clinical picture similar to that of Broca’s aphasia [5].

Patients with large left-hemispheric lesions that result in severe nonfluent aphasia typically do not show a good natural recovery from such an insult, nor do they appear to be as responsive to traditional speech therapy methods as patients with smaller lesions or other types of aphasia. Most post-stroke language interventions used in the subacute and chronic stroke phases are administered by speech therapists who evaluate the patient’s individual needs and then use a combination of techniques tailored to the individual patient’s impairment profile. At present, there are no universally accepted methods or a ‘gold standard’ for the treatment of severe nonfluent aphasia against which new or existing interventions can be compared, nor have any criteria been established for measuring meaningful treatment efficacy. Nevertheless, most therapists, clinicians and researchers in the aphasia field would probably agree that a treatment should be considered effective if a patient shows improvement in speech output that generalizes to untrained language structures and/or contexts [6].

To date, functional imaging of language recovery has largely focused on spontaneous or natural recovery without examining the neural effects of...
a particular intervention. Moreover, the neural mechanisms underlying post-stroke recovery continue to remain unclear, in particular for patients with large left-hemispheric lesions. Some studies emphasize the role of preserved language function in the left hemisphere [7,8], while others propose that language function is restored when right-hemisphere regions compensate for the loss [7,9–14]. Other studies report evidence of a bihemispheric role or even a right hemisphere only role in language recovery following an insult, particularly if the insult involves large parts of the left hemisphere [15–19]. Interestingly, only a few studies have examined the neural correlates of an aphasia treatment by contrasting pre- and post-therapy assessments [19–25]. The general consensus is that there are two routes to recovery. In patients with small lesions in the left hemisphere, there tends to be recruitment of both the left-hemispheric perilesional cortex, with variable involvement of right-hemispheric homologous regions during the recovery process [8,16,18,26]. In patients with large left-hemispheric lesions involving language-related regions of the fronto–temporal lobes, the only path to recovery may be through recruitment of homologous language and speech-motor regions in the right hemisphere [16,24]. It has been suggested that recovery via the right hemisphere may be less efficient than recovery via the left hemisphere [8,18], possibly due to patients with relatively large left-hemispheric lesions being generally more impaired and recovering to a lesser degree than patients with smaller left hemisphere lesions. Nevertheless, activation of right-hemispheric regions during speech/language functional MRI (fMRI) tasks has been reported in patients with aphasia, irrespective of their lesion size [16]. For patients with large lesions that cover the language-relevant regions on the left, therapies that specifically engage or stimulate the homologous right-hemispheric regions have the potential to facilitate the language recovery process beyond the limitations of natural recovery [24,25,27].

Intonation-based ‘speech therapy’ facilitates the role of right-hemispheric regions in recovery from nonfluent aphasia

Since the recovery from aphasia can be somewhat incomplete for patients with large left-hemispheric lesions, it is necessary to identify treatments that can better engage brain regions that might drive the recovery process and, ultimately, change the course of natural recovery through neural reorganization. Melodic intonation therapy (MIT) is an intonation-based treatment method for nonfluent or dysfluent aphasic patients that was developed in response to the observation that severely aphasic patients can often produce well-articulated, linguistically accurate words while singing, but not during speech [28–33]. MIT is a hierarchically structured treatment that uses intoned (sung) patterns that exaggerate the normal melodic content of speech across three levels of increasing difficulty. The intonation works by translating prosodic speech patterns (spoken phrases) into melodically intoned patterns using just two pitches. The higher pitch represents the syllables that would naturally be stressed (accented) during speech. At the simplest level, patients learn to intone (sing) a series of two-syllable words/phrases (e.g., ‘water’, ‘ice cream’ or ‘bathroom’) or simple two- or three-syllable social phrases (e.g., ‘thank you’ or ‘I love you’). As each level is mastered, patients move on to the next level, in which phrases gradually increase in length (e.g., ‘I am thirsty’ or ‘a cup of coffee, please’). Beyond the increased phrase length, the primary differences between the three levels of MIT lie in the way the treatment is administered and the level of support that is provided by the therapist.

Compared with nonintonation-based speech therapies, MIT contains two unique components: the melodic intonation (singing), with its inherent continuous voicing, and the rhythmic tapping of each syllable (using the patient’s left hand) while phrases are intoned and repeated. Since the initial account of its successful use in three chronic, nonfluent (Broca’s) aphasic patients [34], reports have outlined a comprehensive program of MIT [35–38] including strict patient selection criteria [37,38] and data that demonstrated significant improvement on the Boston Diagnostic Aphasia Examination (BDAE) [39] after treatment [40,41]. In a case study comparing MIT with a nonintonation-based control therapy [42], the authors found that MIT had a general facilitating effect on articulation and a longer-term effect on phrase production that was specifically attributed to its melodic component. However, the outcomes of that study were measured by the patient’s ability to produce practiced phrases prompted by the therapist, rather than by the transfer of language skills to untrained structures and/or contexts. In one of our previous studies, we compared two patients with similar speech output impairments and similar lesion sizes and locations that were subjected to MIT or a control intervention termed ‘speech repetition therapy’ [24]. Following treatment, both interventions yielded significant improvements.
in propositional speech that generalized to nonpracticed words and phrases; importantly, the MIT-treated patient gains surpassed those of the control-treated patients. Another case series of six patients treated with MIT showed a more than 200% improvement in a measure of spontaneous speech, correct information units (CIUs). In post- versus pre-therapy comparisons, the behavioral improvement was highly significant when compared with baseline variations in repeated test assessments. The change in CIUs showed a strong trend for a correlation (p = 0.08) with the change in the size of the right arcuate fasciculus (AF), a fiber bundle connecting temporal and frontal brain regions [25].

Unlike many therapies administered in the subacute and chronic phase that typically involve between one and three sessions per week, MIT is an intensive treatment program that requires a commitment of up to 1.5 h/day, 5 days/week over a period of several months (as recommended by the developers of MIT and the therapists currently administering the treatment) until the patient has mastered all three levels of MIT. In addition to its two unique components (i.e., intonation and left-hand tapping), there are several other components that not only play an important role in MIT, but are also used in other therapies. Among these components are the intensity of the therapy, the slow rate of vocalization (1 syllable/s) and an administration protocol that includes one-on-one sessions with a therapist who introduces and practices words/phrases using picture cues while giving continuous feedback. These features that MIT shares with other speech therapies must be carefully considered when the efficacy of MIT is tested against a control intervention [24].

The original interpretation of MIT’s path to successful recovery was that it engaged areas for articulation and speech output areas in the right hemisphere [34,40], although this has not been proven to date. Alternatively, MIT may exert its effect by either unmasking existing music/language connections in both hemispheres or by engaging preserved language-capable regions in either or both hemispheres. Since MIT incorporates both the melodic and rhythmic aspects of music [34,35,37,38,40,43,44], it may be unique in its potential for engaging not only auditory–motor regions on the right but also nonlesional regions in the affected left hemisphere. Belin et al. suggested that MIT-facilitated recovery is associated with the reactivation of left-hemisphere regions, most notably the left prefrontal cortex, just anterior to Broca’s region [45]. Although this publication was the first to examine patients treated with an MIT-like intervention using functional neuroimaging, their findings were surprising and somewhat contrary to the hypotheses that had been put forth by the original developers of MIT and those that have been using MIT in a clinical research setting [25,34,40,41]. It is interesting to note that although Belin and colleagues’ primary finding was an activation of left prefrontal regions when participants were asked to repeat intoned words, there is an important aspect of their study that is not often reported. In their analysis comparing the repetition of spoken words with the hearing of those words, they found blood flow changes that occurred predominantly in the right hemisphere (including the right temporal lobe and the right central operculum), which is consistent with some of our findings [24,25].

Figure 1 shows fMRI activation maps (superimposed onto the surface projections of a spatially standardized normal brain) of one of our other patients treated with MIT. The results of two imaging contrasts are shown: ‘overt speaking versus silence (control condition)’ and ‘overt speaking versus vowel production’ (p < 0.05 family-wise error) before (Figure 1A) and after (Figure 1B) therapy. Furthermore, a direct voxel-by-voxel comparison of the two acquisitions is shown in Figure 1C. The color codes represent different magnitudes of activation: the color yellow indicates stronger activation than the color red. The pronounced differences in activation observed in the post- versus pre-therapy comparison shows that there is more activation in the right temporal, premotor and posterior inferior frontal region after therapy. For more details on the fMRI tasks and the analysis of these fMRI datasets, see [46].

Figure 2 shows a diffusion tensor imaging study of one of our patients before and after 75 sessions of MIT. The AF is a fiber tract that connects the superior temporal lobe with the posterior inferior frontal gyrus in a reciprocal manner and it is typically not as strongly developed in the right hemisphere as it is in the left hemisphere in right-handed individuals [47]. The treatment-induced increase is evident when the pre- (Figure 2A) and post-therapy (Figure 2B) images are compared. The AF is very important for auditory–motor mapping and therefore plays a crucial role in language development. Furthermore, it may well be the remodeling of this fiber tract as a result of the intense therapy that supports long-term therapy effects in severely nonfluent patients [25]. The most likely explanation for these imaging results are changes in myelination [48–50]; however, changes
in axon diameter owing to changes in myelin or axon density, themselves owing to axonal sprouting, are also possible [51–53]. In patients with large left-hemispheric lesions, the right AF may play a crucial role in facilitating the mapping of sounds to motor actions and its feedback control. However, the right AF is usually not as well developed as the left AF in right-handed individuals with left-hemispheric language dominance [47]. However, as illustrated by our data, the right AF can be modified to support and facilitate speech output in patients with nonfluent aphasia.

**Assessing behavioral & brain effects of experimental aphasia therapy studies**

Most patients undergo a battery of assessments, such as the BDAE [39], to:

- Help classify an aphasic disorder as fluent or nonfluent and specific aphasic syndromes;
- Assess the severity of impairment in comprehension, naming, repetition, reading and writing.

However, the BDAE may not be ideal for assessing an improvement in speech production in a quantitative and objective way in patients with moderate-to-severe nonfluent aphasia [6,24]. Several studies have used investigator-assembled testing batteries to quantitatively measure changes in speech production or improvement in spontaneous speech in response to an intervention. Among those measures are:

- Conversational interviews: eliciting patients’ responses to questions regarding biographical data, medical history, daily activities, descriptions of familiar procedures (e.g., cooking a favorite dish, working on a hobby or doing routine repair work), etc.;
- Descriptions of complex pictures;
- Naming tasks (i.e., naming frequent and infrequent picture items) [54,55].

Patient responses on the conversational interview, the description of common and well known (to the patient) procedures and the description of complex pictures can be quantified by calculating the average number of CIUs per minute and the average number of syllables per phrase [56]. All meaningless utterances, inappropriate exclamations, incorrect responses (inaccurate information) and/or perseverations are excluded prior to scoring.

Functional imaging studies have used a variety of experimental paradigms to examine the effects in the brain of therapeutic interventions ranging from covert to overt naming and word-stem completion tasks [16]. We have used a list of
bisyllabic words/phrases that patients are capable of saying at baseline and then repeated that same list of words/phrases at all imaging time-points prior to and after the therapeutic intervention. It is important to note that the rate of speaking/singing remains the same for all assessments, so true treatment-induced changes in the brain network can be detected rather than changes in the rate of production. Furthermore, in order to isolate the neural mechanisms that change in response to therapy and distinguish them from those that control basic sensorimotor operations of articulation, the use of an appropriate control condition is critically important (Figure 1). We estimate that between eight and ten fMRI datasets would be needed to show significant within-group fMRI changes over time and between ten and 12 fMRI datasets to show between-group differences in fMRI changes over time. We are using a fMRI method referred to as sparse temporal sampling that has been previously shown to be ideal for overt vocalization tasks in the scanner environment since it uses the natural delay in the neurovascular coupling to separate the period of greatest movement from the period of speaking-induced neural activity changes [46,57,58].

**Possible mechanisms explaining the effects of an intonation-based speech therapy**

The traditional explanation for the dissociation between speaking and singing in aphasic patients is the presence of two routes for word articulation: one for spoken words through the brain's left hemisphere and a separate route for sung words that uses either the right or both hemispheres. The small amount of empirical data available supports a bihemispheric role in the execution and sensorimotor control of vocal production for both speaking and singing [46,59–62], with a tendency for greater left-lateralization for speaking under normal physiological conditions (i.e., faster rates of production during speaking than singing). Furthermore, the representation of the sensory elements of music and language may be either separate or in different locations with smaller degrees of overlap (for more details, see [63–66]). Nevertheless, if there is a bihemispheric representation for speech production, then the question of why an intervention that uses singing or a form of singing such as MIT has the potential to facilitate syllable and word production still remains. In theory, there are four possible mechanisms by which MIT’s therapeutic effect could be achieved, as described in the following sections.

**Reduction of speed**

In singing, words are articulated at a slower rate than in speaking, thereby reducing dependence on the left hemisphere. The rate of 1 syllable/s is the rate suggested by the developers of MIT [37]. Although we have made some adjustments to the original MIT protocol [38], we did find the slow rate of vocalization particularly useful as a starting rate for our patients. In fact, many of our patients are so severely impaired at baseline that 1 syllable/s is initially too fast for them. Although rate is not used as an outcome measure in the daily sessions, we do train our therapists to adhere to the 1 syllable/s rate. When patients reach the ‘advanced level’ of MIT, the rate is gradually increased to approximately 2 syllables/s as they are transitioned from singing back to speaking.

**Syllable lengthening**

Syllable lengthening provides the opportunity to distinguish the individual phonemes that together form words and phrases while the continuous vocalization inherent in singing ‘strings’ the sound together and thereby encourages fluency. This connected segmentation (i.e., overemphasizing the individual phonemes but still connecting them into meaningful words and phrases), coupled with the reduction of speed in singing, may help nonfluent aphasic patients practice auditory–motor mapping under feedback control and thereby increase fluency, possibly by receiving greater support from right-hemispheric structures.

**Syllable ‘chunking’**

Prosodic features such as intonation, change in pitch and syllabic stress may help patients group syllables into words and words into phrases, and this ‘chunking’ may also enlist more right-hemispheric support.

**Left-hand tapping (1 tap/syllable & 1 syllable/s)**

Left-hand tapping is likely to engage a right-hemispheric, sensorimotor network that may, in turn, provide an impulse for verbal production in much the same way that a metronome has been shown to serve as a ‘pacemaker’ in other motor activities (e.g., rhythmic anticipation or rhythmic entrainment) [67,68]. In addition, there may be a set of shared neural correlates that control both hand movements and articulatory movements [69–72], and furthermore, the sound produced by the tapping may encourage auditory–motor coupling [73]. In theory,
reduction of speed, syllable lengthening and syllable chunking can be applied to nonintonation-based speech techniques. However, these components are not often featured in the traditional therapeutic context.

How might MIT facilitate recovery and what do its unique elements contribute to the process? Functional imaging tasks targeting the perception of musical components that require a more global than local processing strategy (e.g., melodic contour, musical phrasing and/or meter) tend to elicit greater activity in right- than left-hemispheric brain regions. It has also been demonstrated that tasks emphasizing spectral information over temporal information have shown more right- than left-hemispheric activation [74,75]. Furthermore, patients with right-hemispheric lesions have greater difficulty with global processing (e.g., melody and contour processing) than those with left-hemispheric lesions [76,77]. Thus, it is possible that the melodic element of MIT engages the right hemisphere, particularly the right temporal lobe, more than therapies that do not make use of pitch or melody.

The effects of the left-hand tapping should be considered in the same context. Once the right temporal lobe is specifically engaged by the melodic intonation and contour, it is conceivable that the role of the left-hand tapping could be the activation and priming of a right-hemispheric sensorimotor network for articulation. Since concurrent speech and hand use occurs in daily life and gestures are frequently used during speech, hand movements, possibly in synchrony with articulatory movements, may have a facilitating effect on speech production, but the precise role of this facilitation is unknown. We hypothesize that tapping the left hand may engage a right-hemispheric sensorimotor network that coordinates not only hand movements but also orofacial and articulatory movements as well, and may facilitate speech production through rhythmic anticipation, rhythmic entrainment or auditory–motor coupling [67,68,73,78,79].

**Conclusion**

The clinical observation that patients with non-fluent aphasia are better at singing lyrics than they are at speaking the same words inspired the development of MIT. Despite several small case series, the efficacy of MIT has not been substantiated and its neural correlates remain largely unexplored. Because of its potential to engage or unmask language—in particular, speech-motor regions in the unaffected right hemisphere—MIT is well suited for patients with large left-hemispheric lesions whose only chance to recover is through recruitment of the right hemisphere. The observed brain changes following treatment indicate that MIT’s unique engagement of predominantly right-hemispheric brain regions.

---

**Executive summary**

**Paths to recovery from aphasia**
- There are two principal pathways of recovery:
  - Patients with small lesions in the language-dominant hemisphere and milder forms of aphasia are more likely to recover through recruitment of the perilesional cortex;
  - Patients with large lesions in the language-dominant hemisphere and moderate-to-severe forms of aphasia are most likely to recover through recruitment and training of rudimentary language-capable structures in the right hemisphere.

**Singing but not speaking in patients with moderate & severe nonfluent aphasia**
- Anecdotal reports and small case series of aphasic patients who are unable to speak but are able to sing words can be found in the literature as long as 100 years ago.
- This suggests two somewhat duplicate systems for vocal production. Typically, the language-dominant hemisphere has a very elaborate system of support for expressive language functions. The nondominant hemisphere has a separate, most likely rudimentary system in place that can support speech-motor functions.

**Behavioral & neural correlates of melodic intonation therapy**
- An intense course of melodic intonation therapy (MIT) engages a right fronto–temporal network through two unique components of MIT: melodic intonation and left-hand tapping.
- An intense course of MIT leads to improvement in spontaneous language skills.
- Pilot imaging data suggest that functional and structural imaging changes occur in a right fronto–temporal network in patients undergoing an intense course of MIT.

**Mechanisms of the therapeutic effects of melodic intonation therapy**
- Reduction in speed, syllable lengthening, syllable chunking and the engagement of a right-hemispheric articulatory sensorimotor network are the main mechanisms that may underlie the therapeutic effect of MIT.
- Long-lasting speech improvements are supported by structural brain changes, in particular in tracts connecting fronto–temporal brain regions.
(including the superior temporal region, the primary sensorimotor and premotor cortices and the inferior frontal gyrus) and the connections between these regions (mainly through the AF), accounts for its facilitating effect.

**Future perspective**

Although approximately 1,000,000 people in the USA suffer from aphasia [101], reliable and standard treatment methods have not been established for this disorder. While a meta-analysis by Robey determined that an array of treatment methods for aphasia is, on average, beneficial [80], effect sizes vary widely [80,81]. Some of this variability is undoubtedly due to the differences in treatment approaches used across studies. Furthermore, a meta-analysis by Bhogal et al. concluded that aphasia treatments are more likely to achieve positive results if the total amount of therapy exceeds 55 h [82]. To maximize both the success and the efficiency of new treatment methods, it is important to first understand which components of the treatment contribute to its effectiveness and then apply these components to a treatment study in an intense manner. In this regard, our ongoing randomized clinical trial (RO1DC008796, NCT00903266) is comparing MIT with a matched control treatment (i.e., speech repetition therapy) that does not include the two unique elements of MIT (i.e., intonation and rhythmic tapping with the left hand) but shares other therapy components. This large-scale study represents an example of testing the efficacy of an experimental intervention by isolating its critical components through the selection of an appropriate control intervention and by randomizing patients using a stratification scheme based on impairment level and/or lesion size and location (more details of this trial can be found at [102,103]). The selection of meaningful and quantifiable outcome measures with high test–retest reliability is also an important component of this study.

**Financial & competing interests disclosure**

The authors gratefully acknowledge support from the NIH (1RO1 DC008796, 3R01DC008796–02S1, R01 DC009823–01), the Rosalyn and Richard Siffta Family Foundation, the Grammy Foundation and the Matina R Proctor Foundation. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

**Bibliography**

Papers of special note have been highlighted as:

- of interest
- of considerable interest

20. First publication to show dynamic changes in language reorganization after stroke.

**Special Report**

**From singing to speaking: facilitating recovery from nonfluent aphasia**
First description of MIT.


First series of chronic aphasic patients, all treated with MIT.


Important publication that provides a framework: treatment of underlying forms.

Functional MRI study showing shared and distinct regions of activation when comparing singing and speaking with the same rate of production.


Excellent description of the language pathways in humans.


Important publication that provides a comprehensive overview of repair mechanisms after stroke.


Describes an important quantitative measure of propositional speech.


From singing to speaking: facilitating recovery from nonfluent aphasia

**Special Report**


**Important meta-analysis establishing the efficacy of speech therapy.**


Websites

101. The National Aphasia Society
   [www.aphasia.org/Aphasia%20Facts/aphasia_faq.html](http://www.aphasia.org/Aphasia%20Facts/aphasia_faq.html)

102. Music, Neuroimaging and Stroke Recovery Laboratories
   [www.musicianbrain.com/#aphasia](http://www.musicianbrain.com/#aphasia)

103. Melodic intonation therapy and speech repetition therapy for patients with nonfluent aphasia