Aphasiology

Routledge
Taylor & Francis Group

ISSN: 0268-7038 (Print) 1464-5041 (Online) Journal homepage: https://www.tandfonline.com/loi/paph20

Perspectives from case studies in obtaining evidence for music interventions in aphasia

Dawn L. Merrett, Chris Tailby, Graeme D. Jackson & Sarah J. Wilson

APHASIOLOGY
2019, VOL. 33, NO. 4, 429–448
https://doi.org/10.1080/02687038.2018.1428729

Perspectives from case studies in obtaining evidence for music interventions in aphasia

Dawn L. Merrett[a], Chris Tailby[a,b], Graeme D. Jackson[b] and Sarah J. Wilson[a,b]

[a]Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, VIC, Australia; [b]The Florey Institute of Neuroscience and Mental Health, Melbourne, VIC, Australia

ABSTRACT

Background: Music interventions for aphasia, such as Melodic Intonation Therapy, have often been criticised for a lack of high-quality evidence regarding their efficacy and mechanisms of action. However, attempts to evaluate these interventions and produce an evidence base for or against their use have proven challenging.

Aims: We discuss both the challenges in obtaining research evidence and some possible solutions, taking into perspective differences between clinical and research approaches in their aims, orientation, and methodology. Research is generally focused on standardisation, generalisability, and the provision of adequately powered and statistically sound evidence. In contrast, clinical work is usually client-centric, requiring flexibility to address the needs of the individual patient. To illustrate these points, we present case studies of two individuals with chronic post-stroke aphasia, who were pilot participants for a music intervention study.

Methods and Procedures: These patients received research-oriented treatment with a standardised audio-visual Melodic Intonation Therapy protocol delivered via DVD over 6 weeks. They underwent comprehensive language assessments before and after therapy, which included functional neuroimaging for one individual.

Outcomes and Results: This standardised approach provided modest clinical benefit in one case, although this was not captured in standardised outcome assessments. For the second, more severely aphasic participant, there was no observable benefit, possibly because the standardised approach did not provide the flexibility needed to deal with the severity of his deficit.

Conclusions: Through presentation of these case examples, we highlight how heterogeneous clinical presentations and individual differences pose challenges to standardised research designs. We then offer suggestions for how these factors might be accommodated within rigorous research designs to provide a better evidence base for aphasia interventions.

ARTICLE HISTORY
Received 31 August 2017
Accepted 11 January 2018

KEYWORDS
Evidence-based practice; Melodic Intonation Therapy; standardisation; aphasia rehabilitation

CONTACT Sarah J. Wilson sarahw@unimelb.edu.au Melbourne School of Psychological Sciences, The University of Melbourne, 12th Floor, Redmond Barry Building, Melbourne, VIC 3010, Australia

Supplemental data for this article can be accessed here.

Introduction

Evidence-based practice has been the goal of clinical work in most health-related fields, including aphasia rehabilitation, for many years now (Patterson & Coppens, 2017). Yet, a recurring criticism of numerous aphasia treatments is the lack of a sound evidence base. This is particularly apparent in the domain of music-based interventions for language impairments. Despite a growing research interest in the topic, obtaining appropriate evidence for various musical therapeutic approaches has been somewhat elusive. For example, Melodic Intonation Therapy (MIT; Albert, Sparks, & Helm, 1973; Sparks, Helm, & Albert, 1974) is a well-known form of music therapy that has been used in the treatment of individuals with non-fluent aphasia and often other speech and language disorders. In 1994, the American Academy of Neurology released a report stating that MIT was a promising treatment for non-fluent aphasia with expert opinion and case series evidence (Class III) for efficacy (Benson, Dobkin, Gonzalez Rothi, Helm-Estabrooks, & Kertesz, 1994). In the intervening decades the therapy has been used clinically, but at this time – more than 40 years since the original publications – is still lacking "gold standard" evidence for efficacy and high-quality, well-powered investigations into its mechanisms of action. While numerous research groups are working to provide these, the challenges that they are encountering in doing so (for instance, van der Meulen, van de Sandt-Koenderman, Heijenbrok, Visch-Brink, & Ribbers, 2016) make it worthwhile asking what impediments are contributing to the difficulty in researching these types of treatments.

To start, music-based interventions should be considered within their multi-disciplinary context. Our current understanding of music-based aphasia treatment is dependent on contributions from specialists in music psychology and music therapy, speech and language pathology, neuroimaging, neurology, and neuroscience, and it is likely within this multi-disciplinary framework that new insights will continue to be formed. However, in order to maximise the potential of cross-disciplinary work, it is necessary to acknowledge some of the tensions that arise when the disciplines involved have different orientations, aims, and methodologies (Aagaard-Hansen, 2007). This is perhaps most obvious in the differences between clinically- and research-focused disciplines, or even between clinical and research foci within a given discipline. The inherent tensions between research and clinical work are often unacknowledged or overlooked, even in the domain of "clinical research", and may have a great deal of explanatory power regarding the difficulty in obtaining adequate evidence for therapeutic efficacy. These tensions include standardisation versus customisation, generalisation versus individual differences, and the role of the group versus the individual in providing new scientific insights. These are discussed here in the context of aphasia rehabilitation, but are more broadly applicable to other clinical and research spheres.

A major component of most research is the use of a standardised research protocol, ensuring that any participants or groups are treated or managed in a systematic way that can be reliably replicated. This approach minimises the variance associated with the experiment and the experimenters, and increases confidence in the statistical measures that usually form the basis of research outcomes. Despite its importance, standardisation naturally limits flexibility, and some degree of flexibility is useful, and perhaps ethically necessary, within a clinical context. Clinicians such as speech-language pathologists and music therapists are expected to apply clinical judgment in conjunction with evidence-based treatment protocols to ensure that they are adequately addressing the needs of their patients. Although modifying a prescribed therapy protocol for an individual participant in a research context might invalidate the research findings, modifying the protocol in a clinical context might lead to a better outcome for the individual, which is the primary aim of clinical work. For this reason, many clinicians engage in customisation, even of rather tightly prescribed therapy protocols, for their individual patients. While customisation can take many forms, it often involves (1) changing elements to make a protocol easier for a severely affected patient, (2) increasing or decreasing the intensity or duration of a therapy, or (3) adapting the therapy to the preferences of the patient for better engagement. For instance, MIT has been customised in each of these ways (see recent reviews by Merrett, Peretz, & Wilson, 2014 and Zumbansen, Peretz, & Hébert, 2014a for examples). Many of these customisations appear to have been helpful for the patient(s) to whom they were applied, with positive outcomes reported in many case studies of customised MIT. Unfortunately, this also means that within the research literature, there are almost as many different MIT protocols as there are reported cases/studies, with no real consensus as to what might constitute an optimal protocol within a given context. Although best practice guidelines for customisation, i.e., standardised customisation, have been imbedded into some aphasia therapies, MIT does not yet have a sufficient evidence base in this regard.

Another aim of research is generalisability, which is the capacity to apply the conclusions from the group of participants under study to the larger population. This requires that the group being studied is representative of the larger population. It also typically requires that the group is randomly selected and large enough to average out the naturally occurring differences between individuals that might influence the results. Realistically, these requirements for generalisability are both theoretically and practically difficult to achieve within a therapy framework. Clinical research samples are often highly selective and tend to be case studies, case series, or small groups. This is due to the challenge of finding representative and/or homogeneous clinical groups, given the incredible diversity of clinical presentations, as well as the time commitment and cost involved in large clinical studies. Despite attempts to overcome these hurdles, with some degree of success, music-based aphasia therapy has often been criticised for the lack of well-powered group studies to provide a statistically sound and generalisable evidence base.

The importance of a solid evidence base for rehabilitation practice cannot be overstated. Yet because of the intrinsic patient-centric nature of clinical work, there is a definite tension between an individual's current need for personalised treatment and the need to develop best-practice interventions at a group level that are statistically likely to generalise to other individuals. A focus on the individual does not preclude group-level statistical comparisons that form the basis of empirical studies, but it does complicate them. If successful clinical practice involves customisation and responsiveness to individual differences, then imposing the typical research framework of standardisation and generalisability to aphasia therapy investigations could lead to situations where the therapy under study no longer approximates clinical practice. Flow-on effects might include inaccuracy of the evidence base, underestimation of therapeutic efficacy, and/or use of less-than-ideal modifications of therapy protocols in practice.

To illustrate these issues and frame further discussion, we present two case studies of individuals who were pilot participants for a longitudinal aphasia rehabilitation study investigating the neurobiological mechanisms of MIT. Approaching this study primarily from a music neuroscience research perspective, the aim was to standardise the therapy and assessment and to eliminate known sources of variance to the greatest degree possible. Given the challenges in providing time-intensive therapy to large numbers of patients, an additional aim was to enable this study to be conducted on a larger scale with a small research team. To this end, we developed a DVD-based (digital video disc) protocol using the principles of MIT, including humming and singing with a therapist, a slow rate of intonation, hand-tapping, and a hierarchical progression across time towards longer phrases and natural speech (see *Methods: DVD Training Protocol* for further details). The DVD therapy was unable to accommodate the "backing-up" procedure often implemented in MIT for unsuccessful attempts at phrases (see Sparks, 2008), but was otherwise highly similar in form and content to typically reported MIT practice.

Methods

Participants

The pilot participants were two individuals with chronic non-fluent aphasia following left hemisphere stroke, recruited through community referrals. This study received approval from the Austin Health Human Research Ethics Committee, and all participants gave written informed consent in accordance with the Declaration of Helsinki.

Case A

Case A was a 66-year-old right-handed male with severe chronic non-fluent aphasia following a left middle cerebral artery infarct 2.5 years previously. He also presented with mild dysarthria, moderate right hemiparesis, a right homonymous hemianopia, post-stroke depression, disinhibition, and executive dysfunction. Although his receptive language was mostly intact, his communication skills were poor, and his expressive language was extremely limited. He had previously been given extensive formal speech therapy as an inpatient, including unsuccessful attempts to use a communication book to increase his communicative capacity, and he had participated in some outpatient group communication sessions, with poor results. Case A was a unilingual English speaker who left secondary school prior to graduation. He had no previous training in music, although he listened to music frequently prior to the stroke and enjoyed singing. Post-stroke, he was still able to sing the lyrics of some songs, such as "Happy Birthday" and snippets of old folk songs that were highly familiar.

Case B

Case B was a 62-year-old right-handed male with severe chronic non-fluent aphasia and apraxia following a major stroke during surgery involving the left carotid artery 18 months previously. The affected neural areas were primarily left frontal, particularly subcortical white matter and basal ganglia structures, with stroke-related cortical atrophy also extending into the parietal lobe. Sequelae included right hemiparesis, severely affecting the upper limb, and a decline in cognitive function. Case B was well educated

(24 years total education) and extremely high functioning prior to the stroke. He was a native English speaker with rudimentary fluency in a second language, with 1 year of formal training in music and some experience singing in a choir. Of note, Case B closely resembled an "ideal" MIT candidate, as described in some of the therapy creators' reports (Helm-Estabrooks & Albert, 2004; Sparks et al., 1974), with one exception. In particular, he presented with a unilateral lesion, nonsense stereotyped output, well-preserved auditory comprehension, high motivation, emotional stability, and a good attention span; however, he was unable to produce any accurate words when singing familiar songs. He occasionally was able to approximate vowel sounds, but mostly hummed the tunes. Also, it was not possible to disentangle the influence of co-morbid apraxia versus non-fluent aphasia on Case B's speech output, which consisted entirely of nonsense babble and two cued words. Given that MIT is thought to act not only on aphasic language function, but also on articulatory motor control (Zumbansen et al., 2014a), Case B was enrolled as a pilot participant despite his severe co-morbid apraxia.

Procedure

Both participants engaged in 6 weeks of intensive DVD-based MIT, with comprehensive behavioural assessments pre- and post-therapy. Case A was a pilot participant for the behavioural aspects of our project only, while Case B also underwent functional neuroimaging before and after therapy to examine any potential change at the neural level. The protocol consisted of 3 DVDs which progressed through different stages of MIT (as described by Sparks, 2008). Each DVD was rehearsed 10 times over 2 weeks for a total of approximately 30 h of MIT over 6 weeks. Four sessions each week were conducted at home, with minor assistance from caregivers to set up the DVD, to ensure that the sound was set at a comfortable volume and that the viewing distance and angle were appropriate, and to record adherence to the protocol in a practice diary. The fifth session each week was conducted with the first-named researcher, to ensure that the protocol was followed correctly.

Materials

Behavioural assessments

Pre- and post-MIT behavioural assessments consisted of a battery of standardised assessments covering the domains of speech, language, mood, cognition, and music. In particular, speech and language function was assessed using the Boston Diagnostic Aphasia Examination, 3rd Ed (BDAE; Goodglass, Kaplan, & Barresi, 2001), the Frenchay Dysarthria Assessment, 2nd Ed (Enderby & Palmer, 2008), the Apraxia Battery for Adults, 2nd Ed, (Dabul, 2000), word retrieval/generation tasks (FAS, Animals; Strauss, Sherman, & Spreen, 2006), and spoken phrases specific to the MIT training in the current study.

For the trained MIT phrases, repetition accuracy was scored at the syllable and phrase level. The participant heard each phrase only once, unless no response was attempted, at which point the phrase was repeated. The participant's first attempt was scored, although within phrase corrections of whole syllables (for example, "I li … love you") were allowed for scoring purposes. In order for a syllable to be counted as correct, each phoneme of the syllable had to sound correct when the audio recording of the task was reviewed at full speed. For a phrase to be counted as correct, all syllables of the phrase had to be correct. Repetition accuracy was rated by two independent raters, one of whom was unfamiliar with the participant. The interclass correlation coefficient estimate and its 95% confidence interval were computed using IBM SPSS 24 (https://www.ibm.com/products/spss-statistics/) based on an absolute-agreement 2-way mixed effects model, showing an excellent inter-rater reliability of .981 (95% CI: .969–.988).

Since verbal aspects of intelligence are not readily testable in aphasic participants, the Perceptual Reasoning Index of the Wechsler Adult Intelligence Scale IV (Pearson PsycCorp, Sydney, Australia) was used to obtain a measure of non-verbal intellectual functioning. However, it should be noted that the Perceptual Reasoning Index includes tasks that have motor and visual components, which may reduce the validity of this measure in stroke patients with motor or visual impairments. Auditory attention was assessed using the Brief Test of Attention (Psychological Assessment Resources, Lutz, FL), and an adaptation of the Rivermead Behavioural Memory Test for aphasic individuals (Cockburn, Wilson, Baddeley, & Hiorns, 1990) was used to probe basic memory functions.

The Profile of Mood States (McNair, Lorr, & Droppleman, 1971) was used to assess mood and affect for Case A prior to MIT, but even with assistance he found it extremely difficult to complete and was not willing to repeat it post-MIT. Given that other aphasic individuals could have similar difficulties, simple visual analogue mood scales were chosen to replace the Profile of Mood States and were used with Case B. They provided an additional advantage in that they could be completed prior to and after each MIT session with the researcher, allowing within-session effects of MIT to be investigated using paired t-tests. Finally, the Montreal Battery of Evaluation of Amusia (Peretz, Champod, & Hyde, 2003) was administered to determine whether the participants had any music processing anomalies in pitch, rhythm, and musical memory domains. Since musical functions are often disrupted after stroke (Särkämö et al., 2009), it was considered important to take any potential ongoing music processing difficulties into account when administering the music-based intervention.

DVD training protocol

Each DVD contained 15 phrases to be rehearsed. For measurement purposes, five of the initial phrases were retained throughout the training, while five of the phrases were extended and five novel phrases were introduced in each of the second and third DVDs, for a total of 35 trained phrases. In keeping with the protocol described by Norton, Zipse, Marchina, and Schlaug (2009), the phrases became progressively longer with each stage. In the first DVD, the phrases ranged from two to four syllables, and included phrases such as “Watch TV”, “Good morning”, and “I need help”. In the second DVD, the extended and novel phrases varied from four to six syllables, and in the third DVD, they ranged from six to nine syllables, for example, “Listen to classical music”, “I would like to buy a stamp”, and “Let’s go to the supermarket”. The phrases consisted of both functional and content words and covered a wide variety of consonant types, consonant clusters, vowel sounds, and diphthongs.

Each of the MIT phrases was set to a unique tune that was a rough exaggeration of the phrase’s normal prosodic contours. This is in contrast to the MIT method that was developed by Helm-Estabrooks (1983) in which only two different notes–a high note and low note–are used, typically a minor third apart (Norton et al., 2009). Although singing on a minor third has become fairly common practice in MIT, the current protocol used a greater variety of melodic patterns, consistent with the original concept of MIT (Sparks et al., 1974). It was felt that this would make the intensive training more enjoyable, and that it might help promote access to/retrieval of the phrases from memory after training, a hypothesised mechanism for MIT efficacy (Wilson, Parsons, & Reutens, 2006). Syllables that would be accented in English were usually placed on higher notes and unaccented syllables on lower notes. The tunes used a fairly restricted pitch range, with the range of all tunes equal to or less than a major 6th.

The DVD training included other specific aspects of contemporary MIT protocols such as left hand-tapping, a one-second per syllable rate of intonation, a progression to longer and more linguistically complex phrases, and a gradual shift to spoken language at later stages of the training. Participants were reminded with each phrase to tap along with their left hand, while also watching the audiovisual model tap with the right hand (as in a mirror). The phrases were introduced by humming the tune of the phrase, and then asking the participant to hum along. Next, the words of the phrase were sung, and then the participant was asked to sing along. In the next step, the participant was asked to sing along, and the audiovisual model gradually faded out, leaving the participant to sing independently. Next, the participant was asked to copy the phrase sung by the model. In the final step, the model sang the question, "What did you say?", followed by a pause for the participant to try to respond with the sung phrase. Over the second and third DVDs, the time between hearing the model and copying the phrase and the time between hearing the sung question and initiating independent production was modulated with hand gestures (Helm-Estabrooks & Albert, 2004). Delaying the response in this way requires controlled initiation from the participant. In the third and final DVD, normal prosody was introduced first through the use of *sprechgesang*, a vocal style between speech and singing (Sparks, 2008), and then by having the participant both copy and initiate a response to a question using speech rather than singing.

Practice diaries

To canvass details regarding protocol adherence for at-home DVD sessions, practice diaries were created to enable participants and/or their carers to record their practice dates and times and to record any irregularities (e.g., interruptions), observations (e.g., engagement level, fatigue), or questions/concerns. Given the physical limitations of the participants, the carers were expected to do most of the recording and were therefore carefully instructed along with the participants in the use of the diaries. The diaries were reviewed weekly during the in-clinic sessions and retained at the end of the protocol.

Functional imaging protocol

For Case B, both overt speaking and singing were evaluated during separate functional runs in imaging sessions prior to and following the 6 weeks of MIT. One speaking task and two different singing tasks were used, with one singing task trained intensively during MIT (sing-trained) and the other singing task not trained (sing-untrained), to allow the investigation of potential training and generalisation effects. The phrases used for the speaking and singing tasks consisted of 5 syllables each, with one phrase for each task ("pretty little bird", "fluffy little chick", "gentle little duck"). They were performed at a slow rate of approximately 1.7 syllables/s. These phrases were introduced to the participant in a standardised format the day before the baseline imaging session. A mock scan was also conducted at this time to acquaint the participant with the equipment, the scanner noise, and the feeling of singing and speaking in the scanner. On the day of the scanning sessions, Case B was asked to rehearse the tasks outside of the scanner. During rehearsal and during the actual scanning session, templates of the phrases (both sung and spoken) were played at the beginning of each run to remind him of the correct task for that particular run and to promote a consistent rate of execution. Recordings of Case B's in-scanner performances were obtained to measure the extent to which he was able to complete the task correctly.

The fMRI protocol design was a mixed (block and event-related) sparse temporal sampling design, where task blocks comprised jittered singing or speaking events that occurred between scans. This design was chosen to minimise the possibility of movement artefacts associated with overt vocal production in the scanner. Each sung or spoken phrase was repeated four times within each block, and four task blocks alternated with four rest blocks in each run. The tasks were jittered in relation to image acquisition by precise cueing using a visual display controlled by Presentation software (Presentation 15.0, Neurobehavioral Systems, Berkeley, CA, www.neurobs.com). Images were acquired on a 3-tesla Siemens Trio system (Siemens, Erlanger, Germany) using a multi-slice Echo Planar Imaging (EPI) sequence (single shot gradient recalled echo) providing T2*-weighted blood-oxygenation-level dependent contrast. A single whole brain volume was acquired in 3 s during each 12 s effective repetition time (TR), with a total of 33 volumes in each run. A high resolution T1-weighted anatomical image was obtained on which to overlay the functional images after co-registration.

The functional images were pre-processed and analysed using Statistical Parametric Mapping Software (SPM8r6313; Wellcome Trust Centre for Neuroimaging, London, UK) with the aid of the iBrain^{TM} analysis toolbox for SPM (Abbott et al., 2011). Due to T1-bleedthrough between adjacent interleaved slices of each volume, the slices obtained in the second half of each TR (alternate slices) were excluded from the analysis, effectively creating a 3mm gap between slices. Pre-processing included realignment, 2D motion correction, and 2D smoothing. Since standard slice timing correction cannot be applied in sparse temporal sampling, differences in the acquisition time of each slice were addressed by performing a separate statistical analysis for each slice. The individually modelled slices were recompiled to create the statistical parametric maps for each task (singing minus rest and speaking minus rest) at each time point (pre- and post-MIT). Each map was initially thresholded at an alpha level of .001, and then corrected by controlling the topological false discovery rate (FDRc). The cluster size required to retain above-threshold voxels was determined by setting the FDRc to p = .05.

Given the potential variability of the haemodynamic response after stroke and aphasia rehabilitation therapy (Peck et al., 2004), and the considerable in-scanner head motion of Case B (see below), functional runs were also analysed using independent component analysis (ICA) – a data-driven multivariate analysis technique. The data were prepared for ICA by denoising each session with high pass filtering (period = 128s) and motion regression. ICA was implemented using MELODIC v3.14 in FSL (FMRIB Software Library, www.fmrib.ox.ac.uk/fsl; Beckmann & Smith, 2004; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012), using the temporal concatenation approach. To identify potential task-related components, a multisession GLM (multiple regression) was fit to the time course for each component, modelled as a separate active-vs.-rest boxcar function for each session. Contrasts were then calculated on the resulting parameter estimates.

Motion parameters for Case B, obtained through rigid body realignment to correct for head motion during one of the singing tasks, were compared to motion parameters from a group of healthy controls performing a similar singing task using one-sample t-tests. The alpha level was set to p = .05, with Bonferroni correction for multiple comparisons.

Results

Case A initially struggled to get into the routine of using the DVDs and to concentrate throughout, but improved markedly in this regard from weeks 4 through 6. He completed 31 DVD sessions in total, with fewer repetitions than requested for the first 2 weeks (8/10 for DVD 1) and more repetitions than requested for weeks 3 and 4 (13/10 for DVD 2). During initial sessions with the researcher and also at home (as reported by his carer), Case A initiated frequent interruptions to the DVD protocol; however, this was no longer the case by the fourth week. Around this time, he began enjoying the DVDs to the extent that he asked to do the DVDs multiple times per day. He followed the instructions on the DVDs reasonably well, but often forgot to perform the tapping component of the protocol. This could be related to executive dysfunction, or it could possibly be the result of a restricted field of view due to his hemianopia. Although he seemed to compensate well for the hemianopia during the sessions, he may not have been able to fully see the tapping of the model and would have had to rely more on the auditory reminders to tap.

Case A exhibited nearly static scores on the BDAE and other standardised tests of speech, language, and cognitive function before and after MIT (Table 1). However, he showed objective improvement in his ability to speak the phrases learned in training, and in the generation of trained phrases in response to questions. Before training, he was able to correctly repeat 17% of the phrases used in training, with 52% of the syllables in those phrases generated correctly. After training, his repetition accuracy on these same phrases increased to 54%, with 81% of the syllables generated correctly. Given that all of Case A's utterances were empty or subclausal, an alternative scoring method for discourse analysis (Nicholas & Brookshire, 1993) was applied to the free conversation and picture description segments of the BDAE. This analysis suggested that the percentage of correct and informative words increased after training, even though fewer total words were spoken after training than at baseline.

Case A also showed subjective improvements in mood, attention, and the use of verbal communication in daily life, noted by both the researchers and caregivers. Although he refused to complete the formal post-MIT mood assessment, increased positive mood and decreased negative mood were clinically evident. He also demonstrated a greater capacity to attend to the DVDs and to sustain attention in social situations. Examples were noted of him using phrases learned on the DVDs in appropriate conversational contexts. Given the extreme paucity of speech output Case A displayed prior to MIT, the expansion of his vocabulary to include phrases that he learned on the DVDs was in itself a highly salient outcome. Zumbansen et al. (2014a) refer to this as *palliative* use of MIT, with the therapy providing a trained functional

Table 1. Behavioural assessment results for Cases A and B before and after DVD-based Melodic Intonation Therapy.

	Case A		Case B	
	Pre-MIT	*Post-MIT*	*Pre-MIT*	*Post-MIT*
SPEECH & LANGUAGE ASSESSMENTS				
Boston Diagnostic Aphasia Examination, 3rd Ed				
Severity rating (0–5 scale)[a]	1–2	1–2	0–1	0–1
Social responses (/7)	4.5	4.5	0.5	0.5
Free conversation (% CIUs of total words)	42.1	56.1	0	0
Picture description (% CIUs of total words)	23.5	53.3	0	0
Auditory comprehension (% correct)	90.5	97.3	81.1	86.5
Oral agility (% correct)	50	41.7	50	50
Verbal agility (% correct)	7.1	14.3	0	0
Automatised sequences (% correct)	50	75	0	0
Repetition (% correct)	30	30	0	0
Boston Naming Test (% correct)	38.3	38.3	DNC	DNC
Frenchay Dysarthria Assessment, 2nd Ed[b]				
Tongue – general tone & movement	c	c	b-c	b-c
Tongue – movement in speech	c	c	e	e
Intelligibility	d	d	e	e
Apraxia Battery for Adults, 2nd Ed[c]				
Diadochokinetic rate	Mild	Mild	Moderate	Mild
Increasing word length	Severe	Severe	DNC	DNC
Limb apraxia	None	None	Moderate	Mild
Oral apraxia	None	None	Moderate	Mild
Repeated trials	Moderate	Moderate	DNC	DNC
Word retrieval/generation (FAS, Animals)				
Orthographic lexical retrieval (# words)	0[e]	0[e]	0[e]	0[e]
Semantic retrieval (# words)	3[e]	5[e]	0[e]	0[e]
Spoken repetition of phrases (Trained)				
Correct syllables (/165)	85	134	0	2
Correct phrases (/35)	6	19	0	0
MOOD ASSESSMENTS				
Profile of Mood States				
Total Mood Disturbance	54[e]	DNC	-	-
Visual analogue mood scales[d]				
Happy (/100)	-	-	53.4	70
Sad (/100)			24.6	19.1
COGNITIVE ASSESSMENTS (Baseline only)				
Wechsler Adult Intelligence Scale IV				
Perceptual Reasoning Index	67 (Extremely Low)		81 (Low Average)	
Brief Test of Attention	Impaired		Impaired	
Rivermead Behavioural Memory Test, 1st Ed	Poor; intact recognition & prospective memory		Poor; intact recognition & prospective memory	
Montreal Battery of Evaluation of Amusia				
Pitch composite score (/30)	17.3[e]		22.7	
Rhythm (/30)	21[e]		24	
Metre (/30)	24		21	
Memory (/30)	16[e]		24	

[a]Aphasia Severity Rating Scale: 0 = No usable speech – 5 = Minimal discernible speech difficulty;
[b]Scoring for Frenchay Dysarthria Assessment, 2nd ed: a = normal for age; b = mild abnormality noticeable to a skilled observer; c = abnormality obvious, but can perform task/movements with reasonable approximation; d = some production of task, but poor in quality, unable to sustain, inaccurate, or extremely laboured; e = unable to undertake task/movement/sound.
[c]Impairment Levels for Apraxia Battery for Adults, 2nd ed, subtests: None, Mild, Moderate, Severe.
[d]Composite scores for weekly pre-session and post-session (not pre- and post-MIT);
CIUs: correct information units (see Nicholas & Brookshire, 1993, for this scoring method); DNC: did not complete;
[e]More than 2 standard deviations from norms.

vocabulary rather than actual rehabilitation of propositional language. Overall, Case A and his caregivers felt that the DVDs were of great benefit to him, although this was not truly reflected in the standardised battery used to assess his outcome.

Case B was highly motivated to learn to speak again and to use MIT. He practised faithfully and effortfully with the DVDs over 6 weeks, completing all of the prescribed sessions, but was unable to accurately produce any of the sung phrases at any point in the training (Table 1). Over the 6 weeks of training, he became able to sing a few syllables that he could not produce at baseline, but the overall picture was quite static. Speech and language assessments suggested some possible improvement in apraxia symptoms (from moderate to mild using the scoring system of the ABA, 2nd ed.), but no changes in aphasia symptoms. Despite the lack of speech/language modulation, at the behavioural level the DVD training influenced his mood (Table 1), with significant increases in positive mood after MIT sessions, $t_{(6)} = -2.463$, $p = .049$.

Case B was also unable to produce the sung and spoken phrases used in the functional scanning protocol either before or after training. Although he attempted to speak and sing in the scanner, he was only able to produce stereotyped babble during the spoken condition and humming or sung babble during the singing condition. The conventional SPM imaging analysis showed limited significant activation or deactivation for most conditions and time points (Figure 1(a–d)); however, these analyses appear to be confounded by excess motion in the scanner, particularly in the z-plane, during effortful attempts to speak and sing. Case B produced significantly greater translational movements in the x- and z-planes and for rotations about the x-axis compared to healthy controls performing a similar singing task (Supplementary Data). He also had greater maximum movements for all motion parameters (x, y, and z translations and rotations around those planes), with the exception of rotations around the z-axis, compared to healthy controls (Supplementary Data). His largest motion in the z-plane was more than three times greater than the average maximum z-plane motion in healthy controls, and rotations around the x-axis were also nearly 3 times greater.

A data-driven multivariate technique, spatial ICA, was also used to investigate whether there might be separable task-related patterns of activity that were obscured in conventional analyses by noise (due to factors such as motion or poor model specification related to abnormal haemodynamic response function shapes). The first contrasts calculated on the parameter estimates resulting from a multisession GLM (see Methods) compared pre-MIT and post-MIT, with the alpha-level set to $p = .0005$ after Bonferroni correction (.05/96, given 16 components and six tests on each component). Specifically, we evaluated the following contrasts: sing-trained minus rest pre vs. sing-trained minus rest post, sing-untrained minus rest pre vs. sing-untrained minus rest post, and speak minus rest pre vs. speak minus rest post. None of these comparisons were significant (all p-values > .07). We therefore collapsed across time (average parameter estimates from the pre- and post-MIT sessions) and investigated (1) the average effects across all singing sessions, (2) the average effects across all speaking sessions, and (3) the average effect of singing versus the average effect of speaking.

Three components showed significant task-related activity in either the sing or speak conditions. One of these resembled the venous system (Component 7), so is not considered further. The remaining two components showed activation in the singing, but not speaking, runs (Figure 1(e)). One of these (Component 1) showed singing-related deactivation in left frontal cortex and right cerebellum, accompanied by singing-related activation in right frontal and bilateral parietal regions; activity during the speaking task was not significant, nor was the contrast comparing singing and speaking directly. The

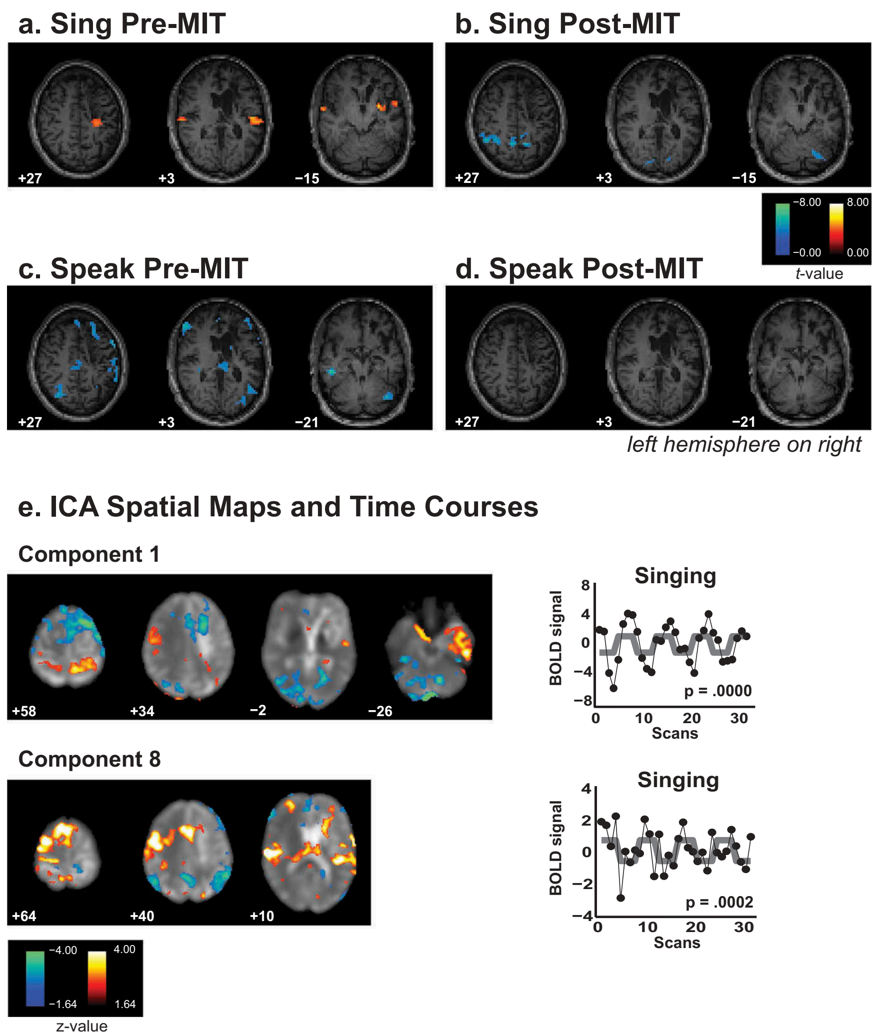

Figure 1. Case B's brain activity for overt attempts to speak and sing short novel phrases (compared to rest) before and after standardised Melodic Intonation Therapy (MIT). (a–d) Conventional SPM analyses. (a) Sing-rest at baseline. (b) Sing-rest after 6 weeks of DVD-based MIT. (c) Speak-rest at baseline. (d) Speak-rest after 6 weeks of DVD-based MIT. $p < .001$, FDRc = .05; de/activation clusters overlaid on a T1-weighted anatomical image. (e) ICA analyses. Spatial maps of the ICA components extracted for the singing and speaking tasks that showed task-modulated activity, with their BOLD (Blood-oxygen-level-dependent) signal time courses (mean-centred data). Spatial maps are overlaid on the mean of the smoothed functional image (average across sessions), thresholded at voxelwise $p < 0.001$. In the time-course plots, the grey solid lines represent the boxcar functions corresponding to cycles of task and rest that were fitted to each time course using multiple regression. The p-values show the significance of the contrasts for singing calculated on the parameter estimates obtained from the multiple regression.

other (Component 8) consisted of activation in bilateral auditory cortex, right frontal cortex, and superior medial frontal cortex, accompanied by deactivation in medial and lateral parietal cortex. This component resembles the task positive and task negative networks (Fox et al., 2005), with the left MCA territory regions of these networks obliterated by the stroke.

Discussion

For both of these cases, standardised testing did not indicate significant changes in propositional language production following 6 weeks of DVD-based MIT. These quantitative behavioural results might imply that this MIT protocol was not an effective therapeutic approach. However, Case A showed striking clinical changes in his attempts to communicate, as well as in mood and attention, following the training. This suggests that some positive effects of the training were not captured by the standardised assessment batteries. Since the stated goal of MIT is the restoration of propositional language, our language assessments were tailored towards the evaluation of propositional language. Case A highlights the relevance of quantifying other possible outcomes, such as the use of trained phrases to enhance functional communication. This raises the important question of what defines "successful" aphasia therapy. Should improving a patient's functional vocabulary be considered a successful MIT outcome, given its impact on the daily functioning of the individual, even if it is not possible to restore propositional language? Improving the functional communication of Case A represented greater progress than had been possible previously, despite extensive speech and language rehabilitation.

In contrast, Case B did not show similar clinical improvements, apart from changes in mood. Again, one might conclude that MIT was not an effective strategy for Case B or that he was not a suitable candidate, given that his presenting profile included severe comorbid apraxia and a lack of language facilitation through singing. However, an alternative explanation is that the inflexibility of this particular approach to MIT was the problem. Following our study, Case B decided to undertake further singing-based language rehabilitation with a music therapist, including MIT with an easier and more flexible protocol, and he regained the ability to sing a number of high-frequency phrases. Our standardised six-week protocol was not optimised for the severity of Case B's aphasia and apraxia, as even the content of the first DVD was beyond his capacity, whereas a less inflexible and more tailored clinical approach was effective for this participant.

Our results suggest that standardisation of therapy within the research context might limit the clinical validity of MIT and may reduce its efficacy and generalisability. The standardisation of assessment also presents challenges for interpreting the efficacy of MIT. In the first case, our standardised assessment battery did not adequately capture the benefits of MIT for the participant, potentially leading to the likely incorrect conclusion that MIT was not an effective treatment. Nickels (2005) argues that using aphasia batteries to measure language rehabilitation success may not be appropriate, given their lack of sensitivity to change and the high degree of variability in the performance of aphasic individuals. However, the alternative hypothesis-driven assessment approach recommended by Nickels is highly individualised and presents distinct challenges for group research. Hypothesis-driven assessment often gives rise to tailored assessments and therefore different outcome measures for each individual, complicating statistical analyses at the group level. In the second case, the standardised protocol was found to be too difficult given the severity of Case B's aphasia and apraxia. In a clinical context, the protocol would have been customised to address this patient's individual rehabilitation needs, whereas this was not possible in our research context. Due to this inflexibility, the selection criteria for participants in larger group studies would need to be narrowed to exclude any participants with more extreme or diverse presentations, thereby limiting the generalisability of any conclusions drawn.

In addition to the standardised therapy protocol and assessments, the imaging protocol piloted with Case B was also standardised, with set phrases to be spoken or sung. Unfortunately, despite the simplistic nature of these phrases, Case B was unable to produce them either before or after training. Many aphasia rehabilitation studies have used overt repetition of high-frequency phrases, overt naming, or automatic speech tasks to assess functional activation patterns for language. These types of tasks represent non-propositional speech, and as such are not particularly suitable for assessing changes in propositional language activation. However, the advantage of these non-propositional tasks is that all but the most severely affected aphasic participants can usually complete them. For tasks that more closely approximate propositional speech, such as orthographic lexical retrieval or the generation of novel, untrained phrases, it is quite likely that some participants will not be able to successfully perform the tasks in scanner. The greater the focus on ensuring that all aphasic participants in a group study can complete the same task in-scanner, the more likely that the overt responses will need to be highly trained, non-propositional language production. Clearly, the design of functional imaging protocols for aphasia rehabilitation research presents some distinct challenges and provides further demonstration of the trade-off that occurs between clinical validity and experimental standardisation.

In the current study, Case B's imaging results are difficult to interpret; they may be due in part to his inability to accurately perform the tasks, especially the speaking task. Unfortunately, his effortful yet unsuccessful attempts to speak and sing in the scanner were accompanied by significant movement, which in turn affected the conventional SPM model fits. Although the sparse temporal sampling procedure used here has been recommended for overt production tasks in aphasia (Crosson et al., 2007), in this case it was still unable to entirely mitigate the confound of motion. In their recent review of motion correction issues in resting state fMRI, Power, Schlaggar, and Petersen (2015) indicate that patients, the elderly and children have a propensity to move more during scanning. Given the deleterious effects this movement has on imaging results, fMRI in neurological populations poses distinct challenges. One of the ways to address these is through alternate data-driven analysis methods, such as ICA as used here. ICA was able to uncover task-modulated activity during singing not detected in the conventional analyses, including deactivation of peri-lesional regions in the left hemisphere and activation of right frontal, superior medial frontal and bilateral auditory regions. This suggests that Case B may have been deactivating disordered left frontal regions and recruiting right hemisphere regions of the language and singing networks (Wilson, Abbott, Lusher, Gentle, & Jackson, 2011) in order to attempt the singing tasks. The fact that Case B showed significant task-related activity for singing attempts but not for speaking might reflect that he was able to "hum", but was not able to speak or sing any words. It also raises the conjecture that singing therapy (in this case, MIT) would provide more potential benefit than a speech-based therapy if it could promote utilisation of this intact portion of the singing network. This idea is perhaps corroborated by the finding that in both of our cases, MIT apparently led to more progress (during our study for Case A and following our study for Case B) than extensive speech-based therapy in the chronic stage.

Given the experimental methods used in our case studies, we would be remiss not to raise the question of whether standardised electronic versions of interventions, such as that used here, could ever be comparable to customisable clinician-administered therapy. Computer-based aphasia rehabilitation is becoming more common and is fairly well-established, with demonstrated efficacy for a number of different treatments (Zheng, Lynch, & Taylor, 2016). To date, however, electronically-delivered music interventions have not been investigated in the literature. Although two small randomised controlled trials of MIT (in subacute and chronic aphasia, respectively) included the use of an iPod app to increase treatment intensity, the majority of the therapy was provided face-to-face by speech-language therapists, and the efficacy of the electronic component was not specifically investigated (van der Meulen et al., 2016; van der Meulen, van de Sandt-Koenderman, Heijenbrok-Kal, Visch-Brink, & Ribbers, 2014). A number of factors motivated the use of electronic media to present the therapy in the current study, including the desire to standardise both the instruction and the practice. The DVDs provided a consistent audiovisual model for the participant to follow and allowed tight control of protocol fidelity and therapy parameters (speed, number of repetitions, etc.). It also improved the feasibility of the research by reducing the necessary hours of face-to-face therapy, and in principal would greatly expand the scope for clinical treatments given the reduced cost of implementing such therapy.

Our DVD-based MIT protocol allowed for excellent standardisation of an intensive therapy protocol within a research context and was well-accepted by our pilot participants, but was not able to respond to the needs of some individual aphasic presentations. Limitations of this particular protocol include the generic nature of the phrases chosen, the omission of backup procedures that allow the patient to have additional opportunities to work on a particular phrase, an invariant difficulty level regardless of impairment severity, and the limited number of phrases for training. For example, the content of phrases trained in MIT is thought to be significant, with suggestions that the material should be relevant and even personalised for each individual with aphasia (Sparks, 2008). With regards to the number of phrases used in training, the original MIT protocol was designed to include a large number of phrases to avoid repetition and practice of the same phrases until overlearned, with recent research supporting this recommendation (Zumbansen et al., 2014b). An additional limitation of any electronic protocol is the reduction or elimination of interactions between clinicians and patients, thereby reducing opportunities for the development of therapeutic relationships or alliances, which may bolster rehabilitation outcomes (Kayes & McPherson, 2012). Interaction with a live therapist could possibly have differential effects on motivation and even on the ease of speech movement imitation, although this has yet to be investigated.

With the exception of potential therapist effects, the limitations of our MIT protocol outlined above are not unique to our study and its use of DVDs; however, overcoming these types of limitations presents distinct challenges for electronic or computer-based therapies. Nonetheless, we would suggest that it is theoretically possible to address these limitations within an electronic framework, although it would likely require greater initial investment in production. Increased storage capacity and processing speed in relatively inexpensive small computers, in addition to innovations in fields such as machine learning and artificial intelligence, have made it feasible to design more full-featured and customisable computer-based approaches. Collaboration with computer scientists and software developers might allow music-based interventions to take advantage of the standardisation and resource-use optimisation provided by electronic protocols, without sacrificing all flexibility and responsiveness. Such tools would likely provide significant advantages within both clinical and research domains. When evaluated from a psychological standpoint, our pilot data suggest that electronic music therapy warrants further investigation. Should further research demonstrate its efficacy, significant accessibility benefits would be realised with this approach.

Conclusions

In light of the tensions between the research and clinical approaches that we have described here and that are illustrated by our two case studies, it is important to consider how these tensions can best be resolved. Greater awareness and acknowledgement of the differences in orientation, aims, and methodology of the two approaches is a good start, so that wherever possible the strengths of both can be harnessed to conduct high-quality clinical research. Here, we present a few suggestions to be considered as the field moves forward, in the hope that these perspectives will spark ongoing discussion about music intervention research in aphasia and clinical research more generally.

First, we suggest that outcome measurement should be carefully considered. Both clinical observations and quantitative data should be valued, and sensitivity and flexibility of assessments across a variety of participants should be the goal. This may mean a focus on the development of new assessment options that measure ecologically valid and patient-centred outcomes with both adequate sensitivity and test-retest reliability. In contrast, the use of existing options that have not been optimised or the inappropriate adaptation of existing clinical assessments to novel research contexts simply because they are available may hamper further progress in the field. For example, many functional outcome measures lack the nuances and psychometric properties necessary to address specific research hypotheses, yet targeting functional outcomes with valid measures of everyday communication skills could provide the type of generic and flexible endpoints required to accommodate heterogeneous clinical presentations in group studies. In addition, if functional imaging tasks are to be used to understand the underlying changes in brain function that accompany clinical changes, the use of innovative task design and analysis methods such as event-related ICA should be considered (Masterton, Jackson, & Abbott, 2013; Tailby, Rayner, Wilson, & Jackson, 2017).

Second, we suggest that the often lamented heterogeneity of clinical samples be exploited. Highly valuable information can be obtained through the systematic exploration of individual differences, rather than the elimination of differences through averaging across groups (Vogel & Awh, 2008). This is an area in which a clinical approach is highly useful. Clinical studies, particularly detailed single-subject experimental research, typically gather more data (in both depth and breadth) from each individual than can be obtained in standard group studies. If the strengths and validity of this detailed approach can be combined with larger samples where funding permits, then the results are likely to provide the most useful evidence base, addressing both efficacy and possible mechanistic explanations. Given the many variants of music intervention protocols currently in use, the ongoing use of case studies and even of controlled single-subject experimental designs without adequate replication will not be helpful in addressing unanswered questions about MIT and other music interventions, such as issues of efficacy, generalisability, therapeutic mechanisms, and protocol best practices. However, hybrid approaches that draw on the strengths of both clinically-oriented single-subject experimental designs (Thompson, 2006) and standardised group designs would allow individual differences to be examined in a meaningful way.

Third, we suggest that in some research contexts, it would be appropriate to place less emphasis on the standardisation of therapy protocols, in order to allow them to resemble actual clinical practice. Magee and Stewart (2015) discuss examples of flexible protocols in recent music therapy studies, indicating that it is certainly feasible. For instance, a randomised controlled trial of improvisational music therapy for children with autism used a well-described but adaptable protocol, with ongoing monitoring of protocol adherence through clinical supervision, documentation, and video recording of treatment sessions (Geretsegger, Holck, & Gold, 2012). Within the speech and language rehabilitation domain, trials that evaluate somewhat personalised interventions are also becoming more common (for example, Bowen et al., 2012). Music interventions for aphasia could adopt similar procedures. With good documentation, some of the individual differences in treatment and response to treatment could be disentangled and would add to, rather than detract from, the validity of the findings. Even more importantly, it would ensure that (a) patients are not disadvantaged in any way when being treated in the context of clinical research and (b) the evidence created is appropriate and reliable in the clinical context.

There are significant challenges to achieving quality group-level research for music-based aphasia interventions. Despite these challenges, the goal certainly remains to carefully assess therapeutic efficacy, optimise protocols, and understand how and why successful therapy works. It should be noted that an exclusive focus on obtaining gold standard research evidence, that is, randomised controlled trials, as the foundation for evidence-based practice may lead us to forget that we should still be using the best available evidence even if randomised controlled trials are not available, and that clinical expertise is also part of the equation. It is clinical expertise that allows for necessary customisation, because "even excellent external evidence may be inapplicable to or inappropriate for an individual patient" (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, *p*.72). We hope that these considerations will assist researchers and clinicians working in aphasia rehabilitation as they pursue evidence-based practice and seek to maximise its potential for patient outcomes.

Acknowledgements

We thank David Abbott for assistance with the neuroimaging design and analysis.

Disclosure statement

The authors report no conflicts of interest

Funding

This work was supported by the Austin Medical Research Foundation and the Operational Infrastructure Support Program of the State Government of Victoria, Australia.

References

Aagaard-Hansen, J. (2007). The challenges of cross-disciplinary research. *Social Epistemology*, *21*, 425–438. doi:10.1080/02691720701746540

Abbott, D. F., Masterton, R. A. J., Waites, A. B., Bhaganagarapu, K., Pell, G. S., Harvey, M. R., … Jackson, G. D. (2011). The iBrain^{TM} analysis toolbox for SPM. In *Proceedings for the 17th annual meeting of the Organization for Human Brain Mapping, Quebec City, Canada,* 364 WTh.

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology*, *29*, 130–131. doi:10.1001/archneur.1973.00490260074018

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, *23*, 137–152. doi:10.1109/TMI.2003.822821

Benson, D. F., Dobkin, B. H., Gonzalez Rothi, L. J., Helm-Estabrooks, N., & Kertesz, A. (1994). Assessment: Melodic intonation therapy. *Neurology*, *44*, 566–568.

Bowen, A., Hesketh, A., Patchick, E., Young, A., Davies, L., Vail, A., … Tyrrell, P. (2012). Effectiveness of enhanced communication therapy in the first four months after stroke for aphasia and dysarthria: A randomised controlled trial. *BMJ: British Medical Journal*, *345*, e4407. doi:10.1136/bmj.e4407

Cockburn, J., Wilson, B., Baddeley, A., & Hiorns, R. (1990). Assessing everyday memory in patients with dysphasia. *The British Journal of Clinical Psychology*, *29*, 353–360. doi:10.1111/bjc.1990.29.issue-4

Crosson, B., McGregor, K., Gopinath, K. S., Conway, T. W., Benjamin, M., Chang, Y.-L., … White, K. D. (2007). Functional MRI of language in aphasia: A review of the literature and the methodological challenges. *Neuropsychology Review*, *17*, 157–177. doi:10.1007/s11065-007-9024-z

Dabul, B. L. (2000). *Apraxia battery for adults: Examiner's manual* (2nd ed.). Austin, TX: Pro-Ed.

Enderby, P., & Palmer, R. (2008). *Frenchay dysarthria assessment: Examiner's manual* (2nd ed.). Austin, TX: Pro-Ed.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 9673–9678. doi:10.1073/pnas.0504136102

Geretsegger, M., Holck, U., & Gold, C. (2012). Randomised controlled trial of improvisational music therapy's effectiveness for children with autism spectrum disorders (TIME-A): Study protocol. *BMC Pediatrics*, *12*, 2. doi:10.1186/1471-2431-12-2

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston diagnostic aphasia examination* (3rd ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.

Helm-Estabrooks, N. (1983). Exploiting the right hemisphere for language rehabilitation: Melodic intonation therapy. In E. Perecman (Ed.), *Cognitive processing in the right hemisphere* (pp. 229–240). New York, NY: Academic Press.

Helm-Estabrooks, N., & Albert, M. L. (2004). Melodic intonation therapy. In *Manual of aphasia and aphasia therapy* (2nd ed., pp. 221–233). Austin, TX: Pro Ed.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*, 782–790. doi:10.1016/j.neuroimage.2011.09.015

Kayes, N. M., & McPherson, K. M. (2012). Human technologies in rehabilitation: 'Who' and 'How' we are with our clients. *Disability and Rehabilitation*, *34*, 1907–1911. doi:10.3109/09638288.2012.670044

Magee, W. L., & Stewart, L. (2015). The challenges and benefits of a genuine partnership between Music Therapy and Neuroscience: A dialog between scientist and therapist. *Frontiers in Human Neuroscience*, *9*, 223. doi:10.3389/fnhum.2015.00223

Masterton, R. A. J., Jackson, G. D., & Abbott, D. F. (2013). Mapping brain activity using event-related independent components analysis (eICA): Specific advantages for EEG-fMRI. *NeuroImage*, *70*, 164–174. doi:10.1016/j.neuroimage.2012.12.025

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *EdITS manual for the profile of mood states (POMS)*. San Diego, CA: Educational and Industrial Testing Service.

Merrett, D. L., Peretz, I., & Wilson, S. J. (2014). Neurobiological, cognitive, and emotional mechanisms in Melodic Intonation Therapy. *Frontiers in Human Neuroscience*, *8*, 401. doi:10.3389/fnhum.2014.00401

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research*, *36*, 338–350. doi:10.1044/jshr.3602.338

Nickels, L. (2005). Tried, tested and trusted? Language assessment for rehabilitation. In P. W. Halligan & D. T. Wade (Eds.), *The effectiveness of rehabilitation for cognitive deficits* (pp. 169–184). Oxford: Oxford University Press.

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy: Shared insights on how it is done and why it might help. *Annals of the New York Academy of Sciences*, *1169*, 431–436. doi:10.1111/j.1749-6632.2009.04859.x

Patterson, J., & Coppens, P. (2017). Integrating principles of evidence-based practice into aphasia rehabilitation. In P. Coppens & J. Patterson (Eds.), *Aphasia rehabilitation: Clinical challenges* (pp. 355–392). Burlington, MA: Jones & Bartlett Learning.

Peck, K. K., Moore, A. B., Crosson, B. A., Gaiefsky, M., Gopinath, K. S., White, K., & Briggs, R. W. (2004). Functional magnetic resonance imaging before and after aphasia therapy: Shifts in hemodynamic time to peak during an overt language task. *Stroke*, *35*, 554–559. doi:10.1161/01.STR.0000110983.50753.9D

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75. doi:10.1196/annals.1284.006

Power, J. D., Schlaggar, B. L., & Petersen, S. E. (2015). Recent progress and outstanding issues in motion correction in resting state fMRI. *NeuroImage*, *105*, 536–551. doi:10.1016/j.neuroimage.2014.10.044

Sackett, D. L., Rosenberg, W. M., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *Bmj*, *312*, 71–72. doi:10.1136/bmj.312.7023.71

Särkämö, T., Tervaniemi, M., Soinila, S., Autti, T., Silvennoinen, H. M., Laine, M., & Hietanen, M. (2009). Cognitive deficits associated with acquired amusia after stroke: A neuropsychological follow-up study. *Neuropsychologia*, *47*, 2642–2651. doi:10.1016/j.neuropsychologia.2009.05.015

Sparks, R. (2008). Melodic intonation therapy. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (5th ed., pp. 837–851). Baltimore, MD: Lippincott Williams & Wilkins.

Sparks, R., Helm, N., & Albert, M. (1974). Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex*, *10*, 303–316. doi:10.1016/S0010-9452(74)80024-9

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary*. New York, NY: Oxford University Press.

Tailby, C., Rayner, G., Wilson, S., & Jackson, G. (2017). The spatiotemporal substrates of autobiographical recollection: Using event-related ICA to study cognitive networks in action. *NeuroImage*, *152*, 237–248. doi:10.1016/j.neuroimage.2017.02.088

Thompson, C. K. (2006). Single subject controlled experiments in aphasia: The science and the state of the science. *Journal of Communication Disorders*, *39*, 266–291. doi:10.1016/j.jcomdis.2006.02.003

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*, 533. doi:10.3389/fnhum.2016.00533

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. doi:10.1177/1545968313517753

Vogel, E. K., & Awh, E. (2008). How to exploit diversity for scientific gain using individual differences to constrain cognitive theory. *Current Directions in Psychological Science*, *17*, 171–176. doi:10.1111/j.1467-8721.2008.00569.x

Wilson, S. J., Abbott, D. F., Lusher, D., Gentle, E. C., & Jackson, G. D. (2011). Finding your voice: A singing lesson from functional imaging. *Human Brain Mapping*, *32*, 2115–2130. doi:10.1002/hbm.v32.12

Wilson, S. J., Parsons, K., & Reutens, D. C. (2006). Preserved singing in aphasia: A case study of the efficacy of Melodic Intonation Therapy. *Music Perception*, *24*, 23–36. doi:10.1525/mp.2006.24.issue-1

Zheng, C., Lynch, L., & Taylor, N. (2016). Effect of computer therapy in aphasia: A systematic review. *Aphasiology*, *30*, 211–244. doi:10.1080/02687038.2014.996521

Zumbansen, A., Peretz, I., & Hébert, S. (2014a). Melodic Intonation Therapy: Back to basics for future research. *Frontiers in Neurology*, *5*, 7. doi:10.3389/fneur.2014.00007

Zumbansen, A., Peretz, I., & Hébert, S. (2014b). The combination of rhythm and pitch can account for the beneficial effect of melodic intonation therapy on connected speech improvements in Broca's aphasia. *Frontiers in Human Neuroscience*, *8*, 592. doi:10.3389/fnhum.2014.00592