DOI: 10.1111/nyas.14848

REVIEW

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Melodic Intonation Therapy for aphasia: A multi-level meta-analysis of randomized controlled trials and individual participant data

Tudor Popescu[1,2,3] | **Benjamin Stahl**[4,5,6,7] | **Brenton M. Wiernik**[8] | **Felix Haiduk**[1] | **Michaela Zemanek**[9] | **Hannah Helm**[4] | **Theresa Matzinger**[1,10,11] | **Roland Beisteiner**[2] | **W. Tecumseh Fitch**[1]

[1]Department of Behavioural and Cognitive Biology, University of Vienna, Vienna, Austria

[2]Department of Neurology, Medical University of Vienna, Vienna, Austria

[3]Centre for Mind and Brain Sciences, University of Trento, Rovereto, Italy

[4]Faculty of Science, Medical School Berlin, Berlin, Germany

[5]Department of Neurology, Charité Universitätsmedizin, Berlin, Germany

[6]Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[7]Department of Neurology, Universitätsmedizin Greifswald, Greifswald, Germany

[8]Department of Psychology, University of South Florida, Tampa, Florida, USA

[9]University Library, University of Vienna, Vienna, Austria

[10]Department of English, University of Vienna, Vienna, Austria

[11]Centre of Language Evolution Studies, Nicolaus Copernicus University Toruń, Toruń, Poland

Correspondence
Tudor Popescu and W. Tecumseh Fitch, Department of Behavioural and Cognitive Biology, University of Vienna, A-1030 Vienna, Austria.
Email: tudor.popescu@univie.ac.at and tecumseh.fitch@univie.ac.at

Benjamin Stahl, Faculty of Science, Medical School Berlin, D-14197 Berlin, Germany.
Email: benjamin.stahl@medicalschool-berlin.de

Funding information
Medical University of Vienna and University of Vienna, Grant/Award Number: SO10300020

Abstract

Melodic Intonation Therapy (MIT) is a prominent rehabilitation program for individuals with post-stroke aphasia. Our meta-analysis investigated the efficacy of MIT while considering quality of outcomes, experimental design, influence of spontaneous recovery, MIT protocol variant, and level of generalization. Extensive literature search identified 606 studies in major databases and trial registers; of those, 22 studies—overall 129 participants—met all eligibility criteria. Multi-level mixed- and random-effects models served to separately meta-analyze randomized controlled trial (RCT) and non-RCT data. RCT evidence on validated outcomes revealed a small-to-moderate standardized effect in noncommunicative language expression for MIT—with substantial uncertainty. Unvalidated outcomes attenuated MIT's effect size compared to validated tests. MIT's effect size was 5.7 times larger for non-RCT data compared to RCT data ($\bar{g}_{case\ report} = 2.01$ vs. $\bar{g}_{RCT} = 0.35$ for validated Non-Communicative Language Expression measures). Effect size for non-RCT data decreased with number of months post-stroke, suggesting confound through spontaneous recovery. Deviation from the

original MIT protocol did not systematically alter benefit from treatment. Progress on validated tests arose mainly from gains in repetition tasks rather than other domains of verbal expression, such as everyday communication ability. Our results confirm the promising role of MIT in improving trained and untrained performance on unvalidated outcomes, alongside validated repetition tasks, and highlight possible limitations in promoting everyday communication ability.

KEYWORDS
Melodic Intonation Therapy, meta-analysis, post-stroke aphasia, apraxia of speech, singing, rhythmic pacing, formulaic language, experimental design

INTRODUCTION

Stroke survivors often experience a profound loss of communication skills, among them a syndrome known as aphasia. This syndrome may manifest as severe difficulty in verbal expression, referred to as "non-fluent aphasia." In addition, stroke survivors frequently suffer from impaired speech-motor planning. Known as "apraxia of speech," this syndrome often occurs in combination with aphasia. Although about one-third of individuals with neurological communication disorders do not recover completely,[1] rehabilitation programs can improve language performance even in the chronic stage of symptoms.[2]

Melodic Intonation Therapy (MIT) is a prominent rehabilitation program originally developed for individuals with non-fluent aphasia.[3] Drawing on the observation that individuals with neurological communication disorders are sometimes able to sing entire pieces of text fluently,[4–6] MIT uses melody, rhythmic pacing, vocal expression in unison and alone, left-hand tapping, formulaic and non-formulaic verbal utterances, as well as other therapeutic elements in a hierarchically structured protocol.[7] Hypotheses on MIT's neural mechanisms have been discussed.[8]

To date, randomized controlled trial (RCT) data have confirmed the efficacy of MIT on validated outcomes in the *late subacute* or *consolidation* stage of aphasia (i.e., up to 12 months after stroke),[9] but not in the *chronic* stage of symptoms (i.e., more than 6–12 months after stroke).[10] From a methodological perspective, influences of spontaneous recovery are generally lower in the *chronic* stage of aphasia, as suggested by RCT data[11] and meta-analyses.[12] This points out the need to carefully consider the stage of symptoms post-stroke in research on MIT. Importantly, speech-language therapy seeks to promote performance on untrained items. Consistent with this goal, the present work distinguishes progress on *trained* items—that is, learning resulting from using the same set of utterances both during treatment and subsequent assessment—from the more desirable goal of attaining generalization to *untrained* items, ideally in the context of everyday communication, to ensure ecological validity.[13]

So far, there have been several systematic reviews[14,15] and meta-analyses on MIT.[16–18] Existing meta-analyses reflect a relatively limited amount of RCT data,[16] dichotomize posttreatment improvement in a way that prevents specific effect size estimates,[17] or do not operationalize quality of outcomes (psychometrically validated vs. unvalidated tests), experimental design (presence vs. absence of randomization and control group), influence of spontaneous recovery (quantified as number of months post-stroke), MIT protocol applied (original vs. modified), and level of generalization (performance on trained vs. untrained items).[18] Given the substantial burden of disease associated with aphasia, the current meta-analysis attempts to provide a deeper understanding regarding the clinical potential and possible limitations of MIT. To achieve this goal, our evaluation synthesizes available studies on MIT to address five research questions, focusing on whether the effect size of the rehabilitation program is systematically altered by:

1. Psychometric quality (i.e., use of validated vs. unvalidated outcomes);
2. Experimental design (i.e., RCT vs. non-RCT data);
3. Confound through spontaneous recovery (i.e., decreasing with number of months post-onset of stroke, MPO);
4. Deviation in protocol (i.e., original vs. slightly modified MIT variants);
5. Degree of generalizability (i.e., performance on trained vs. untrained items).

METHODS

Eligibility criteria

We defined the following basic inclusion criteria for primary studies to be considered for the present meta-analysis:

1. Empirical study with or without a control group that administered MIT to adult individuals with aphasia (aged at least 18 years);
2. Language-related outcomes in prepost assessment.

We chose to include case reports with individual participant data (IPD) to increase the pool of evidence. To determine the influence of experimental design on treatment outcome, we analyzed RCT and non-RCT data separately and comparatively.

After removal of duplicate items (see Supplementary Materials, Section 1), the following exclusion criteria were applied to remaining studies (in this order):

1. Publication in non-peer-reviewed or predatory journal;
2. Unvalidated outcomes (i.e., no published or otherwise accessible work confirming the psychometric properties of a particular test battery); exceptionally, if a study included both trained and untrained items for an unvalidated outcome, we included this work to determine the degree of generalization by comparing performance on trained and untrained items;
3. Other essential data not reported and/or not retrievable, even after contacting the authors (e.g., no sample size or standard error, insufficient information to compute an effect size);
4. Substantial variation from original MIT protocol;[3] we accepted *minor* changes to the MIT protocol (to examine the effect of the categorical variable: original vs. modified MIT), as long as the treatment had all of the following features:
   (i) melody-based vocal expression both in unison and alone;
   (ii) some form of rhythmic pacing (e.g., left-hand tapping);
   (iii) use of verbal utterances known from everyday communicative interaction.

   Aside from the original version of MIT, seven modified MIT protocols were reported across studies initially considered before applying our protocol-related exclusion criteria. Applying these exclusion criteria resulted in four MIT protocol variants finally included (citations indicate the first description of the protocol itself, where available, or studies employing it):
   (i) Modified Melodic Intonation Therapy (MMIT);[19]
   (ii) Singing, Intonation, Prosody, "Atmung" (German for "breathing"), Rhythm, and Improvisation (SIPARI);[20]
   (iii) Speech-Music Therapy for Aphasia (SMTA);[21]
   (iv) "singing therapy."[22,23]

   Excluded protocols were (each with the specific reason):
   (i) Metrical Pacing Technique (no melodic intonation alongside rhythmic pacing);
   (ii) aphasia choirs (unison singing only; use of regular song lyrics rather than verbal utterances known from everyday communicative interaction);
   (iii) music therapy delivered in addition to, but separately from SLT speech-language therapy (no integrative melody-based vocal expression of verbal utterances known from everyday communicative interaction).

Taken together, the included studies comprised 129 treated participants (59 in RCTs; 70 in IPDs) and 62 control participants (all in RCTs). The full list of included and excluded studies can be found in Tables S1 and S2.

Search strategy

To ensure high sensitivity, we used both free text and subject headings in databases for our search, not restricting language or publication form.[24] Figure 1 shows the PRISMA statement chart (Preferred Reporting Items for Systematic Reviews and Meta-Analyses[25]), which summarizes the study counts in all stages of the search. The full counts are given in the Supplementary Materials, Section 1, which also documents the entire literature search procedure, including search terms, databases used, and attempts made to access "gray literature." Furthermore, we followed the guidelines and standards in the Methodological Expectations of the Cochrane Intervention Reviews handbook, and those in the PRISMA checklist (see Supplementary Materials).

Study coding and double-coding

All studies were coded by the first author (T.P.). Two of the authors (F.H. and T.M.) recoded all the studies, verifying cross-coder consistency. Agreement among the three coders occurred in a majority of cases. Discrepancies between coding sheets were resolved by consensus. Intra-class correlations were >.9 in the remaining cases, which amounted to errors arising from numerical estimates of data reported in plot format only.

Outcomes of primary studies

Test batteries of primary studies and their validation status are shown in Table S3. Table S4 details each subtest per battery and the corresponding linguistic Ability; the associated Target Syndrome (aphasia or apraxia of speech); and the hierarchical categorization scheme that determined the dependent variable meta-analyzed (Domain). An abridged, tree-form version of this categorization scheme is illustrated in Figure 2.

Meta-analysis methods

Computed outcome metric

To maximize comparability of effects across studies, we used change scores from pre- to post-test as outcome metric, expressed in *z*-scores. For group-level studies (the RCTs in the current analyses), we standardized *z*-scores using pooled pre-test standard deviation across control and treatment groups. For IPD studies (the case reports in the current analyses), we computed *z*-scores in one of three ways. For studies that reported results as *z*-scores, we used the *z*-scores directly. For studies that reported results as percentile scores, we converted these to *z*-scores using the quantiles of the standard normal distribution. For other studies, we estimated *z*-scores using the following procedure. First, we normalized[a] raw scores to reflect the

[a] For a small number of studies, it was not possible to determine the maximum or minimum possible scores. For these studies, we computed POMP scores using the maximum and minimum *observed* scores in the sample. Results did not change meaningfully if we excluded these studies from results.

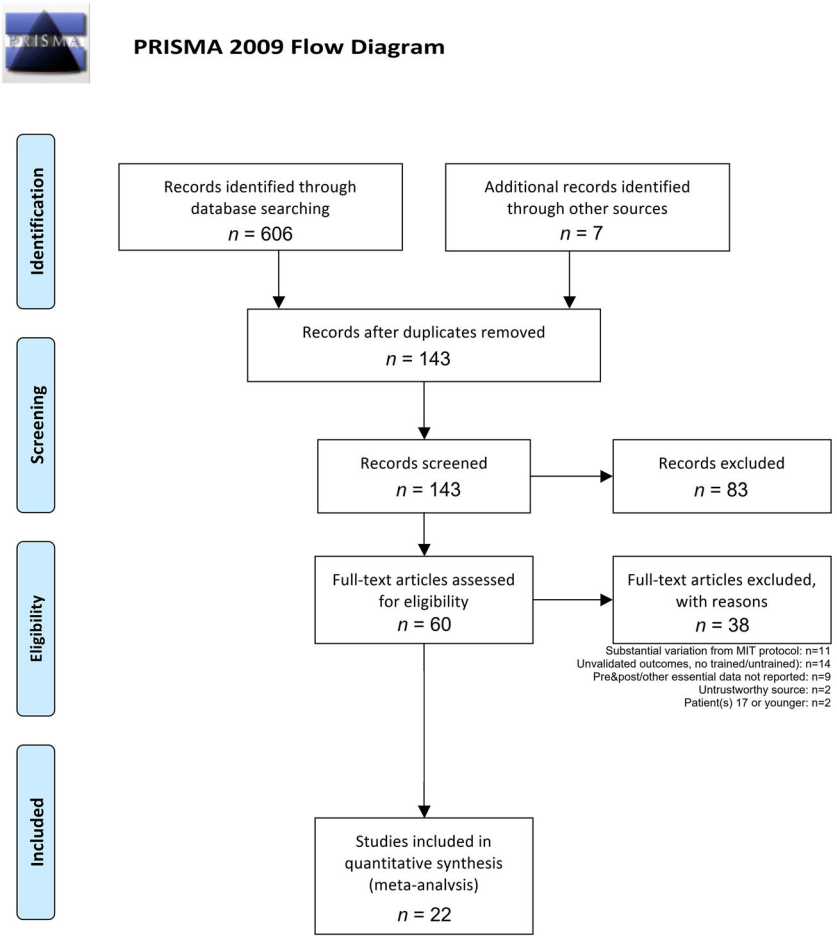

FIGURE 1 Flow diagram from the PRISMA statement. Included and excluded studies are shown in Tables S1 and S2.

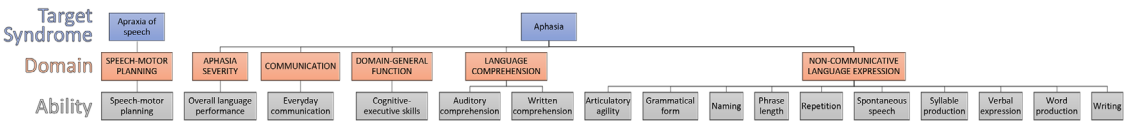

FIGURE 2 Hierarchical categorization scheme, showing how Abilities nest into Domains, within each Target Syndrome considered. All meta-analyses were done at the Domain level of abstraction.

FIGURE 3 Nested multi-level model employed. Standard deviations (τ) are shown at every level: measures (✎), nested within patients (👤), nested within studies (🖳).

proportion of the maximum possible score, POMP.[26] Second, we estimated a three-level random-intercept model for the pretest POMP scores, as detailed in Figure 3. Third, we used the population intercept from these models as the estimated POMP score *mean*, and the participant-level random effects standard deviation as the estimated POMP score (τ). Finally, we used the resulting *mean* and *SD* values to standardize the pre- and post-test POMP scores. (for models specifically fitted to RCT and case report data, see Supplementary Materials, Section 4, where a detailed explanation is also provided for the statistical rationale behind our aggregating of different outcomes).

Moderator analyses

For the RCT meta-analyses, we fitted a meta-regression model with the moderators (1) Domain; (2) whether the study used validated or unvalidated tests as outcomes (for unvalidated measures, we treated trained and untrained items as separate groups to avoid confounding validation with training effects); and (3) the Domain × Validated interaction. To test the effect of time since stroke, we fit another model with additional moderators of (1) mean MPO across treatment and control groups; and (2) difference in mean MPO between treatment and control groups.

FIGURE 4 Meta-analytical results. Data points (circles, triangles, or squares) are study-level standardized mean pre- and post-test difference scores, either adjusted for a control group (g_{ppc}) or not (g_{pp}). Points of different colors are drawn from different studies. Large black points refer to mean $g_{pp(c)}$ values for validated outcomes, with 66% (thick bar) and 95% (thin bar) confidence intervals and *t*-distribution confidence densities. For case reports, one study focusing on Aphasia Severity is not displayed ($g_{ppc} = -4.88$).

For case-report meta-analyses, we initially fitted a similar IPD meta-regression model with Domain, Validated, and Domain × Validated moderators. We tested further moderators by fitting two additional models, adding one moderator at a time. First, we fit a model adding individual-level MPO. Second, we fit a model adding whether a study used the original or a modified MIT protocol.

Data availability

All data generated during the preparation of the present work are accessible online, including raw materials, coding sheets, analysis scripts, and supplementary materials (https://osf.io/gcjqr/).

RESULTS

Study-level standardized mean difference scores and meta-analytic mean differences by Domain are shown in Figure 4. Full meta-regression results are reported in Tables S12–S19.

RCT data

Overall, RCT data showed a small-to-moderate pre- and post-test effect of MIT on aphasia outcomes, after accounting for the control group ($\bar{g} = 0.31$, 95% CI: [−0.01, 0.63]). These results were primarily based on Non-Communicative Language Expression measures (i.e., focus on verbal utterances *per se*, such as in tasks requiring repetition of words and sentences; $k = 3$, $n_{treat} = 176$, $n_{control} = 188$[b]). Other abilities were less commonly assessed, including Communication (i.e., verbal utterances used for social interaction in everyday situations; $k = 2$, $n_{treat} = 39$, $n_{control} = 42$) and Language Comprehension (i.e., understanding the meaning of verbal utterances; $k = 2$, $n_{treat} = 36$, $n_{control} = 37$). In moderator analyses, effects appeared to be much weaker for Communication and Language Comprehension tasks than for Non-Communicative Language Expression, but confidence intervals for these differences were wide (Figure 4). Effects were estimated to be somewhat heterogeneous across studies (random effects standard deviation, $\tau = 0.33$, 95% CI: [0.15, 1.01]).

Two included RCTs have several unvalidated Non-Communicative Language Expression measures. For these unvalidated outcomes, treatment effects for untrained items were somewhat smaller than those for validated outcomes, although the associated confidence interval was fairly wide ($\Delta\bar{g} = -0.15$, 95% CI: [−0.46, 0.15]). As expected, estimated treatment effects were much larger when participants were tested using trained items compared to validated outcomes ($\Delta\bar{g} = 0.99$, 95% CI: [0.60, 1.39]; trained vs. untrained items contrast for unvalidated outcomes: $\Delta\bar{g} = 1.15$, 95% CI: [0.74, 1.56]). Consistent with the statistical literature on measurement error,[27–29] smaller effect sizes for (untrained) unvalidated outcomes ($k = 2$, $n_{treat} = 39$, $n_{control} = 42$) may be attributable to poorer reliability compared to validated tests ($k = 3$, $n_{treat} = 173$, $n_{control} = 183$).

When aphasia stage (MPO) was added to the RCT model ($k = 3$, $n_{treat} = 251$, $n_{control} = 267$), neither mean MPO across groups ($\Delta\bar{g}$ per month = −0.008, 95% CI: [−0.024, 0.008]) nor difference in mean MPO between MIT and control groups ($\Delta\bar{g}$ per month = −0.004, 95% CI: [−0.020, 0.011]) showed meaningful relationships with treatment effects of MIT. Importantly, effect sizes for RCT analyses were drawn from only three studies, so these group-level MPO analyses have limited power to estimate the impact of MPO on treatment effects.

Case report data

Compared to RCT data, case reports without control group yielded much larger effects of MIT ($\bar{g} = 1.72$, 95% CI: [1.00, 2.42]). As with RCT data, these results were primarily based on Non-Communicative Language Expression (repetition) tasks. Overall, Aphasia Severity and Language Comprehension appeared to show somewhat smaller effects, but confidence intervals on these differences were very wide. Effects

[b] k = number of studies, n = number of participants. For a complete list of case numbers (studies and participants) entering into each type of analysis, please see Section 5.1 of the Supplementary Materials (Tables S5 through S11).

were estimated to be highly heterogeneous across studies (τ [between-studies] = 1.41, 95% CI: [0.89, 2.05]), to the degree that MIT was even estimated to be harmful in a small proportion of settings; for instance, the 95% normal-theory prediction interval for Non-Communicative Language Expression ranged from −0.88 to +4.90.[30]

Four case reports included several unvalidated Non-Communicative Language Expression measures (total $n = 10$). As with RCT data, treatment effects for untrained items on unvalidated outcomes appeared to be smaller than for validated outcomes, with a wide confidence interval ($\Delta\bar{g} = -0.47$, 95% CI: [−2.40, 1.46]). Also similar to RCTs, apparent treatment effects were much larger for trained items compared with validated outcomes ($\Delta\bar{g} = 2.37$, 95% CI: [0.44, 4.31]; trained vs. untrained items contrast on unvalidated outcomes: $\Delta\bar{g} = 2.84$, 95% CI: [1.21, 4.48]).

When aphasia stage (MPO) was added to the case-report model ($k = 16$, $n = 246$), MPO showed a moderate negative relationship with treatment effects ($\Delta\bar{g}$ per month = −0.02, 95% CI: [−0.03, −0.01]; estimated effect for 12 months = −0.18, 95% CI: [−0.30, −0.07]; estimated effect for 24 months = −0.37, 95% CI: [−0.61, −0.14]).

Compared to studies using the original MIT protocol ($k = 9$, $n = 131$), modified protocols ($k = 10$, $n = 210$) appeared to show somewhat larger treatment effects, although the associated confidence interval was very wide ($\Delta\bar{g} = 0.56$, 95% CI: [−0.92, 2.03]).

DISCUSSION

The present meta-analysis aimed to investigate the efficacy of MIT while accounting for crucial methodological aspects of primary studies, such as availability of control comparisons, randomized group allocation, use of validated outcomes, and variance in MIT protocol. The meta-analysis also examined the possibility of confound through spontaneous recovery, and the degree to which MIT's effect generalizes to untrained items.

Overall, we found that MIT had a limited positive effect in specific domains, mainly repetition tasks, in line with previous meta-analyses. However, our results reveal that poor methodology may introduce substantial bias into estimated effects. Concerning RCT data of Non-Communicative Language Expression, the use of unvalidated outcomes for untrained items may attenuate MIT's effect size by about 43% when compared to validated tests ($\bar{g}_{unvalidated} = 0.20$ vs. $\bar{g}_{validated} = 0.35$). Holding domain and outcome validity constant, MIT's effect size proved to be 5.7 times larger for non-RCT data compared to RCT data ($\bar{g}_{case\ report} = 2.01$ vs. $\bar{g}_{RCT} = 0.35$ for validated Non-Communicative Language Expression measures). Implications and potential causes of these findings are discussed below.

Research implications

The current results indicate that appropriate study design can help reduce confound to obtain more realistic effect size estimates. In particular, these results re-affirm the importance of setting up and adjusting for adequate control interventions. Otherwise, most of the changes observed in case reports—evident as inflated estimates of efficacy (the 5.7 factor reported)[c]—are inseparable from phenomena of spontaneous recovery, and ultimately regression to the mean, none of which emerge from the treatment itself.

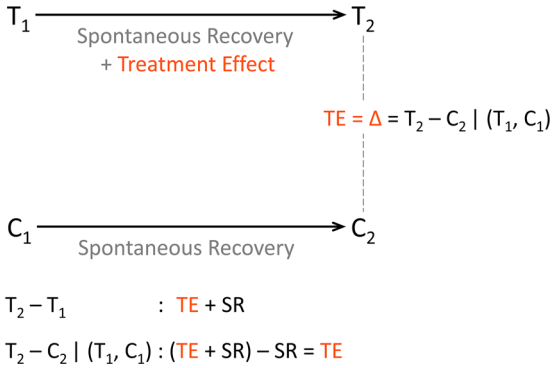

FIGURE 5 Treatment effects (TE) and spontaneous recovery (SR) in interventions. T_1 and T_2 represent the treatment group at time 1 and time 2 (pre- and post-treatment). C_1 and C_2 represent the control group at the same two time points. The | operator denotes "accounting for." The first equation shows how TE may be confounded with SR. The second equation shows how the confounding effect of SR can be removed (for a causal diagram, see also Figure S1).

Figure 5 schematically illustrates the need for a control group to estimate the treatment effect, net of possible influences resulting from time post-stroke only, such as the impact of spontaneous recovery. Case series report T_2–T_1 and tend to interpret it as the treatment effect. However, this confounds treatment effect with spontaneous recovery. To isolate the treatment effect, a control group is needed; from it, we compute T_2–C_2, accounting for baseline differences at T_1. Figure S1 further illustrates this issue.

Control interventions were drastically different among our three RCTs, namely, "control therapy not aiming at language production but using linguistic tasks often trained in severe nonfluent aphasia, such as written language production, language comprehension, and nonverbal communication strategies";[9] "no individual treatment offered, [i.e., only] social interaction as well as low-intensity group therapy to support verbal and nonverbal communication";[10] and "none, or waiting list treatment."[31] Future research and its aggregation in meta-analyses would benefit from control groups with standardized and empirically validated types of intervention, number of weekly sessions, and hours of daily training.

Effect sizes were found to decrease with number of MPO for IPD studies, indicating that progress in language performance reported in the late subacute or consolidation stage of aphasia may arise from influences of spontaneous recovery.[d] Currently available data do not allow concluding whether MIT's effect size increases or decreases with MPO, given the general lack of positive RCT evidence on

[c] See also Table S11, which reports RCT meta-analyses only considering the change in control groups.

[d] See also Table S12, which reports IPD meta-analyses with MPO as a moderator for pre-test scores only.

speech-language therapy in subacute aphasia relative to a comparator.[16] Taken together, these results suggest that validated outcomes, RCT designs, and inclusion of individuals with chronic aphasia are essential prerequisites to determine the efficacy of MIT in a reliable way.

The chronically re-organized language system post-stroke involves undamaged perilesional tissue in the left hemisphere, as well as homotopic areas in the right hemisphere.[32] However, the concrete distribution of activity depends on the time elapsed after stroke. In our meta-analysis, the influence of time post-onset of stroke was seen only for IPD, not for RCT data. We submit two potential reasons for this differential finding: first, it may be a statistical artifact (fewer data points in the RCT category); second, substantial heterogeneity may exist among post-stroke recovery trajectories, which does not "subtract out" during a treatment-versus-control comparison.[33] The latter conjecture would require consolidation by behavioral data in analogy to neurobiological models of language recovery.[34]

Clinical implications

According to the present meta-analysis, MIT leads to gains mainly in repetition tasks that reflect the ability to re-produce prior utterances in exactly the same form. Although this ability may facilitate the acquisition of novel words, it is not entirely clear to what extent it ultimately affects verbal behavior in everyday communicative situations.[35] Our RCT results indicate negligible progress on validated outcomes of everyday communication ability with MIT. The number of non-repetition outcomes was comparatively small, regardless of experimental design, implying that benefits from MIT cannot be ruled out completely; nonetheless, current evidence does not support them. In contrast, large-scale RCT data demonstrate that combining selected non-MIT methods *can* lead to moderate gains on validated outcomes of communication ability.[2] This finding suggests that individuals with aphasia should not rely exclusively on MIT if the primary goal is to improve everyday communication. Still, our meta-analysis should not be taken to downplay the importance of MIT-mediated progress on trained items. In individuals with severe forms of aphasia, "palliative" use of MIT may entail a substantial increase in quality of life.[15] Critically, individuals with aphasia may perceive notable progress in language performance irrespective of statistically significant gains on validated outcomes. Known as "minimal clinically important difference,"[36] this diagnostic approach may be especially valuable for individuals, where MIT can help establish a repertoire of trained phrases to convey basic needs in daily life.[37] In addition, future studies will hopefully explore the impact of MIT on quality of life, emotional well-being, and severity of post-stroke depression in individuals with neurological communication disorders.[38]

Limitations and future directions

As with any meta-analysis, the strength of the results depends on the quantity and quality of the source material. Rigorous eligibility criteria left us with a relatively low number of included studies ($n = 22$) and number of participants ($n = 129$). This small sample size in turn may be responsible for large confidence intervals, which necessarily limit the conclusiveness. Future meta-analyses will be able to make recommendations with greater certainty, provided that subsequent studies overcome methodological issues, pointed out above.

As it stands, methodological issues in our included studies call for caution in interpreting the results. To address these issues, we considered various aspects largely neglected in previous work. In particular, our meta-analysis carefully determined the psychometric quality of each outcome, relative to recently defined standards in aphasia research.[39] Moreover, our evaluation accounted for quality of the research design in terms of using control interventions and group randomization to reflect unspecific influences, including bias due to placebo effects. Our results confirm the overall efficacy of MIT in repetition tasks, albeit to a smaller degree than previously reported.

Interestingly, deviation from the original MIT protocol did not systematically alter the effect size. This finding casts doubt on the notion that the original composition and hierarchical structure of MIT are indispensable for improving language performance. A number of studies included in our meta-analysis employed a modified MIT protocol, and their individual effects are heterogeneous. Therefore, our results cannot express certainty about the impact of MIT protocol variation, and instead highlight the need for high-quality research on the influence of specific modifications.

Based on unvalidated outcomes, cross-sectional and longitudinal multiple-case studies have examined the role of different MIT elements: melody and rhythm,[40] vocal expression in unison or alone,[41] left-hand tapping,[42] and formulaicity of verbal utterances.[43] Possible methodological reasons for seemingly contradictory data, as well as conjectured mechanisms of MIT, have been discussed.[44] Obviously, the present results do not offer insight into any of these mechanisms. However, our results may encourage future research to optimize the composition of the treatment to increase its efficacy in the rehabilitation of neurological communication disorders. For example, individuals with apraxia of speech may benefit from several elements of MIT, such as rhythmic pacing[45] and language formulaicity.[46] For now, the assumed benefit of MIT in individuals with impaired speech-motor planning remains a highly plausible hypothesis that our meta-analysis cannot properly address, given the lack of RCT data for the mentioned patient population ($n = 8$ individuals across $k = 2$ IPD studies). We still believe that pursuing this particular avenue in future work may be extremely promising.

CONCLUSION

We here present the first meta-analysis on MIT that attempts to monitor the effects of various methodological caveats in interpreting the outcome of previous studies, such as lack of validated outcomes, control group, or randomization. Accounting for each of these issues in a rigorous way, the results of our meta-analysis confirm the promising role of MIT in improving trained and untrained performance on

unvalidated outcomes, alongside validated repetition tasks, and highlight possible limitations in promoting everyday communication ability. We hope that the current work will help individuals with aphasia, their families and clinicians make informed treatment decisions in the context of MIT.

ACKNOWLEDGMENTS

We gratefully acknowledge Ajay Halai and Yina Quique Buitrago for their very helpful comments on an early draft of this work; Joost Hurkmans for providing unpublished work; Sarah Wallace, Luisa Krein, and Emily Braun for providing test-related information; and two anonymous reviewers for their constructive criticism on the initial manuscript. First author T.P. accepts the responsibility for the integrity of the data analyzed.

B.M.W. is currently an independent researcher and Research Scientist at Meta, Demography and Survey Science. The current research was conducted while he was at the University of South Florida.

COMPETING INTERESTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization: T.P., B.S., and W.T.F. Methodology: T.P. and B.M.W. Software: T.P. and B.M.W. Validation: T.P. Formal analysis: T.P. and B.M.W. Investigation: T.P. Resources: T.P., B.S., and W.T.F. Literature search and curation: T.P., H.H., and M.Z. Data curation: T.P., B.M.W., F.H., and T.M. Writing—original draft: T.P. and B.S. Writing—review and editing: T.P., B.S., B.M.W., and F.H. Visualization: T.P. and B.M.W. Supervision: T.P. and B.S. Project administration: T.P. Funding acquisition: R.B. and W.T.F.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/nyas.14848

ORCID

Tudor Popescu https://orcid.org/0000-0001-5890-1520
Benjamin Stahl https://orcid.org/0000-0003-3957-1495
Brenton M. Wiernik https://orcid.org/0000-0001-9560-6336
Theresa Matzinger https://orcid.org/0000-0001-5414-7962
W. Tecumseh Fitch https://orcid.org/0000-0003-1830-0928

REFERENCES

1. Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., Gutzwiller, F., & Lyrer, P. A. (2006). Epidemiology of aphasia attributable to first ischemic stroke: Incidence, severity, fluency, etiology, and thrombolysis. *Stroke*, *37*, 1379–1384.
2. Breitenstein, C., Grewe, T., Flöel, A., Ziegler, W., Springer, L., Martus, P., Huber, W., Willmes, K., Ringelstein, E. B., Haeusler, K. G., Abel, S., Glindemann, R., Domahs, F., Regenbrecht, F., Schlenck, K.-J., Thomas, M., Obrig, H., De Langen, E., Rocker, R., …, Bamborschke, S. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: A randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *Lancet*, *389*, 1528–1538.
3. Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic Intonation Therapy for aphasia. *Archives of Neurology*, *29*, 130–131.
4. Gerstman, H. L. (1964). A case of aphasia. *Journal of Speech and Hearing Disorders*, *29*, 89–91.
5. Mills, C. K. (1904). Treatment of aphasia by training. *Journal of the American Medical Association*, *XLIII*, 1940–1949.
6. Yamadori, A., Osumi, Y., Masuhara, S., & Okubo, M. (1977). Preservation of singing in Broca's aphasia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *40*, 221–224.
7. Helm-Estabrooks, N., Morgan, A. R., & Nicholas, M. (1989). *Melodic Intonation Therapy*. Austin, TX: Pro-Ed., Inc.
8. Merrett, D. L., Peretz, I., & Wilson, S. J. (2014). Neurobiological, cognitive, and emotional mechanisms in Melodic Intonation Therapy. *Frontiers in Human Neuroscience*, *8*, 1–11.
9. Van Der Meulen, I., Van De Sandt-Koenderman, W. M. E., Heijenbrok-Kal, M. H., Visch-Brink, E. G., & Ribbers, G. M. (2014). The efficacy and timing of Melodic Intonation Therapy in subacute aphasia. *Neurorehabilitation and Neural Repair*, *28*, 536–544.
10. Van Der Meulen, I., Van De Sandt-Koenderman, M. W. M. E., Heijenbrok, M. H., Visch-Brink, E., & Ribbers, G. M. (2016). Melodic Intonation Therapy in chronic aphasia: Evidence from a pilot randomized controlled trial. *Frontiers in Human Neuroscience*, *10*, 1–9.
11. Doppelbauer, L., Mohr, B., Dreyer, F. R., Stahl, B., Büscher, V., & Pulvermüller, F. (2021). Long-term stability of short-term intensive language-action therapy in chronic aphasia: A 1–2 year follow-up study. *Neurorehabilitation and Neural Repair*, *35*, 861–870.
12. RELEASE Collaborators. (2021). Predictors of poststroke aphasia recovery: A systematic review-informed individual participant data meta-analysis. *Stroke*, *52*, 1778–1787.
13. Blomert, L., Kean, M. L., Koster, C., & Schokker, J. (1994). Amsterdam—Nijmegen everyday language test: Construction, reliability and validity. *Aphasiology*, *8*, 381–407.
14. Van Der Meulen, I., Van De Sandt-Koenderman, M. E., & Ribbers, G. M. (2012). Melodic Intonation Therapy: Present controversies and future opportunities. *Archives of Physical Medicine and Rehabilitation*, *93*, S46–S52.
15. Zumbansen, A., Peretz, I., & Hébert, S. (2014). Melodic Intonation Therapy: Back to basics for future research. *Frontiers in Neurology*, *5*, 1–11.
16. Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews*.
17. Zumbansen, A., & Tremblay, P. (2018). Music-based interventions for aphasia could act through a motor-speech mechanism: A systematic review and case–control analysis of published individual participant data. *Aphasiology*, *33*, 466–497.
18. Haro-Martínez, A., Pérez-Araujo, C. M., Sanchez-Caro, J. M., Fuentes, B., & Díez-Tejedor, E. (2021). Melodic Intonation Therapy for post-stroke non-fluent aphasia: Systematic review and meta-analysis. *Frontiers in Neurology*, *12*, 1303.
19. Baker, F. A. (2000). Modifying the Melodic Intonation Therapy program for adults with severe non-fluent aphasia. *Music Therapy Perspectives*, *18*, 110–114.
20. Jungblut, M. (2009). SIPARI(R): A music therapy intervention for patients suffering with chronic, nonfluent aphasia. *Music and Medicine*, *1*, 102–105.
21. de Bruijn, M., Zielman, T., & Hurkmans, J. (2005). *Speech-music therapy for aphasia (SMTA)*. Beetsterzwaag: Revalidatie Friesland.
22. Akanuma, K., Meguro, K., Satoh, M., Tashiro, M., & Itoh, M. (2016). Singing can improve speech function in aphasics associated with intact right basal ganglia and preserve right temporal glucose metabolism: Implications for singing therapy indication. *International Journal of Neuroscience*, *126*, 39–45.
23. Stahl, B., Henseler, I., Turner, R., Geyer, S., & Kotz, S. A. (2013). How to engage the right brain hemisphere in aphasics without even singing:

Evidence for two paths of speech recovery. *Frontiers in Human Neuroscience*, *7*, 1–12.
24. Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2021). *Cochrane handbook for systematic reviews of interventions*. Cochrane.
25. Moher, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264.
26. Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*, 315–346.
27. Ivanova, M. V., & Hallowell, B. (2013). A tutorial on aphasia test development in any language: Key substantive and psychometric considerations. *Aphasiology*, *27*, 891–920.
28. Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science*, *3*, 94–123.
29. Van Smeden, M., Lash, T. L., & Groenwold, R. H. H. (2020). Reflection on modern methods: Five myths about measurement error in epidemiological research. *International Journal of Epidemiology*, *49*, 338–347.
30. Inthout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, *6*, e010247.
31. Haro-Martínez, A. M., Lubrini, G., Madero-Jarabo, R., Díez-Tejedor, E., & Fuentes, B. (2019). Melodic Intonation Therapy in post-stroke nonfluent aphasia: A randomized pilot trial. *Clinical Rehabilitation*, *33*, 44–53.
32. Turkeltaub, P. E., Messing, S., Norise, C., & Hamilton, R. H. (2011). Are networks for residual language function and recovery consistent across aphasic patients? *Neurology*, *76*, 1726–1734.
33. Doppelbauer, L., Mohr, B., Dreyer, F. R., Stahl, B., Büscher, V., & Pulvermüller, F. (2021). Long-term stability of short-term intensive language–action therapy in chronic aphasia: A 1–2 year follow-up study. *Neurorehabilitation and Neural Repair*, *35*, 861–870.
34. Saur, D. (2006). Dynamics of language reorganization after stroke. *Brain*, *129*, 1371–1384.
35. Stahl, B., Mohr, B., Dreyer, F. R., Lucchese, G., & Pulvermüller, F. (2017). Communicative-pragmatic assessment is sensitive and time-effective in measuring the outcome of aphasia therapy. *Frontiers in Human Neuroscience*, *11*, 1–9.
36. Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, *61*, 102–109.
37. Van Lancker Sidtis, D. (2021). *Foundations of familiar language: Formulaic expressions, lexical bundles, and collocations at work and play* (1st ed.). Hoboken, NJ: Wiley-Blackwell.
38. Stahl, B. (2021). *Music and social interaction in the treatment of post-stroke aphasia*. Thesis for: Habilitation. Medical School Berlin.
39. Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A.-C., Marshall, J., …, Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke*, *14*, 180–185.
40. Kershenbaum, A., Nicholas, M. L., Hunsaker, E., & Zipse, L. (2019). Speak along without the song: What promotes fluency in people with aphasia? *Aphasiology*, *33*, 405–428.
41. Racette, A. (2006). Making non-fluent aphasics speak: Sing along! *Brain*, *129*, 2571–2584.
42. Zipse, L., Worek, A., Guarino, A. J., & Shattuck-Hufnagel, S. (2014). Tapped out: Do people with aphasia have rhythm processing deficits? *Journal of Speech, Language, and Hearing Research*, *57*, 2234–2245.
43. Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: Why singing may not hold the key to recovery from aphasia. *Brain*, *134*, 3083–3093.
44. Stahl, B., & Kotz, S. A. (2014). Facing the music: Three issues in current research on singing and aphasia. *Frontiers in Psychology*, *5*, 1033.
45. Aichert, I., Späth, M., & Ziegler, W. (2016). The role of metrical information in apraxia of speech. Perceptual and acoustic analyses of word stress. *Neuropsychologia*, *82*, 171–178.
46. Stahl, B., Gawron, B., Regenbrecht, F., Flöel, A., & Kotz, S. A. (2020). Formulaic language resources may help overcome difficulties in speech-motor planning after stroke. *PLoS One*, *15*, e0233608.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Popescu, T., Stahl, B., Wiernik, B. M., Haiduk, F., Zemanek, M., Helm, H., Matzinger, T., Beisteiner, R., & Fitch, W. T. (2022). Melodic Intonation Therapy for aphasia: A multi-level meta-analysis of randomized controlled trials and individual participant data. *Ann NY Acad Sci.*, *1516*, 76–84. https://doi.org/10.1111/nyas.14848