Melodic Intonation Therapy

Shared Insights on How It Is Done and Why It Might Help

Andrea Norton, Lauryn Zipse, Sarah Marchina, and Gottfried Schlaug

Music, Stroke Recovery, and Neuroimaging Laboratory, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, Massachusetts, USA

For more than 100 years, clinicians have noted that patients with nonfluent aphasia are capable of singing words that they cannot speak. Thus, the use of melody and rhythm has long been recommended for improving aphasic patients' fluency, but it was not until 1973 that a music-based treatment [*Melodic Intonation Therapy* (MIT)] was developed. Our ongoing investigation of MIT's efficacy has provided valuable insight into this therapy's effect on language recovery. Here we share those observations, our additions to the protocol that aim to enhance MIT's benefit, and the rationale that supports them.

Key words: **Melodic Intonation Therapy; nonfluent aphasia; language recovery; brain plasticity; music therapy**

Introduction

According to the National Institutes for Health (*NINDS Aphasia Information Page: NINDS*, 2008), approximately 1 in 272 Americans suffer from aphasia, a disorder characterized by the loss of ability to produce and/or comprehend language. Despite its prevalence, the neural processes that underlie recovery remain largely unknown and thus have not been specifically targeted by aphasia therapies. One of the few accepted treatments for severe, nonfluent aphasia is *Melodic Intonation Therapy* (MIT),[1–6] a treatment that uses the musical elements of speech (melody and rhythm) to improve expressive language by capitalizing on preserved function (singing) and engaging language-capable regions in the undamaged right hemisphere. In this chapter, we describe how to administer MIT (based on the descriptions of Helm-Estabrooks *et al.*[3,4]), share our additions to the protocol, and explain how it may exert is therapeutic effect.

What Exactly Is MIT?

The original program of music intonation therapy is designed to lead nonfluent aphasic patients (Fig. 1) from intoning (singing) simple, 2–3 syllable phrases, to speaking phrases of 5 or more syllables[1,3,4] across three levels of treatment. Each level consists of 20 high-probability words (e.g., "water") or social phrases (e.g., "I love you"*)* presented with visual cues. Phrases are intoned on just two pitches, "melodies" are determined by the phrases' natural prosody [e.g., stressed syllables are sung on the higher of the 2 pitches, unaccented syllables on the lower pitch (Fig. 2)], and the patient's left hand is tapped 1× per syllable. Although it may appear that the primary difference between the levels is phrase length, the more important

Address for correspondence: Gottfried Schlaug, M.D., Ph.D., Beth Israel Deaconess Medical Center, Department of Neurology – Palmer 127, 330 Brookline Avenue, Boston, MA 02215. Voice: 617-632-8917; fax: 617-632-8920. gschlaug@bidmc.harvard.edu

The Neurosciences and Music III: Disorders and Plasticity: Ann. N.Y. Acad. Sci. 1169: 431–436 (2009).
doi: 10.1111/j.1749-6632.2009.04859.x

A good response to MIT is defined as improvement in conversational speech skills. Patients who are most likely to respond well to this treatment have most or all of the following characteristics:

- a unilateral, left-hemisphere stroke
- poorly articulated, nonfluent, or severely restricted speech output
- ability to produce some intelligible words while singing familiar songs
- poor repetition, even for single words
- moderately well preserved auditory comprehension
- poorly articulated attempts at speech
- good motivation, emotional stability, and good attention span

Figure 1. Ideal candidates for Melodic Intonation Therapy. Summarized from Helm-Estabrooks *et al*.[3]

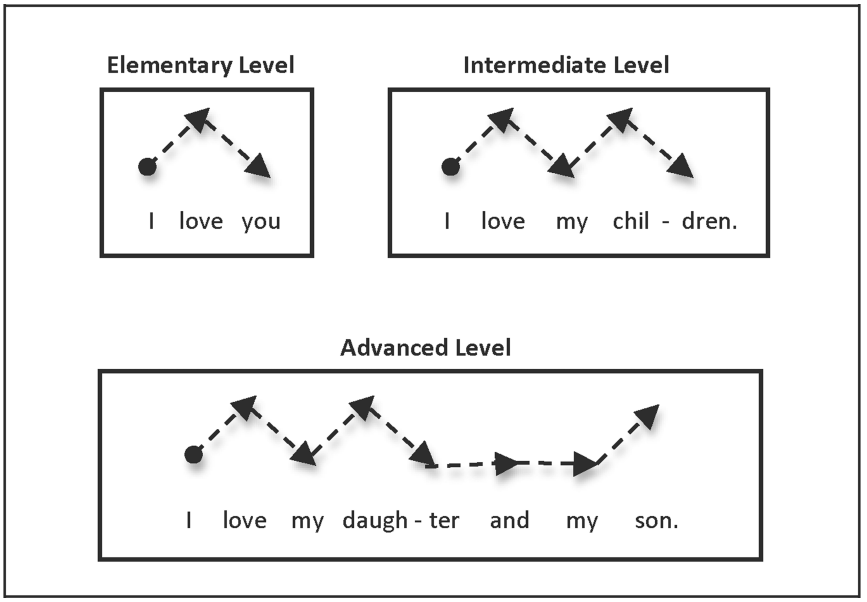

Figure 2. Melodic phrase construction: Phrases are sung on just two pitches; melodic contour is determined by the natural prosody of speech (e.g., stressed syllables are sung on the higher of the two pitches); phrases increase in length and difficulty (Elementary: 2–3 syllables; Intermediate: 4–6 syllables; Advanced: 6–9 syllables) as patients progress through the three levels of treatment. Summarized from Helm-Estabrooks *et al*.[3]

distinctions are the administration of the treatment and degree of support provided by the therapist (Figs. 3–5).

Interestingly, there appear to be almost as many interpretations of the original protocol as there are people using it. While early reports[6,7] depict phrases using 3 pitches rather than the originally specified 2, anecdotal evidence (DVDs from prospective patients across the United States) shows a number of therapists using the technique, and no two sessions are alike. Some use 2 pitches separated by a perfect 4th or 5th, while others write a new tune for each phrase using as many as 7–8 pitches in a specified key. Still others accompany their patients on the piano, use familiar song melodies, or rapidly "play" 4–5 notes up and down the patient's arm as they sing words

Step	Procedure
1	***Humming*** Therapist introduces the target phrase by showing a visual cue, humming the phrase 1x at a rate of 1 syllable/sec., then intoning (singing) the phrase 2x while tapping the patient's left hand 1x per syllable.
2	***Unison intoning*** Therapist and patient intone (sing) the target phrase together while the therapist taps the patient's left hand (1x/syllable).
3	***Unison intoning with fading*** Therapist and patient begin to intone (sing) and tap the target phrase together, but halfway through, the therapist fades out while the patient continues to sing the rest of the phrase accompanied by hand-tapping, but with no further verbal or oral/facial cueing.
4	***Immediate Repetition*** Therapist intones and taps the target phrase while the patient listens. The patient immediately repeats the phrase assisted only by the tapping of the left hand.
5	***Response to a probe question*** Immediately following the patient's successful repetition of the target phrase (Step 4), the therapist quickly intones a question (e.g., "*What did you say?*") and the patient answers by intoning the target phrase. Hand-tapping is the only assistance allowed.

Figure 3. Elementary Level steps and procedures in MIT. Summarized from Helm-Estabrooks *et al*.[3]

or phrases. While all such variations might have the potential to engage right-hemisphere regions capable of supporting speech, it may be just such complex interpretations of the protocol that prevent therapists with little or no musical background from using the treatment. Thus, we aim to simplify the process so any therapist can administer it, and well-trained patients and caregivers can learn to apply the method when intensive treatment ends. Because the focus is not on performance, one does not need to be a musician or even a good singer to administer or participate in this treatment. The goal is to uncover the inherent melody in speech to gain fluency and increase expressive output.

Getting Started

Seated across a table from the patient, the therapist shows a visual cue and introduces a word/phrase (e.g., "**Thank** you"). The accented/stressed syllable(s) will be sung on the higher of the 2 pitches, unaccented syllable(s) on the lower pitch (Fig. 2). The starting pitch should rest comfortably in the patient's voice range, and the other pitch should be a minor 3rd (3 semitones) above or below (*middle C* and the *A* just below it works well for most people). For those unfamiliar with this terminology, think of the children's taunt, "*Naa-naa – Naa-naa.*" These 2 pitches create the interval of a minor 3rd, which is universally familiar, requires no special singing skill, and provides a good approximation of the prosody of speech that still falls into the category of singing.

What Else Is New?

While it has been shown that MIT in its original form leads to greater fluency in small case series,[5] sustaining treatment effects can be a challenge for any intervention. Thus, we have instituted the use of *Inner Rehearsal* and *Auditory-Motor Feedback Training* to help patients gain

Step	Procedure
1	***Phrase Introduction*** Therapist shows a visual cue and intones the phrase 2x (1 syllable/sec.) while tapping the patient's left hand 1x per syllable.
2	***Unison with fading*** Therapist and patient begin to intone and tap the target phrase together, but halfway through, the therapist fades out while the patient continues to sing the rest of the phrase accompanied by hand-tapping, but with no further verbal or oral/facial cueing.
3	***Delayed Repetition*** Therapist intones and taps the target phrase while the patient listens. After a 6-second delay, the patient repeats the phrase assisted only by the tapping of the left hand. No verbal assistance may be given.
4	***Response to a probe question*** Following the patient's successful repetition of the target phrase (Step 3), the therapist waits 6 seconds, then quickly intones a question (e.g., "*What did you say?*") and the patient answers by intoning the target phrase. No assistance is allowed.

Figure 4. Intermediate Level steps and procedures in MIT.
Summarized from Helm-Estabrooks *et al*.[3]

Step	Procedure
1	***Delayed Repetition*** Therapist intones and taps the target phrase while the patient listens. After a 6-second delay, the patient repeats the phrase assisted only by tapping the left hand. No verbal assistance may be given.
2	***Introducing Sprechgesang*** Therapist presents the target phrase in sprech-gesang 2x (accompanied by hand-tapping) while the patient listens. The words should not be sung, but instead, should be presented slowly with exaggerated emphasis on rhythm and stressed (accented) syllables.
3	***Sprechgesang with fading*** Therapist and patient begin the target phrase together (with hand-tapping), but halfway through, the therapist fades out while the patient completes the phrase alone.
4	***Delayed Spoken Repetition*** Therapist presents the target phrase using normal speech prosody (no hand-tapping) while the patient listens. After a 6-second delay, the patient repeats the phrase using normal speech.
5	***Response to a Probe Question*** After a 6-second delay, the therapist asks a question to elicit the target phrase using normal speech. The patient answers by speaking the target phrase without assistance of any kind.

Figure 5. Advanced Level steps and procedures in MIT.
Summarized from Helm-Estabrooks *et al*.[3]

maintainable independence as they improve expressive speech.

Inner Rehearsal

So that patients learn to establish their own "target" phrases, the therapist models the process of *Inner Rehearsal* by slowly tapping the patient's hand (1 syllable per second) while humming the melody, then softly singing the words, explaining that s/he is "hearing" the phrase sung "inside." If the patient has trouble understanding how to do this, s/he is asked to imagine hearing someone sing "*Happy Birthday*" or a parent's voice saying, "*Do your homework*." Once the concept is understood, the therapist taps while softly singing the phrase and indicating that the patient should hear his/her own voice singing the phrase "inside." This inner rehearsal (covert production) of the phrase creates an auditory "target" with which the overtly produced phrase can be compared. Those who master this technique can eventually transfer the skill from practiced MIT phrases to expressive speech initiated with little or no assistance.

Auditory-Motor Feedback Training

Because re-learning to identify and produce individual speech sounds is essential to patients' success, training them to hear the difference between the target phrase and their own speech is a key aspect of the recovery process. In the early phases of treatment, patients listen as the therapist sings the target, and learn to compare their own output as they repeat the words/phrases. Sounds identified as incorrect become the focus of remediation. Once a problem is corrected, the process of singing, listening, and repeating begins again. As patients learn to create their own target through *Inner Rehearsal*, *Auditory-Motor Feedback Training* allows them to self-monitor as thoughts are sung aloud. Over time, they learn to use the auditory-motor feedback "loop" to hear their own speech objectively, identify problems and adjust to correct them as they speak, and thereby decrease dependence on the therapist.

How Does MIT Work?

Preliminary data comparing MIT to an equally intense control therapy that uses no intoning or left-hand tapping indicate that those two elements add greatly to MIT's effectiveness.[5] While its developers suggested that tapping and intoning could engage homologous language regions in the right hemisphere, they did not explain how this would occur.[2] Additionally, since ideal candidates for MIT[2] (Fig. 1) are patients with Broca's aphasia, a population with both linguistic and motor speech impairments, the extent to which MIT addresses aphasia, apraxia, or both is not yet clear. Below is a brief discussion of MIT's critical elements and how they may contribute its therapeutic effect.

Intonation

The intonation at the heart of MIT was originally intended to engage the right hemisphere, given its dominant role in processing spectral information, global features of music, and prosody.[2,8,9] The right hemisphere may be better suited for processing slowly modulated signals, while the left hemisphere may be more sensitive to rapidly modulated signals.[10] Therefore, it is possible that the slower rate of articulation and continuous voicing that increases connectedness between syllables and words in singing may reduce dependence on the left hemisphere.

Left-hand Tapping

Tapping the left hand may engage a right-hemisphere sensorimotor network that controls both hand and mouth movements.[11] It may also facilitate sound-motor mapping, which is a critical component of meaningful vocal communication.[12] Furthermore,

tapping, like a metronome, may pace the speaker and provide continuous cueing for syllable production.

Inner Rehearsal

Inner rehearsal may be particularly effective for addressing apraxia (an impairment of the ability to sequence and implement serial and higher-order motor commands). Silently intoning the target phrase may reinitiate a cascade of activation from a higher level in the cognitive-linguistic architecture (e.g., from the level of a prosodic or phonologic representation), thus giving the speaker another attempt to correctly sequence the motor commands.

Auditory-Motor Feedback Training

In speech, phonemes occur so quickly that it is difficult for severely aphasic and/or apraxic patients to process auditory feedback in time to self-correct. However, when words are sung, phonemes are isolated and thus, can be heard distinctly while remaining connected to the word. In addition, sustained vowel sounds provide time to "think ahead" about the next sound, make internal comparisons to the target, and self-correct when sounds produced begin to go awry.

Whether learning to sing, play an instrument, recover language after stroke, or acquire any new skill, the key to mastery is in the process. Previous small, open-label studies examining patients with moderate to severely nonfluent aphasia have found MIT to be a promising path to fluency, and we have outlined and discussed its critical elements and our enhancements to the protocol here. However, the efficacy of this technique and its potential application to other disorders will require further exploration, and more research will be necessary to fully understand the neural processes that underlie MIT's effect.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Albert, M.L., R.W. Sparks & N.A. Helm. 1973. Melodic intonation therapy for aphasia. *Arch. Neurol.* **29:** 130–131.
2. Sparks, R., N. Helm & M. Albert. 1974. Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex* **10:** 303–316.
3. Helm-Estabrooks, N., M. Nicholas & A. Morgan. 1989. *Melodic Intonation Therapy*. Pro-Ed., Inc. Austin, TX.
4. Helm-Estabrooks, N. & M.L. Albert. 2004. *Melodic Intonation Therapy: Manual of Aphasia and Aphasia Therapy*, 2nd ed., chapt.16, pp. 221–233. Pro-Ed. Austin, TX.
5. Schlaug, G., S. Marchina & A. Norton. 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.
6. Marshall, N. & P. Holtzapple. 1976. Melodic intonation therapy: variations on a theme. In *Clinical Aphasiology Conference*, vol. 6. R.H. Brookshire, Ed.: 115–141. BRK Publishers. Minneapolis, MN.
7. Sparks, R.W. & A.L. Holland. 1976. Method: melodic intonation therapy for aphasia. *J. Speech Hear. Disord.* **41:** 287–297.
8. Zatorre, R.J. & P. Belin. 2001. Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **11:** 946–953.
9. Schuppert, M., T.F. Münte, B.M. Wieringa & E. Altenmüller. 2000. Receptive amusia: evidence for cross-hemispheric neural networks underling music processing strategies. *Brain* **123:** 546–559.
10. Poeppel, D., W.J. Idsardi & V. van Wassenhove. 2008. Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond. Ser. B* **363:** 1071–1086.
11. Gentilucci, M. & R. Dalla Volta. 2008. Spoken language and arm gestures are controlled by the same motor control system. *Q. J. Exp. Psychol. (Colchester)* **61:** 944–957.
12. Lahav, A., E. Saltzman & G. Schlaug. 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.* **27:** 308–314.