2016

Music and multimodal stimulation (M-STIM) : a dynamic approach to increasing expressive and receptive language in severe global aphasia

Diane M. Henning

ABSTRACT

MUSIC AND MULTIMODAL STIMULATION (M-STIM): A DYNAMIC APPROACH TO INCREASING EXPRESSIVE AND RECEPTIVE LANGUAGE IN SEVERE GLOBAL APHASIA

Diane Henning, MA
School of Allied Health and Communicative Disorders
Northern Illinois University, 2016
Thesis Director: Jamie Mayer, Ph.D., CCC-SLP

This treatment study (M-STIM) represented the second phase of development of a modified Melodic Intonation Therapy (MMIT) protocol based on Melodic Intonation Therapy (MIT: a well-known treatment protocol for Broca's aphasia), with the goal of increasing expressive and receptive language function in an individual with severe global aphasia.

In a previous study (MMIT Phase), it was shown that trained, functional phrases set to unique melodies (MMIT) were learned at a faster rate and were more durable over time compared to phrases taught using traditional MIT. However, generalization of both types of stimuli (MIT and MMIT) to everyday, functional use was limited; the authors hypothesized that this was due to persistent auditory comprehension deficits.

Therefore, the current study (M-STIM) utilized previously trained MMIT stimuli with the following modifications: 1) increased melodic uniqueness, 2) increased salience, and 3) multimodal stimulation (collectively referred to as M-STIM: i.e., music and multimodal stimulation). We found that targeting both receptive and expressive language in tandem (i.e., M-STIM), resulted in increased expressive and receptive language subtest scores on the Western Aphasia Battery, improved within-treatment phrase production, and application of trained phrases to everyday life situations, compared to targeting expressive language alone (MMIT).

NORTHERN ILLINOIS UNIVERSITY

DE KALB, ILLINOIS

MAY 2016

MUSIC AND MULTIMODAL STIMULATION (M-STIM): A DYNAMIC APPROACH TO INCREASING EXPRESSIVE AND RECEPTIVE LANGUAGE IN SEVERE GLOBAL APHASIA

BY

DIANE M HENNING

A THESIS SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF ARTS

SCHOOL OF ALLIED HEALTH AND COMMUNICATIVE DISORDERS

Thesis Director:

Jamie F. Mayer, Ph.D., CCC-SLP

ACKNOWLEDGEMENTS

I want to thank all of the individuals who have made this thesis possible, especially my mentor, Dr. Jamie Mayer, who introduced me to the world of aphasia research and helped to spark my interest in applying music to rehabilitation of neurogenic disorders. If it were not for her enthusiasm, encouragement, and support at every step of this process, I would have never had the opportunity to complete this project. I also want to thank our participant, KC, for teaching me far more about life and communication than I ever could have taught her. I also thank her family for their support and giving us the opportunity to work with KC. Additionally, I want to thank my committee for their feedback and understanding throughout this process, the undergraduate students who helped with data collection, and Northern Illinois University for its support of student research. Lastly, I want to thank my family and loved ones for their unconditional love and support.

DEDICATION

I dedicate this work to all of the individuals in the world who have difficulty verbally communicating their most basic wants, needs, and thoughts, no matter the reason, and hope that this thesis in some small way adds to the breadth of research that will eventually lead to breaking their communication barriers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1. INTRODUCTION

Melodic Intonation Therapy (MIT) is a well-known treatment protocol designed to increase expressive language skills for individuals with Broca's aphasia - i.e., moderate to severe, nonfluent expressive language deficits in the presence of relatively intact comprehension (Albert, Sparks, & Helm, 1973; Helm-Estabrooks & Albert, 2004). Unfortunately, many individuals with left-hemisphere lesions large enough to cause moderate to severe expressive language deficits experience concomitant receptive language deficits as well (e.g., mixed or global aphasia). In fact, very few options exist for this population that have been shown to increase expressive and receptive language.

A possible solution for these individuals is to modify traditional MIT, which relies on a two-note, intoned way of speaking to increase verbal expression, in such a way as to focus on singing over intonation (e.g., Baker, 2000; Hartley et al., 2010). While these modifications have been shown to increase access to verbal language in patients with severe nonfluent aphasia, presumably through stimulating the right hemisphere (e.g., Schlaug, Marchina, & Norton, 2008; Schlaug et al., 2009), they do not address the receptive language deficits found in a large portion of this patient population.

Baker (2000) described a modified version of a MIT protocol, MMIT, wherein the two-pitch MIT intonation pattern was modified such that more musical and melodically unique phrases were taught to two individuals with severe nonfluent aphasia after traditional MIT therapy had proven unsuccessful in facilitating increases in verbal expression. The increasing musicality and individuality of stimuli were rationalized to facilitate the encoding/retrieval of these target phrases. Notably, the individuals in this study, who were unable to generate any words independently following traditional MIT, acquired and independently generated words and phrases following this modified version. Baker's success with MMIT is of particular interest in our current study because the second of her two participants presented with "poor ability to comprehend verbal language, requiring several cues such as pictures, written words and gesture to comprehend a simple message" (p. 113). The improvement in verbal expression in the second participant demonstrates the potential for MMIT to meet the needs of patients for whom traditional MIT is ineffective due to deficits in auditory comprehension. Because this study used a case report design, however; the lack of empirical data render interpretation and generalization difficult at best.

Thus, use of music/melodies in aphasia treatment appears promising but hard empirical data are lacking. For example, in a systematic review of music therapy techniques for use with patients with neurogenic language disorders, namely aphasia, research supporting music's effectiveness in therapy was lacking due to poor treatment designs that omitted treatment fidelity, validity and reliability, significance, pre- and post-testing, and the therapy's functional impact on participants, along with other important factors necessary to score higher methodological ratings upon review (Hurkmans et al., 2012).

Given the promising reports from several case studies, however, we proposed that modified MIT (MMIT), focusing on singing to elicit language from individuals with severe aphasia, should be explored as a therapeutic protocol in its own right. Therefore, we followed the Five-Phase Outcomes Research Model developed by Robey (2004) to determine the efficacy, effectiveness, and efficiency of MMIT for global aphasia (see Table 1). Phase 1 involved hypothesis development and defining the therapeutic effect (e.g., beneficial physiological changes at the impairment, activities, and participation level).

Table 1

Robey's Five-Phase Model of Clinical Outcome Research

Phase 1	Hypothesis development and selection of therapeutic effect
Phase 2	Exploration of dimensions of therapeutic effect and make necessary preparation for conduction of a clinical trial
Phase 3	Conduction of clinical trial to assess efficacy
Phase 4	Field research to test effectiveness in the target population, in specific sub-populations, under variations of service-delivery models, variations of the treatment protocol, and conduction of meta-analyses of efficacy studies
Phase 5	Determination of who benefits from the treatment protocol and at what cost; assessment of costs and values through cost-values effectiveness studies

Toward that end, Mayer, Jeppson, and Henning (2014), utilized a single-subject, alternating-treatment paradigm to compare empirically a version of MMIT (modified from Baker, 2000) to the traditional MIT protocol, for increasing expressive language in an individual with global aphasia. The current study built directly on Mayer et al. (2014) and consisted of a continuation of Phase 1 research following Robey's model, i.e. refining the hypotheses and therapeutic effect. At this point, our goals included identifying if a therapeutic effect was present and estimating its magnitude in order to “make point and interval estimates of effect size, make a first approximation to population definitions, make a first approximation of the treatment protocol, estimate appropriate dose, specify the therapeutic effect and how it is to be indexed, [and] generate and refine hypotheses" (p. 404).

Results of Mayer et al. (2014) showed superiority of MMIT (i.e., treatment stimuli set to music) over traditional, two-toned MIT in terms of rate of learning (i.e., phrases taught using MMIT were learned more quickly than those taught using traditional MIT). The MMIT Phase also revealed increased durability over time of MMIT compared to MIT stimuli, in that the former were produced more accurately and with less cueing in a two-month follow-up session. However, the study participant demonstrated very limited application of both MIT and MMIT phrases to functional communication in activities of daily living; Mayer et al. hypothesized that this lack of generalization was secondary to persistent auditory comprehension deficits. Therefore, the current study (M-STIM Phase) refined the hypotheses from Mayer et al.'s (2014) study by utilizing MMIT along with three additional strategies to maximize receptive language skills: 1) increased melodic uniqueness, 2) increased salience, and 3) multimodal stimulation: collectively referred to as M-STIM, i.e., music and multimodal stimulation. These strategies were applied to the treatment of the same participant in Mayer et al.'s study, with the hope that these modifications would lead to increased receptive language skills, at least for trained stimuli, and thus to the ability to apply trained phrases to a variety of cued and uncued contexts, maximizing stimulus generalization and ecological validity.

Research Questions and Predictions

The purpose of the current study (M-STIM) was to increase the treatment efficacy of Mayer et al.'s protocol for global aphasia by focusing on receptive language skills, in addition to the verbal expression skills that were the focus of MMIT, utilizing the three strategies (melodic uniqueness, salience, and multi-modal stimulation) outlined above. Toward this end, the following research questions and predictions were developed:

1. Does utilization of M-STIM lead to improved standardized language scores?

    Prediction 1: KC's standardized auditory comprehension score will improve from the pre-to-post-M-STIM Phase.

    Prediction 2: KC will continue to improve standardized expressive language scores from the pre-to-post-M-STIM Phase.

2. Does one particular strategy of M-STIM (increased melodic uniqueness, increased salience to everyday life, or increased multimodal stimulation) demonstrate significantly greater improvements in KC's performance of trained functional phrases during within-treatment probes compared to the others?

    Prediction 3: KC will demonstrate increases in percent correct of trained-phrase responses containing increased melodic uniqueness, increased salience to everyday life, and/or added multimodal stimulation during within-treatment probes from the pre-to-post-M-STIM Phase, as measured by an M-STIM scoring system.

    a. KC will demonstrate a positive slope of percent correct for responses to trained phrases during within-treatment probes from pre-to-post-M-STIM Phase.

b. A significant effect size will be measured in percent correct of trained phrases during within-treatment probes from the pre-to-post-M-STIM Phase.

3. Are there any differences in KC's performance outcome measurements when measured by the traditional MIT scoring system versus the M-STIM scoring system (see p. 43)?

   Prediction 4: The M-STIM scoring system will more accurately capture clinically significant progress over time compared to the traditional MIT scoring system.

4. Does KC demonstrate any functional changes in the use of trained phrases outside of therapy as a result of M-STIM?

   Prediction 5: Post-M-STIM, KC will demonstrate functional positive changes in appropriateness and frequency of use of trained phrases outside of therapy, as measured by two caregiver questionnaires.

2. REVIEW OF LITERATURE

Chronic Global Aphasia

Global aphasia is an acquired, neurogenic, multimodal language disorder resulting in deficits in both expressive and receptive language abilities. This is typically the result of a stroke occurring at the level of the middle cerebral artery stem, resulting in a large infarct in the left lateral hemisphere with consequent damage to Broca's and Wernicke's areas, along with damage to the arcuate fasciculus and possibly subcortical structures as well (Alexander & Loverso, 1992; Manasco, 2014). Due to the severity of verbal expression deficits, treatment for global aphasia usually incorporates communication through nonspeech means, such as the use of gestures (e.g., Visual Action Therapy; Helm-Estabrooks, Fitzpatrick, & Barresi,1982; Gestural Communication; Cubelli, Trentini, & Montagna, 1991). The use of an alternative augmentative communication device or system to provide a mode of communication to augment verbal expression is also a common practice (Dietz, et al., 2013; Hough & Johnson, 2009; Johannsen-Horbach, Cegla, Mager, Schempp, & Wallesch, 1985; Lasker & Garrett, 2006).

Unfortunately, due to their severity, functional expressive and receptive language recovery in people with global aphasia can be limited, especially after the first year following stroke (Alexander & Loverso, 1992; Huber et al., 2002; Poeck, Huber, & Willmes, 1989; Weniger, 2003). After this time, persons with aphasia are considered to be in the "chronic stage," in which less spontaneous recovery may occur. Despite this conventional wisdom, however, many researchers have demonstrated positive intervention outcomes for those individuals in the chronic stage of recovery receiving treatment (Baker, 2000; Jungblut, 2009; Jungblut, Huber, Mais, & Schnitker , 2014; Ren da Fontoura, de Carvalho Rodrigues, Brandão, Monção, & Fumagalli de Salles, 2014; Yamaguchi, Akanuma, Hatayama, Otera, & Meguro, 2012). Given these data, Ren da Fontoura et al. (2014) hypothesized that in the chronic stage of recovery, intact neural tissues are recruited; thus, it is possible, for example, that the role of music in therapy is appropriate in this stage, as involvement of the intact right hemisphere is sought.

Research Supporting Music-Based Therapies for Individuals with Acquired Brain Injury

Hurkmans et al. (2012) presented a systematic review containing a widespread search on all articles written before 2010 pertaining to music in treating acquired language disorders. Search results initially found 1250 articles but only 50 articles contained a therapy study and were considered relevant to the review. Of those 50 articles, only 18 were effect-controlled with measurements before and after intervention, and only 15 met all inclusion criteria.

Reviewers used criteria based on the American Speech-Language-Hearing Association's (ASHA) levels of evidence scheme (2001), which lists nine indicators: study design, blinding, sampling, group/participant comparability, treatment fidelity, outcomes, significance, precision, and intention-to-treat (controlled trials only); with quality indicators for each. Reviewers noted that all but one study (Baker, 2000) contained participants whose speech impairment was caused by stroke and participants across all studies were in the chronic phase of recovery.

Nine out of fifteen treatment studies evaluated the effectiveness of MIT (which is described in detail in following sections); however, in most studies, participants attended therapy less frequently than prescribed in Albert et al., 1973. Methodological quality of the studies was rated with a scale from 0-9. Using that scale, five studies received a score of zero, two studies received a four (the highest score given), and the others fell in between that range. Nine studies followed a case series format and none of the studies involved a randomized controlled trial. Eight studies did not report any information on validity, reliability, or provide *p*-values, and only five studies used comprehensive language tests as outcome measures. Of note, the authors stated that "no distinction has been reported in related (speech parameters) and unrelated measures (non-speech parameters like reading, writing and auditory comprehension)," (p. 7); therefore, it is important that future research determines if any relationships exist between these parameters.

Hurkmans et al. (2012) discussed that overall, methodological quality was rated as low. They also mentioned that in one-third of studies, researchers did not report participants' level of education or cognitive functioning, which Hurkmans et al. suggested is a strong indicator of treatment success, as cognitive impairments have the potential to limit recovery. A lack of reporting was also found in relation to participants' musical background, right-or-left-handed dominance, and generalization of targets to everyday life.

The researchers also suggested that future research use standardized methods and parameters relevant to rehabilitation interventions, for example, utilizing Wade's (2005) model based on the World Health Organization's International Classification model or the American Speech-Language Hearing Association's levels-of-evidence scheme. In sum, no conclusions can yet be drawn "with regard to the effect of the use of musical elements in the treatment of individuals with acquired neurological disorders," (p. 16) due to the low methodological quality of the studies.

Since Hurkmans et al.'s review, a number of research studies have been conducted that do contain higher levels of methodological validity. These (Al-Janabi et al., 2014; Conklyn, Novak, Boissy, Bethoux, & Chemali, 2012; Hurkmans et al., 2014; Jungblut et al., 2014; Ren da Fontoura et al., 2014; van der Meulen, van de Sandt-Koenderman, Heijenbrok-Kal, Visch-Brink, & Ribbers, 2014) will be discussed throughout the following sections.

Melodic Intonation Therapy

Melodic Intonation Therapy (MIT) was developed by Albert, Sparks, and Helm (1973) to increase verbal expression in persons with nonfluent Broca's aphasia. Albert et al. developed the protocol using a series of phrases with each syllable in the phrase paired to a tone a minor third apart from the other (Figure 1). The protocol consists of training functional phrases of about four syllables each with the use of a specific cueing hierarchy, as described briefly below.

Each level of the hierarchy increases in difficulty with increasingly less cueing from the clinician, with the ultimate goal of expanding verbal expression and utterance length while developing speech-like prosody or "sprechgesang" (Sparks, 2008, p.707). Sparks (2008) identified eight principles when using MIT, which include: indirectly correcting the patient by simply reverting to an earlier step in the hierarchy when unsuccessful, repeating functional phrases with a high level of intensity and frequency (Sparks recommends twice-daily sessions five times per week), and utilizing increased pauses in specific steps of the hierarchy to promote generalization.

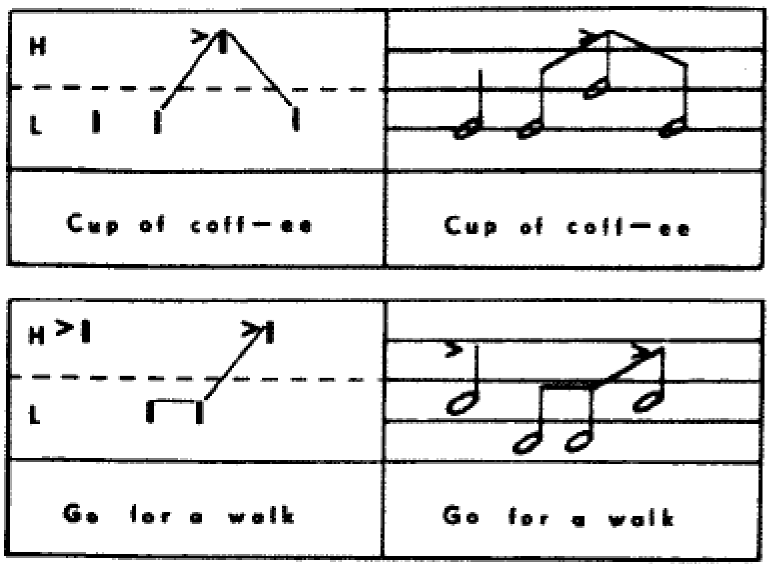


Figure 31–4. Transposition of spoken prosody models to melodic intonation. Key of C in treble cleff is used for illustration. No attempt is made to present accurate musical tempo.

Figure 1. Example of the two-toned intonation pattern of MIT. Extracted from Sparks (2008, p. 707)

This well-known protocol has proven effective at increasing verbal expression with a very specific population of stroke survivors, namely those with nonfluent Broca's aphasia (Albert, Sparks, & Helm, 1973; Norton, Zipse, Marchina, & Schlaug, 2009; Schlaug, Marchina, and Norton, 2009; Sparks, Helm, & Albert, 1974; Sparks & Holland, 1975). In a review of the effectiveness of MIT, Benson et al. (1994) rated MIT as having "Class III evidence," meaning that while more research is needed, current findings suggest it is appropriate for the specified population. The use of Melodic Intonation Therapy (MIT) for persons with global aphasia has not been recommended due to auditory comprehension deficits impacting performance outcomes (Albert, Sparks, & Helm, 1973; Helm-Estabrooks, Albert, & Nicholas, 2014; Sparks, 2008).

Recently, a randomized controlled trial was conducted by van der Meulen, van de Sandt-Koenderman, Heijenbrok-Kal, Visch-Brink, and Ribbers (2014) with 27 right-handed participants with severe, nonfluent aphasia who were in the subacute phase of recovery following a left-hemisphere stroke (2-3 months post-stroke). Importantly, all participants demonstrated relatively intact auditory comprehension. Approximately half of the participants received MIT therapy for six weeks, while the other half received a control therapy that did not include any language production training (e.g., gestural therapy, writing, or auditory comprehension training). Then, after six weeks, the MIT group received the control therapy and the control group received MIT.

MIT was administered over the course of six weeks with treatment sessions occurring five hours per week. The authors reported that clinicians administering treatment followed Sparks' (2008) and Helm-Estabrooks, Nicholas, and Morgan's (1989) MIT protocol. An iPad application, consisting of short video clips (i.e., a mouth singing the target utterances), was developed to support the therapy protocol, with participants expected to view/imitate as homework. The Aachener Aphasie Test (AAT; Graetz, de Blesser, & Willmes, 1991) was given at three points during the study: before treatment, after six weeks of treatment (MIT or control), and six weeks later (MIT or control). The Amsterdam Nijmegen Everyday Language Test (ANELT, Blomert, Koster, & Kean, 1995) and the Sabadel Test (van Eeckhout, 1982), which contains stories to engage participants in conversation, were two assessments given that both examine carryover to functional communication. The authors concluded that the experimental group (those that received MIT first) showed significant improvements on all tasks except on the Sabadel (ANELT, naming on the AAT, repetition on the AAT, and MIT repetition of trained and untrained items), while the control group only showed significant improvements on MIT repetition of untrained phrases. In sum, these results suggested some generalization of trained phrases into functional communication in the experimental group and that more significant increases in expressive language were shown when therapy occurred in the first six-weeks (experimental) as opposed to the second six-weeks (control). The authors point out, however, that a larger sample size would provide more significant effects of the outcome measures, and thus further research in this area should be pursued.

Neurological Basis for Effectiveness of MIT

A seminal study in aphasia literature (Belin et al., 1996) conducted brain imaging (Positron emission tomography, PET) in an attempt to determine activation patterns of linguistic and non-linguistic functions in individuals following stroke. Belin et al. (1996) selected seven right-handed individuals, all who had a unilateral left-hemisphere (LH) middle cerebral artery (MCA) stroke accompanied by severe, nonfluent aphasia (two individuals with Broca's and five with global aphasia), and who had, prior to the study, demonstrated little spontaneous recovery. These participants, however, demonstrated significant improvements in expressive language following a French version of MIT, called Thérapie Mélodique et Rythmique (TMR). Belin et al. selected two types of verbal tasks, one loaded with MIT phrases and one control task (spoken repetition of untrained phrases), to uncover possible differences in brain activation patterns during the PET study. The French version of the Boston Diagnostic Aphasia Examination was used to measure increases in expressive language abilities from pre- to post-TMR, and all study participants showed significant improvements in both expressive and receptive language following intervention.

During Belin et al.'s (1996) PET study, activation tasks included four items: rest, hearing, repetition of typical conversational prosody, and repetition mirroring the two-toned intonation pattern used in MIT. Nine brain regions were selected as areas of focus to search for brain activity during the above-mentioned tasks: Broca's area (and RH homologue), Wernicke's area (and RH homologue), and bilateral Heschl's, anterior superior temporal, and middle temporal gyri; as well as bilateral temporal poles, sensorimotor mouth region, parietal, and prefrontal regions. Cerebral blood flow (CBF) measurements were used to detect levels of brain region activation; the authors found that CBF decreased in the right hemisphere in seven out of nine participants during repetition with MIT as compared to repetition without MIT, whereas CBF significantly increased in Broca's area. Belin et al. also reported two major findings: 1) abnormal activation of the RH occurred during listening tasks and repetition without MIT, and 2) reactivation of Broca's area and left prefrontal cortex occurred during repetition with MIT. The authors hypothesized that abnormal RH activation may account for the persistence of aphasia and may be maladaptive to recovery. The authors acknowledged that these results may appear contrary to previous hypotheses that report the function of MIT is to utilize and strengthen RH structures due to the musical components of MIT; however, they pointed out that MIT is better described as "exaggerating speech prosody" than singing and therefore may better tap into LH language centers, just as a mother exaggerates her speech for her infant (Belin et al., 1996, p. 1510).

Belin et al.'s (1996) findings stand apart from more current research utilizing brain-imaging technology. Schlaug, Marchina, and Norton (2009) hypothesized that intensive MIT therapy applied to patients with chronic nonfluent aphasia would result in structural brain changes. A group study was conducted for six right-handed individuals, all of whom were over a year post-left-hemisphere (LH) stroke had demonstrated moderate to severe nonfluent aphasia in the context of relatively preserved comprehension. Schlaug et al. (2009) utilized high-resolution MRI studies, both before and after 75 sessions of MIT, to assess for neural changes. The arcuate fasciculus (AF) of the right hemisphere (RH) was identified in all six subjects (the AF in the LH could not be identified in any of the subjects, as the majority of the tract had been destroyed post-stroke), and comparison of pre-versus post-treatment measures demonstrated a significant increase in the number of fibers in the RH AF post-treatment. Behavioral measured dovetailed these data, with all participants showing a significant increase in correct information units (CIUs) post-treatment. Schlaug et al. theorized the two elements of MIT that were the strongest contributors to the increase in AF fibers were singing and tapping with the left hand, due to the fact that musical elements have been shown to require more global, as opposed to local, processing and, in turn, activate more right-hemisphere regions than exhibited in speech alone.

Breier, Randle, Maher, and Papanicolaou (2010) used magnetoencephalography (MEG) during covert naming tasks to discover which areas of the brain were utilized during this language-specific task. Both individuals in this study completed two blocks of MIT. One block of therapy lasted three weeks, with two days a week of therapy and two 30-minute therapy sessions in one day, totaling six hours of MIT per block. A three-week break occurred between each block of MIT, and MEG was used to measure brain activity at three time points: pre-MIT, post-block 1, and post-block 2. Both individuals were right-handed males who were two-to-five years post-LH MCA stroke. The Western Aphasia Battery (WAB) was used to assess pre-treatment aphasia severity, with both participants scoring in the moderate-to-severe range. Both individuals demonstrated deficits in expressive and receptive language, but Patient 2 demonstrated more severe auditory comprehension deficits than Patient 1. CIUs were utilized as the primary outcome measure of expressive language change for both participants.

After block 1 of MIT, Patient 1 demonstrated over a 35% increase in CIUs and maintained that increase during the break and after block 2 of MIT. Patient 2 showed no increases in CIUs after block 1 or block 2 of MIT. The MEG studies revealed gradually increased LH activation and decreased RH activation during the covert naming task in Patient 1 after block 1 and 2 of MIT. In Patient 2, LH activation increased after both blocks of MIT, but RH activation also increased with an even greater number of late, language-specific dipoles found during scanning; i.e., that both LH and RH centers were activated during language-specific tasks.

This study concluded with mixed results due to the differing nature of outcomes between Patient 1 and 2. The researchers found that Patient 1, who showed a positive response to MIT through significantly increased CIUs, also increased activation of LH areas that remained intact post-stroke and decreased activation of RH areas during language-specific tasks, suggesting that MIT is associated with increased LH activation and decreased RH activation.

Patient 2, however, showed increases in both LH and RH activation during language tasks, but did not show an increase in the number of CIUs produced. Breier et al. (2010) concluded that Patient 2's lack of positive outcomes from MIT could be due to the severity of auditory comprehension deficits, along with the individual's history of chronic occlusion of the LH MCA and bilateral stenosis of the external carotid artery, suggesting that blood flow to language centers in the LH was very limited. Therefore, it could be implied that attempting to reactivate language centers in the LH would also produce limited positive outcomes.

Breier et al.'s (2010) study differed from Schlaug et al.'s (2009) study in that Patient 2 (of Brier et al.'s study) demonstrated increased RH cortical activation during language activities but did not increase CIUs as a result of intervention, while Schlaug et al.'s participants demonstrated both increases in the number of fibers in the RH AF and increases in CIUs. However, it is difficult to compare directly the results of these two studies due to the differences in methodology used to detect activation of structural changes in the brain.

Despite these differences, one can infer that Breier et al.'s (2010) study better demonstrated the activity in the brain during a linguistic task, while Schlaug et al.'s (2009) study operated under the assumption that an increase in AF fibers would directly lead to increased expressive language; that is, without functional imagery, it is merely speculation that the two variables are so closely related. Furthermore, while Belin et al. (1996) provided evidence that contradicts RH involvement in expressive output using MIT, both Schlaug et al. and Breier et al. noted a positive relationship between expressive language abilities and RH cortical activity in some of the participants.

Variations of Melodic Intonation Therapy and Other Music-Based Interventions

A number of studies have reported results of modifications to MIT, as well as results of music-based interventions for aphasia. Six of these studies were reviewed below.

(1) Baker (2000) reported two cases involving a 32-year-old female and a 30-year-old male, both of whom suffered a traumatic brain injury, which resulted in severe non-fluent aphasia. To aid in recovery of their expressive language abilities, she designed a modified MIT protocol (MMIT) in which she created unique melodies for each trained phrase, all of which were composed around a central harmonic structure. Both participants had received MIT previously with little improvement, but Baker realized the participants' relative strengths in singing versus speaking, so she developed this modified protocol with the goal that these unique melodies would be internalized and used as a mnemonic device to recall and verbalize the target word of each phrase. The concept of only recalling a target word differs from MIT because in MIT it is expected that the participant repeat the entire phrase. Baker explained that due to the severity of nonfluent aphasia, a one-word target is a more appropriate and expected response for these individuals. With each participant, she first sang popular songs with the participant to provide "immediate positive feedback that verbal output is possible, albeit not functional," (p.111). The second step involved the clinician playing and singing the training phrases several

times as the participant listened. Baker noted a tape recording was also created with all of the training phrases for the participant to listen to in between therapy sessions.

Participant 1, Tara, initially received five half-hour sessions per week; therapy was then increased to four hours per week due to initial improvements in expressive language. Her therapy continued for four years, with initial treatment involving the training of ten phrases, and then gradually increasing the number of targets after she was able to successfully generate the ten target words for each phrase upon cueing. Upon discharge from therapy, her verbal expression consisted of 148 words/phrases. Tara was able to produce 124 of these words/phrases independently but she required musical cues for the remainder.

Participant 2, Jeff, received three-to-four sessions of MMIT per week. His therapy protocol began with six phrases, all containing family member's names. Jeff received a much shorter (albeit more realistic) length of therapy, lasting about four months. A tape was created for him to practice the phrases outside of therapy, and upon discharge, he was able to independently generate 30 words. Whereas both participants made significant gains in their ability to functionally communicate, the case series format, along with lack of empirical evidence, make interpretation and generalization of the results difficult at best.

(2) Ren da Fontoura et al. (2014) conducted a study is to determine the therapeutic efficacy of using music to stimulate language rehabilitation in a single participant with Broca's nonfluent aphasia five years post-stroke. Ren da Fontoura et al. noted the participant's comprehension was fairly intact and verbal expression consisted of a variety of dysfluencies, including dyspraxia, anomias, phonological paraphasias, and agrammatism. The modification of

the MIT protocol involved the use of Brazilian popular music to promote increased verbal expression. Psycholinguistic measures, including the Boston Aphasia Diagnostic Exam, Functional Assessment of Communication Skills (ASHA FACS, (Frattali, Thompson, Holland, Wohl, & Ferketic, 1995), and NEUPSILIN-Af (Brief Neuropsychological Assessment Instrument for patients with expressive aphasia, Fonseca, de Salles, & Parente, 2008), were conducted before and after therapy to note any possible increases in overall language function as a result of the intervention.

Treatment frequency consisted of 45-minute twice-weekly sessions for a duration of three months (24 total sessions). During therapy, lyrics of popular, slow-tempo Brazilian songs, containing simple words and sentences that could be carried over into daily use, were trained using the traditional MIT hierarchy. Initially, each song was trained by breaking it apart into smaller phrases and words, accompanied by pictures and the text of the lyrics. Then, utterance length of the phrases was gradually increased as the participant became more proficient. One song was trained at a time, and when the participant could recite the lyrics with prosody typically found in speech, the criterion for beginning a new song was met.

The following outcome measures were used to measure progress: 1) Boston Diagnostic Aphasia Examination-Short Form, 2) Token Test (a measure of auditory comprehension), 3) speech analysis (speech rate in number of correct words per minute), 4) NEUPSILIN-Af, and 5) American Speech-Language-Hearing Association Functional Assessment of Communication (ASHA-FACS). Psycholinguistic measures post-treatment from the Boston Diagnostic Aphasia Examination showed increases in the percentage of correct answers in tasks of oral language understanding, naming of specific categories, word repetition, verbalization of automatic sequences, reading, and written designation. Results of the Token Test revealed an increase in correct answers from 72% to 78%. The speech analysis revealed an increased number of correct words per minute during conversational speech, with the largest increase in speech rate occurring between sessions eight and sixteen, with the rate leveling off after the sixteenth session. Results from the post-treatment administration of the NEUPSILIN-Af revealed improvement in attention, working memory, semantic and episodic verbal memory, prospective memory, naming technique, reading aloud, and spontaneous and dictated writing. On the ASHA-FACS, the participant showed increases in social communication. Overall, results revealed that gains were made not only in expressive language but in executive functioning, especially attention and working memory, and the therapy task of singing popular Brazilian songs kept the participant motivated in therapy.

(3) Conklyn, Novak, Boissy, Bethoux, and Chemali (2012) conducted a pilot study of a modified MIT protocol with a randomized controlled single-blind design. Thirty 30 acute stroke survivors participated, and therapists created individual treatment plans for each participant with novel melodic phrases that closely matched the prosody in pitch and rhythm of spoken phrases. Results demonstrated significant within-subject change, as well as a significant difference between groups, with the greatest improvement shown in those participants receiving the modified MIT treatment. Improvement was measured using a pre- and post-test modeled after the Western Aphasia Battery’s responsive and repetition subsections. The use of novel melodic phrases in this study is comparable to Baker’s (2000) study. However, it is important to note that these participants were in the acute stage of recovery, unlike many of the participants in other studies previously mentioned who were in the chronic stage. Therefore, it is important to

consider the possible role of spontaneous recovery in the outcomes of this intervention, especially given the much shorter dosage of therapy (two-to-five sessions) compared to other studies (e.g., Baker, 2000; Jungblut, 2009; Ren de Fontoura et al., 2014).

(4) Jungblut and colleagues conducted several studies about a music-based intervention called Singen Intonation Prosodie Atmung Rhytmusübungen Improvationen (SIPARI), registered trademark; (Jungblut, 2009; Jungblut & Aldridge, 2004; Jungblut, Huber, Mais, & Schnitker, 2014; Jungblut, Suchanek, & Gerhard , 2009). SIPARI was founded on the tenants of MIT and was created for use with patients with nonfluent Broca's or global aphasia. SIPARI's use with persons with global aphasia is notable, as it appears to provide an effective treatment strategy for them, as opposed to MIT, which has not proven effective for that population. SIPARI, like MIT, combines rhythm, singing, prosody, and intonation as tools to increase verbal expression but is also focuses on breathing exercises and improvisation.

Jungblut et al. (2009) presented a single-subject case study of a 61-year-old male, Mr. Benz (pseudonym), with chronic global aphasia, for whom weekly individual therapy sessions with SIPARI training began three years post-stroke. Jungblut et al. reported that at the onset of therapy, Mr. Benz had extremely limited spontaneous speech that consisted of a few single words and stereotypical utterances. The authors also noted that Mr. Benz's ability to attend and concentrate was severely impaired initially but improved over the course of therapy. A key component of SIPARI that differs from both MIT and MMIT is the use of "musical communication in joint improvisation" (p. 64). Improvisation in this case is the spontaneous creation of speech set to novel melodic phrases, meaning that the clinician and Mr. Benz communicated in a dialogue with spontaneously created phrases set to melody. In eight months, significant improvements were shown in Mr. Benz's language abilities, as well as attention and concentration.

After a year, group therapy sessions were introduced in which Mr. Benz was encouraged to communicate with other individuals with a similar type and severity of aphasia. The Aachener Aphasie Test (AAT; Graetz, de Blesser, & Willmes, 1991) was conducted at four points during the study with results revealing increases in verbal expression (most significantly in repetition and naming). Auditory comprehension increased slightly but did not improve as significantly as verbal expression. Clinically significant improvements in spontaneous speech that carried over into functional communication were discovered and reported at three intervals during the study using the AAT's six rating scales: communicative verbal behavior, articulation and prosody, automatized language, semantic structure, phonemic structure, syntactic structure. Jungblut et al. (2009) also observed that Mr. Benz demonstrated increased participation in social conversation and began talking on the telephone. The carryover into functional communication is significant, as it has not been demonstrated elsewhere. The authors attributed this increase in spontaneous speech not only to the melodic speech elements but also to temporal-rhythmic components.

Jungblut et al. (2014) further explored the efficacy of rhythm in the recovery of spontaneous speech in patients with chronic nonfluent aphasia and apraxia of speech. To begin, a pre-study was conducted with 30 healthy individuals who underwent functional imaging while chanting vowel changes in rhythmic sequences. Data from this study showed correlations to rhythm and temporal processing and sequencing. Rhythm was also found to cause activation of language-specific areas of the brain and was found to factor into lateralization of activation. This discovered activation caused the authors to hypothesize that rhythm plays a large role in the success of the SIPARI voice-training therapy for patients with chronic nonfluent aphasia. Three patients with both apraxia of speech and chronic nonfluent aphasia (two global and one Broca's) participated in the study and underwent the same fMRI procedure as the healthy participants.

Several assessments were administered, including the AAT (Graetz, de Blesser, & Willmes, 1991), pre- and post-therapy. Each participant received 50 individual SIPARI voice-training therapy sessions of a 60-minute duration twice weekly, with the fMRI procedure occurring pre- and post-therapy. The authors noted that none of the stimuli of the fMRI paradigm was trained during SIPARI voice-training therapy.

The results of this study are noteworthy: the participant with Broca's aphasia demonstrated significantly increased left-hemisphere activation in perilesional regions following treatment while the two participants with global aphasia demonstrated increased activation of perilesional regions and homologous areas in the right hemisphere. These results are somewhat consistent with Brier et al. (2010) and Schlaug et al. (2009), but recall that Brier et al. did not see clinical signs of improvement in the participant that showed increased activation of the right hemisphere. Similarly, Belin et al. (1996) hypothesized that increased right hemisphere activation was maladaptive in recovery of expressive language.

In contrast, Jungblut and colleagues (2004, 2009, 2014) noted behavioral improvements in language and speech-motor capabilities following their treatment protocol. The authors hypothesized that improved temporal sequencing must occur first in order to improve speech-motor and language abilities and suggested that since singing involves both linguistic and musical components (i.e., rhythmic-sequencing), it may be the best route for this patient population to access language. One difference between SIPARI and MIT in relation to its possible effect on individuals with global aphasia is its use of improvisation through spontaneous utterances set to unique melodic patterns, hypothetically tapping into propositional speech patterns more quickly than through the MIT hierarchy.

(5) Hurkmans et al. (2014) described an intervention called Speech-Music Therapy for Aphasia (SMTA) that required the input and presence of both a speech therapist and music therapist during treatment in order to utilize the strengths of both fields and maximize treatment effects. In SMTA, the speech therapy portion trained at three hierarchical levels: 1) phonemes/ syllables, 2) words, and 3) sentences; while the music therapy portion included elements of "melody, rhythm, meter, tempo, and dynamics," (p. 942), with both portions of treatment occurring simultaneously. At the word and sentence level, functional words and phrases were trained that are personally relevant to the individual, and phonemic and visual (showing mouth references and gesturing) cueing strategies were also used during treatment. The aspects of music therapy follow a similar hierarchy to the traditional MIT protocol in that the first stage was singing, then "rhythmical chanting" (which appears similar to the MIT term sprechgesang), followed by speaking (p. 943). SMTA also incorporated the use of novel melodies in training functional phrases, as in Baker (2000).

Hurkmans et al.(2014) tested the effectiveness of the SMTA protocol for increasing verbal communication in daily life with five participants, three-to-six months post-left hemisphere-stroke, who had Apraxia of Speech (AoS) and aphasia (three of the participants had Broca's aphasia, one participant had global aphasia, and one participant had Wernicke's aphasia). The AAT (Graetz, de Blesser, & Willmes, 1991) was used to determine aphasia type and severity (with the auditory comprehension portion used as a control), while the Diagnostic Instrument of Apraxia of Speech (DIAS; Feiken & Jonkers, 2012) was used to assess AoS severity, both pre- and post-treatment and at a three-month follow-up.

Outcome measures also included the Amsterdam -- Nijmegen Everyday Language Test (ANELT; Blomert, 1992), which tests functional verbal communication, and the Token Test score on the AAT. Testing, using the Modified Diadochokinesis Test (MDT) and the Psycholinguistic Assessments in Language Processing of Aphasia (PALPA 12; Kay, Lesser, & Coltheart, 1996), was conducted weekly during treatment, with the MDT being the experimental test and the PALPA 12 (repetition of numbers) being the control. All test administrators were blind to the research study and different from the clinicians conducting treatment.

The SMTA protocol consisted of 24, 30-minute sessions, occurring twice weekly. The participants were also assigned 30-minutes of homework to practice three times per week on non-treatment days. The homework consisted of recordings of the target items. During the course of treatment, participants were also enrolled in physical and occupational therapy. Participant 3 (global aphasia), showed improvements with “intelligibility” and “comprehensibility of verbal communication” subtests of the ANELT, articulation of phonemes and DDK on the DIAS, repetition on the AAT, and demonstrated decreased aphasia severity, as measured by the Token Test (subtest of the AAT). No improvement was noted on the auditory comprehension portion of the AAT (the control task).

Of note, Participant four (crossed Broca’s aphasia) did show improvement on both control tasks, which the authors attributed to increased sustained attention and working memory, and which by clinical observation, improved throughout treatment. Whether or not this improvement was directly related to SMTA is unknown. Overall, improvement was noted through positive outcome measures in four-out-of-five participants on the “intelligibility” and “comprehensibility of verbal communication” subtests of the ANELT. All participants displayed improvements in aspects of accuracy, consistency, and fluency of articulation, as measured by the DIAS. Researchers acknowledged the limitations of a small sample size and noted that a randomized controlled trial would provide the most empirical evidence to support the effectiveness of SMTA to treat a variety of types of aphasia and AoS.

(6) Al-Janabi et al. (2014) recognized the relative strength of singing/intoning phrases with patients with global aphasia and conducted a study pairing MIT with brain stimulation to increase verbal expression. This study combined MIT with rTMS (repetitive transcranial magnetic stimulation), which was applied to the right Broca’s speech homolog with the purpose of improving language function in people with aphasia. Al-Janabi et al. explains that Vines et al. (2011) used brain stimulation to the right hemisphere in combination with MIT but the anodal-tDCS electrodes used had poor focality of stimulation, such that it was impossible to determine which specific area in the right hemisphere caused increases in verbal expression. Therefore, the use of rTMS in the current study allowed for better insight into the specific regions of the right hemisphere in which stimulation potentially proved effective.

There were two participants in this study, both male and almost a year post-stroke. The Western Aphasia Battery (Kertesz, 2007) was used to determine aphasia severity. Participant 1, GOE, was categorized as having moderate non-fluent Broca’s aphasia and Participant 2, AMC, was categorized as having moderate-to-severe non-fluent Broca's aphasia with deficits in auditory-verbal comprehension. Language function measures, which consisted of automatic speech tasks and a phrase repetition task, were assessed at two points before treatment, following each treatment session, and one week post-study completion. fMRI data were used to assist in positioning of the electrodes and for tracking any possible cortical reorganization during the study. A naming/reading task was administered using a block-design during fMRI data collection.

The study consisted of two phases (three treatment sessions each), each separated by a one-week break. Each session consisted of a five-minute rTMS session or sham rTMS session, followed by a five-minute break and then 40 minutes of MIT. MIT followed the traditional protocol and each phrase was trained using the two-toned intonation pattern. Participant GOE began with rTMS and AMC began with sham rTMS, to which treating clinicians were blind.

A one-sample *t*-test across five testing points was applied to the percent correct performance outcomes of the phrase repetition task, the verbal fluency task, the automatic speech task, and the naming/reading task. The first participant, GOE, demonstrated significantly greater improvement in the treatment phase for the rTMS treated phrase repetition list but improvement was delayed "such that the treatment gains only became evident at later testing points," (p.5). AMC did not show any significant improvement. Verbal fluency was scored by percentage of items produced in the correct order for each category and scores were used to calculate mean percentage accuracy at each testing point. GOE showed significant improvements in verbal fluency in the treatment phase but AMC did not. Both GOE and AMC showed improvement in the automatic speech task, while their performance in the naming/reading task both decreased.

Region-of-interest (ROI) analyses were applied to the fMRI activations of Broca's area (referred to in two sections, as the pars opercularis and the pars traingularis) pre- and post-treatment. ROI analyses revealed significantly increased activity in the left BA44 (pars opercularis) and significantly decreased activity in its right homolog for participant GOE. Participant AMC, however, demonstrated significantly increased activity in the left and right BA45 (pars triangularis). Researchers reported that both participants only demonstrated modest behavioral effects from treatment, which they related to the fact that only three MIT sessions were administered. The authors also noted that GOE's improvement consisted only of those MIT phrases practiced in combination with rTMS. Al-Janabi et al. explained that "as MIT training is highly repetitive and becomes automatized, neural changes are restricted to mechanisms involved in automatic speech rather than (the more conscious) propositional speech, which is required for the naming/reading task," therefore, MIT may limit the possible generalization of trained tasks to functional environments, (p. 9).

Blank et al. (2002) explored the brain activation patterns of three types of speech: propositional (non-formulaic, spontaneous speech), non-propositional (formulaic, automatic, e.g., counting or nursery rhymes), and covert (internal speech without articulation), using positron emission tomography (PET). Results revealed differing activation patterns for all three types of speech, which suggested that in order to generalize trained phrases of a formulaic nature (e.g., MIT trained phrases), the phrases must transition from an automated speech task to a propositional speech task, or that propositional speech tasks should be targeted in therapy, as opposed to the more repetitive-type speech tasks found in MIT.

Blank et al.'s (2002) findings were further supported by Stahl, Hensler, Turner, Geyer, and Ko's (2013) findings, which also suggested that different cortical patterns are involved in propositional versus non-propositional speech. This proposition begs the question: If trained behaviors become automatized through the repetitive nature of MIT training, do they ever have the ability to become learned, rather than performed, behaviors?

Jungblut et al. (2009) addressed the repetitive nature of music-based intervention by adding an improvisational component in the development of SIPARI, which may have accounted for the study participant's increases in spontaneous speech; however, Jungblut et al. attributed these increases to the temporal-rhythmic components, which contradicts Blank et al.'s (2002) findings. Due to the complex nature and mixed conclusions of what underlying mechanisms affect clinical outcomes in individuals with aphasia, theories of learning were also explored to discover other avenues for increasing use of functional phrases outside of therapy.

One vein of research explored the relationship between trained behaviors and their possible generalization outside of treatment. Titze and Verdolini Abbott (2012) hypothesized that in order for a targeted behavior to generalize, it must first become a learned behavior, as opposed to a performed behavior. Coppens and Patterson (2015) further supported this notion by explaining that a learned behavior is durable over time and transfers to different environments and contexts, while a performed behavior requires that certain conditions be met (such as a specific environment, e.g., a treatment room). Additionally, Fox et al. (2002) stated that learned behaviors rely on internal cues, whereas performed behaviors are initiated by external cues.

The types of behaviors trained in therapy have the potential to affect generalization outcomes. Al-Janabi et al.'s (2014) and Blank et al.'s (2002) findings demonstrated that not all types of verbal expression create the same brain activation patterns. They purported that other regions of the brain, besides Broca's and Wernicke's area, are active during different types of speech, and connections through white matter pathways to and from these other areas could be the cause of expressive language deficits in individuals post-stroke, further adding to the complexity by which language deficits are viewed in relation to brain lesions.

These insights into the types of therapy techniques that facilitated carryover to non-formulaic/ propositional speech are important factors to consider when designing a treatment protocol. The original MIT protocol by Albert et al. (1973) developed a treatment heirarchy (as discussed in further detail previously), containing four levels of difficulty, to decrease gradually reliance on more rhythmic and song-like intonation patterns and systematically transition to the typical prosody found in everyday conversation.

Coppens and Patterson (2015) reviewed existing aphasia literature and suggested the following strategies to promote stimulus generalization: 1) training targets in the context of a sentence (as opposed to training a verb or noun in isolation), 2) training more complex items, 3) training functional targets, 4) pairing treatment with a home practice program, and 5) including discourse/ conversational training in therapy. Further complicating the ability of researchers to apply causal relationships between intervention and generalization; however, are the number of variables that can affect generalization of trained behaviors, including individual variability, aphasia type and severity, cognitive skills (e.g., executive function), level of motivation, amount of practice, dosage of treatment, and the number of exemplars (Coppens & Patterson, 2015).

These theories concerning generalization of treatment were considered in order to determine how other factors could potentially influence outcomes.

Conclusion

With van der Meulen et al.'s (2014) randomized controlled trial, MIT now has proven efficacy with Phase III research (based on Robey's [2004] Five-Phase Model) with individuals with Broca's aphasia in the subacute phase of recovery. In combination with over 40 years' worth of clincial observations, case studies, and smaller sample studies (Phase II research, Robey (2004)), clinicians can be confident in the evidence base for treating the specified population effectively using MIT.

Where evidence is promising, but more empirical validity is warranted, is in the area of treating those with more severe nonfluent aphasia, such as global aphasia, using the tenents of MIT as a framework for treatment. Both Baker (2000) and Jungblut (2009) have demonstrated positive outcomes in treating individuals with auditory comprehension deficits using a music-based intervention, but more research is needed to determine the specific treatment effects in increasing auditory comprehesion alongside verbal expression in individuals with global aphasia. Therefore, the present study (M-STIM) saught to combine the most effective elements of modified MIT (as determined in Mayer, Jeppson, and Henning (2014)) in conjunction with multimodal stimulation to enhance both verbal expression and auditory comprehension in a single participant with global aphasia.

Mayer, Jeppson, & Henning, 2014: Precursor to the Current Study

Procedures: Mayer et al. (2014)

Prior to initiation of therapy, 40 functional phrases were developed as treatment targets with input from KC's family. These phrases were then divided into two sets of 20, with both sets containing an equal number of syllables, and balanced in terms of semantic category (e.g., feelings, activities, and daily routines). To test the treatment efficacy of traditional MIT compared to the novel MMIT protocol, half of these stimuli were trained utilizing the two-toned pattern of MIT, while the other half were given unique melodies (MMIT) such that each melody: (1) ranged from three to nine notes, with each note carrying one syllable of each phrase; (2) was unique from any other in note sequence and rhythm; (3) spanned a 5-note range from C4-G4, to ensure comfortable singing for an untrained female voice; and (4) centered around the C Major scale. Additionally, (5) word painting was utilized to aid in memorization (e.g., question phrases, i.e., "how are you?" had a melody line that went up in pitch; bedtime phrases were written with longer note values to give a "tired, more languid" sensation). The unique melodies for both Mayer et al.'s study and the current stud (M-STIM) were created by the first author, who has a Bachelor of Music degree.

During treatment sessions, the clinician used a piano application on a smartphone to play phrases trained with MMIT. Additionally, a rhythmic component was utilized for both MIT and MMIT trials in the form of hand tapping, in unison with the clinician, during specific steps of the MIT hierarchy. A modification from both Sparks' (2008) and Baker's (2000) protocols included the use of visual aids for all trained phrases. Visual aids were presented during probes and during the training of each phrase. All sessions were digitally recorded for both video and audio.

During each treatment session, all phrases (MIT and MMIT) were trained using the MIT protocol hierarchy (following Sparks, 2008), with each level (from I to IV) in the hierarchy increasing in difficulty (see Table 2). Initially, all phrases were trained at Level II. After 90% accuracy over five consecutive treatment sessions was achieved at Level II, phrases were trained using the Level III protocol.

Table 2

Level II and Level III MIT treatment protocol hierarchy (Sparks, 2008)

Level II	Step 1	the clinician modeled the phrase using the two-toned intonation pattern (MIT phrases) or played/sang the melody (MMIT phrases)
	Step 2	the clinician signaled KC to join in and intoned the phrase along with the clinician, which was also accompanied by hand-tapping. Then, phrase repetition continued in unison with the clinican slowly fading out
	Step 3	The clinician signaled KC to stop, intoned the phrase, and asked KC to repeat it
	Step 4	After successful repetition, the clinician asked KC, “What did you say?” to ilicit a delayed repetition response
Level III	Step 1	the clinician modeled the phrase once and then signaled KC to join in, accompanied by hand tapping and subsequent fading out of the clinician
	Step 2	the clinician intoned the phrase and signaled KC to repeat it, along with continued hand tapping
	Step 3	hand tapping stopped and the clinician asked a question that required a response using the trained phrase (e.g., “What do you drink in the morning?” with the appropriate response being the trained phrase “cup of coffee” or the target word “coffee.”)

In a typical 60-minute treatment session, the first five minutes of treatment consisted of conversation between the clinician, KC, and her family, to establish rapport and informally assess improvement in propositional speech. During most sessions, a primer was used in the form of playing a familiar song (e.g., Don McLean's "American Pie," using YouTube on a smartphone), during which KC was encouraged to sing along (Baker, 2000). Initial probing to test carryover and generalization of trained phrases was then conducted for approximately 10 minutes. Following probes, the remainder of the session was utilized to train each phrase using the previously described MIT/MMIT hierarchy.

During Mayer et al. (2014), KC received 24 weekly 60-minute therapy sessions (February 2014 to November of 2014). KC was also encouraged to practice trained functional phrases daily at home, and all phrases were recorded on her smartphone so that she could do so independently. Family reports indicated, however, that practice occurred only one to two times per week outside of therapy; due to KC's cognitive deficits, she needed consistent cueing to operate the smartphone for practice, thus creating a barrier to practicing trained phrases outside of therapy sessions.

Results: Mayer et al. (2014)

Upon completion of Mayer et al.'s (2014) study, KC demonstrated increased verbal expression in trained phrases, given a verbal cue. These trained phrases, while not generated independently, provided an avenue for verbal expression, which, importantly, was not present before treatment. KC's significantly increased Western Aphasia Battery Aphasia Quotient (WAB AQ) from January 2014 to September 2014 also provided strong evidence suggesting that the

MIT/MMIT protocol aided in her overall progress (see Table 3). Due to KC's marked functional impairments, the primary goal of treatment was not necessarily to re-activate combinatorial, propositional expressive language, but to train a set of functional phrases that KC could use as needed in everyday life situations. However, despite her ability to generate trained phrases when given a cue or prompt from a clinician or or caregiver, KC remained unable to generate any phrases independently outside of therapy. Mayer et al. (2014) hypothesized that this may have been due to her persistent auditory comprehension deficits, as evidenced by the lack of change in KC's WAB auditory comprehesion scores from pre-treatment testing to re-testing in Fall 2014 (see Table 3).

3. METHODS: M-STIM

Participant

The single participant in this study, KC, was a right-handed 51-year-old female, 22 months status-post a large, left-hemisphere stroke in the distribution of the left carotid, resulting in right hemiparesis and global aphasia. KC spent 5 weeks in the intensive care unit following her stroke and 6 weeks in inpatient rehabilitation, with verbal output restricted only to “ling ling ling” and accompanied by severe auditory comprehension deficits. She received outpatient services upon discharge, during which time her stereotypy changed to the phrase, “on a Wednesday,” but was discharged from therapy after just one month due to lack of any additional progress.

KC initially presented to our university clinic nine months following her stroke, seeking additional therapy and support. At this time, a thorough assessment was conducted, and KC was found stimulable for verbal expression through choral singing and the two-note intoned pattern of MIT. Therefore, a treatment protocol was developed utilizing a single-subject design in which half of the family-selected functional phrases were trained with traditional MIT and half were given unique melodies and trained with a modified version of MIT, MMIT. As described previously, results from Mayer, et al.’s (2014) study indicated that compared to stimuli trained using MIT, those trained using MMIT were learned more quickly and retained for a longer period of time; however, deficits in auditory comprehension appeared to prevent generalization of these gains to functional situations. Thus, during the M-STIM treatment phase, which began at

22 months post-stroke, we augmented MMIT therapy by training functional phrases using MMIT in conjunction with strategies to address KC's auditory comprehension deficits (see Figure 2 for treatment timeline).

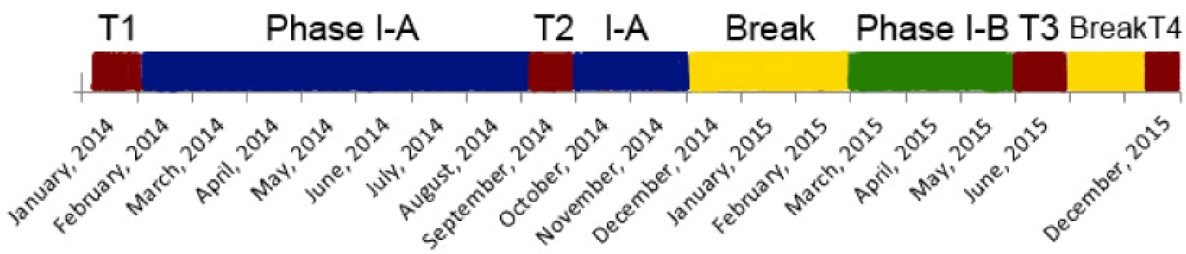


Figure 2. Treatment timeline beginning with initial assessment

T1= Initial assessment of WAB and TONI-3; T2= Evaluation during the Mayer et al. (2014) intervention of WAB; T3= Evaluation after the M-STIM Phase intervention of WAB and TONI-3; T4= Evaluation of WAB six months post-the M-STIM Phase; Break= time with no intervention

KC's previous medical history included a 30-plus year history of a seizure disorder (controlled with medication), alcohol abuse, and tobacco use. At the time of this study, she resided with her husband in a small town southwest of the Chicago suburbs. Her family support network consisted of her husband, grown children, and mother. Following her stroke, she was unable to return to work due to the severity of her deficits; she also possessed few hobbies, and her family reported that she spent most of her days watching television. Social history included a high-school education; she had earned a living as a server at a local diner. At the beginning of outpatient services at the university clinic, KC required a caregiver be present at all times and needed assistance in completion of activities of daily living. KC possessed average musical knowledge, i.e., ability to sing along to popular songs on the radio, but had no formal musical training.

Pre-treatment Assessment

Initial questionnaires, including the American Speech-Language-Hearing Association Quality of Communication Life (ASHA QCL) Scale (Paul, et al., 2004), and family interviews were conducted to determine KC's view of communication, level of desire to communicate and in which contexts. These data were used to create targeted treatment phrases (see Table 4). Standardized pre-treatment testing was conducted using the Western Aphasia Battery (WAB; Kertesz, 2007), to assess spontaneous speech, auditory comprehension, repetition, and naming/ word finding, and the Test of Nonverbal Intelligence-Third Edition (TONI-3) (Brown, Sherbenou, & Johnsen, 1997) , to assess nonverbal abstract problem solving ability. Table 3 shows KC's initial assessment results at 9 months post-stroke and 16 months post-stroke (after 7 months of Mayer et al.'s intervention). Results from the WAB indicated very severe aphasia; the WAB defines global aphasia by the following scores: fluency (spontaneous speech), <5; auditory verbal comprehension, 0-3.9; repetition, 0-4.9; and naming/word finding, <7. The TONI-3 is standardized with a mean of 100 and standard deviation of 15; KC was tested over two separate sessions secondary to fatigue and ranked in the 3rd percentile with a quotient of 72 out of 160, revealing severe cognitive deficits.

Table 3

Standardized Assessment Results

		Max	T1	T2	T3	T4
WAB	Spontaneous Speech Score	20	1	4	1	7
	Auditory Verbal Comprehension Score	10	3.45	3.45	4.65	5.2
	Repetition Score	10	1.1	1.8	2.4	2.7
	Naming and Word Finding Score	10	1.1	1.3	2.1	2.4
	Aphasia Quotient	100	13.4	21.1	24.1	34.6
TONI -3	Quotient	160	72	N/A	90	N/A
	Percentile Rank	100	3	N/A	26	N/A

Notes: WAB= Western Aphasia Battery; TONI-3= Test of Nonverbal Intelligence-Third Edition; Max= maximum score; T1= initial evaluation; T2= evaluation after the MMIT Phase of treatment; T3= evaluation after the M-STIM Phase of treatment; T4= evaluation at 6-months post-the M-STIM Phase follow-up

Study Design

Given the results of Mayer et al.'s (2014) study, M-STIM was developed to continue Mayer et al.'s protocol while specifically and concurrently addressing KC's auditory comprehension deficits to further improve access to language and increase appropriate (cued or uncued) verbal expression. In addition, because results of Mayer et al.'s study demonstrated superiority of MMIT over MIT in terms of rate and durability of learning, additional MMIT stimuli were incorporated into the M-STIM Phase. Framed in Robey's (2004) model, the current study was conducted with the goal of better delineating best practices for a protocol to maximize expressive language, through music and multi-modal stimulation, for individuals with global aphasia.

We utilized a discovery-oriented single-subject multiple-baseline design following Robey's 2004 Five-Phase Model for Clinical Outcome Research, with the present study (M-STIM) consisting of a continuation of Phase I research. This study was completed with Institutional Review Board approval for research with human subjects. The first and second authors digitally recorded for audio and video and conducted all sessions, ensuring treatment fidelity.

Procedure

Treatment sessions followed a similar protocol as in Mayer et al.'s (2014) study. At the beginning of every session, the first five minutes consisted of conversation with KC and her family. KC and the clinician then sang along to a popular song (using YouTube on a smartphone) to act as a primer and confidence-builder for KC, given that singing was one of her relative strengths. Then, probing commenced to test carryover and generalization of trained phrases. After that, all 18 phrases (see Table 4) were trained every session using the MIT/MMIT protocol described above, with the incorporation of multimodal stimulation (M-STIM).

During M-STIM, KC attended 10 weekly, 60-minute treatment sessions over the course of 15 weeks, including two follow-up sessions beginning three weeks post-treatment. KC was also instructed to practice trained phrases twice daily for homework. All of the trained phrases were recorded on her iPAD on one file, and the procedure for how to access the file was rehearsed during therapy with the hopes that KC would learn how to access these files herself, without the aid of a caregiver.

Treatment Protocol

In the current study (M-STIM), 18 functional phrases were developed based on the results and information gathered from Mayer et al.'s (2014) study (see Table 4). In order to equitably compare KC's performance from Mayer et al.'s study to the M-STIM Phase, only one variable per trained phrase was changed in the M-STIM Phase. To address research question #2, the design and context of phrases were modified such that: (1) Pre-selected MIT phrases were given more unique melodies (i.e., MMIT) to determine if the uniqueness of the melody improved recall (e.g., "nice hot shower" became "I want to take a nice hot shower (Figure 3)); (2) Pre-selected MIT and MMIT phrases remained MIT or MMIT respectively (i.e., the two-toned intonation pattern or unique melody was retained), but the specificity of the words were changed to increase salience (e.g., "Can we order food?" became "Italian beef please" (one of KC's favorite foods)), (see Figure 4); or (3) Phrases developed in the MMIT Phase continued to be trained, but multimodal stimulation was added throughout varying contexts to augment auditory comprehension. For example, the trained MIT phrase, "I am cold" was accompanied with a cold icepack that KC could touch in order to incorporate another sensory modality. In addition, during training of this phrase, a picture of snow or two people skiing down a mountain was displayed in order for KC to create multiple associations. These visual aids were varied every week to increase the number of contexts in which KC could (in theory) use and understand the meaning of these trained phrases. Physical objects were also incorporated into the session for select trained phrases. For example, a toothbrush and toothpaste were placed on the table while training the phrase, "I'll brush my teeth." Additionally, for phrases containing "feeling" words, such as "That's really funny" or "I am sad", short (~30 sec)YouTube videos of babies or animals that incorporated the concept of the feeling word were shown to KC before training the phrase. Every week, different videos or pictures were provided during training to continually vary the contexts in which the phrases were trained.

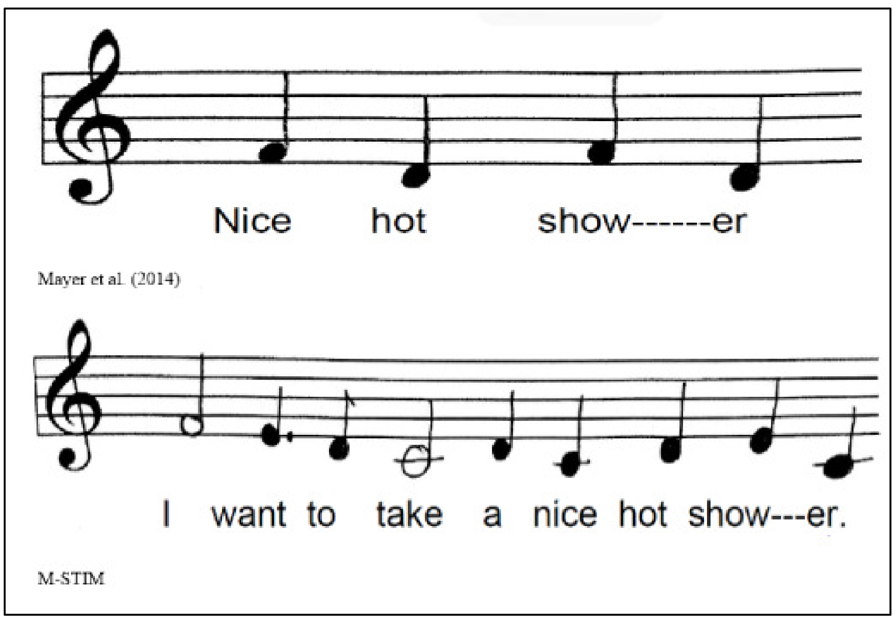


Figure 3. Mayer et al.'s (2014) notation and M-STIM's increased melodic uniqueness modification

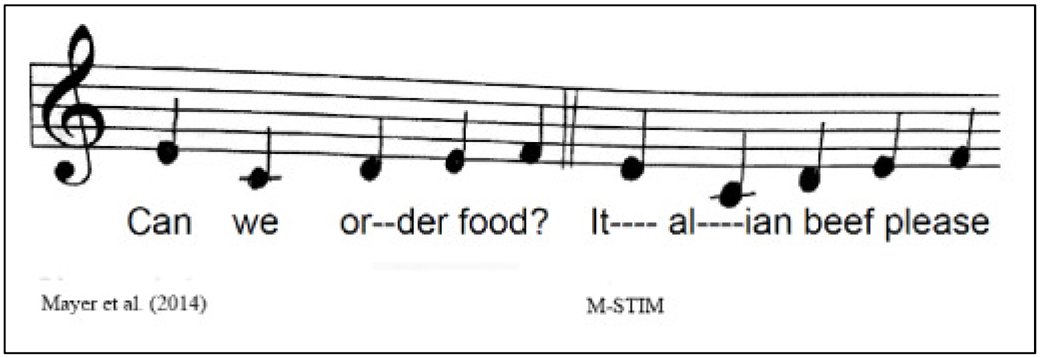


Figure 4. Mayer et al.'s (2014) notation and M-STIM's increased salience modification

Training of all phrases followed Levels II and III of the MIT protocol (Table 2) as described in the synopsis of Mayer et al.'s (2014) study. However, since some phrases were trained during Mayer et al.'s study and some were new in the M-STIM Phase, determination of training level was individualized per phrase based on KC's level of success, meaning that some phrases were trained at Level II, while other phrases were trained at Level III, within the same treatment session. Order of stimulus presentation was randomized at each session to factor out fatigue during the practice of any one phrase.

Table 4

Eighteen functional phrases trained every session in the M-STIM Phase, showing the modifications compared to Mayer et al. (2014)

Phrase	MIT/ MMIT	Physical Object	Tactile Cue	YouTube Video	Increased Salience	Increased Melodic Uniqueness
Cup of coffee	MIT	X				
Please curl my hair	MMIT	X				
I'll brush my teeth	MMIT	X				
Don't do that, please stop	MMIT					X
How are you doing today?	MMIT					X
I want to take a nice hot shower	MMIT					X
Can you get me something please?	MMIT					X
That's really funny	MMIT			X		
I love you	MIT			X		
I am very happy	MIT			X		
I am sad	MMIT			X		
Time for therapy	MMIT					X
Let's go look at pictures	MMIT					X
I am hot	MIT		X			
I am cold	MIT		X			
Watch "Bonanza"	MIT				X	
Italian beef please	MMIT				X	
Let's go to Kohl's	MIT				X	

Outcome Measures

Within-treatment Probes

Probes were completed for all 18 functional trained phrases at the beginning of each treatment session. During probes, the question (a natural verbal cue, asked in Step 3 of the Level III MIT hierarchy) (see Table 2), was used to elicit the appropriate trained response from KC. For example, the appropriate response to the question "What do you drink in the morning?" is "cup of coffee." Due to KC's aphasia severity, a variety of cues were necessary to elicit the desired response; these were tracked and scored (described below) during probes. If KC did not generate an appropriate response, then a physical object or tactile cue would be placed on the table (e.g., a coffee mug). If KC appeared not to understand the question, it was rephrased (e.g., "What do you like to drink?"). If KC was still unable to provide the appropriate response, then a phonemic cue for the first sound in the word or phrase was provided (e.g., /k/ in "cup"). If she was still unable to respond, then the clinician provided the first word (e.g., "cup").

For MMIT phrases, the unique melody was played on a piano (available as a smartphone application), before the phonemic cue was presented, to operate as a type of mnemonic device to help trigger recall of the phrase. The number and variety of cues required for KC to respond appropriately were tracked, with all probe responses recorded using an M-STIM scoring system designed to reflect the amount of cueing required for an appropriate response (see Table 5 for the probe scoring system and examples). This system was developed based on KC's severity level and modeled after Porch's (1967) Porch Index of Communicative Capability (PICA); i.e., a simple "+/-" scoring system would have failed to capture clinically significant progress over time. Thus, KC could earn a minimum of 0 and maximum score of 9 points per probed phrase, with a total of 162 maximum points per probe (i.e., 18 phrases). In addition, due to the severity of KC's aphasia, we counted understandable phonemic paraphasias as correct because these were functional for KC.

Table 5

M-STIM scoring system for within-treatment probes

9	Complete, independent, accurate or close approximation given pictorial cue + question (e.g., Clinician asks, "What do you drink in the morning?" while displaying a picture of a mug; KC responds "cup of coffee" or "dup of coffee")
8	Complete, independent, accurate or close approximation given object/tactile cue + question (e.g., Clinician asks, "What do you drink in the morning?" while handing KC a mug; KC responds "cup of coffee" or "dup of coffee")
7	Rephrase question x1 to increase auditory comprehension
6	Rephrase question x2 to increase auditory comprehension
5	Rephrase question x3 to increase auditory comprehension
4	(MMIT only): Musical cue (Clinician plays the notes of the phrase, without words, on the piano app; KC listens)
4	(MIT): Musical cue/humming (Clinician hums the two-tone intonation pattern for the targeted phrase)
3	Phonemic cue alone (first sound). (Clinician says "/k/…", KC responds "cup of coffee")
2	Phrase + phonemic cue (Clinician: "cup of /k/…")
1	Choral singing (Clinician cues patient: "Sing it with me, together. *Cup of coffee*.")
0	No response or unable to get an approximation even with maximal cueing

Method of deriving scores from multiple raters for within-treatment probes

Eight raters, all undergraduate students studying communicative disorders, were carefully trained by the author to collect pertinent session data in which to apply the M-STIM scoring system.

Data were collected by the raters during nine Mayer et al. (2014) and nine M-STIM treatment sessions by recording KC's observed behaviors during within-treatment probes through a tallying system. The recorded observed behaviors were converted to the M-STIM scoring system by the author and specific precautions were taken to avoid pitfalls and discrepancies among raters' observations. The author also met with the raters in a group training session with the raters given opportunities to observe all of the sought-after behaviors in a recorded treatment session in order to collect the data with the author present, who provided feedback on the raters' performance.

To reduce the possibility of observer drift, meaning that observers have a tendency to change the applications of a definition of an observed behavior over time, the author implemented the following , as suggested by Kazdin (1982): The author trained eight raters who were assigned to view four-to-five treatment sessions, with the order of sessions randomized, such that the raters were blind to KC's progression through the M-STIM protocol. For example, rater 1 observed and rated sessions 1, 5, 8, and 13 and did not view them in any specific order.

Kazdin (1982) stated that the complexity of the observation could affect the level of interrater agreement. He explained that the number of different responses scored in a given period and the range of client behaviors performed within that period could increase complexity, such that the greater number of different behaviors the client performs, the lower the inter-observer agreement would likely be. Kazdin recommended that since agreement for a given response may be influenced by the number of different behaviors and other types of responses being observed in a given timeframe, observers needed to be trained at higher levels of agreement for each of the codes. In order to address the possible influence of complexity on the raw data collected by the observers, the author video-recorded all sessions in order for the raters to pause the DVD to record data and re-watch portions in which the raters questioned the observed behavior.

Additionally, due to the highly complex nature and nuances of the observed behaviors, the author developed the following system by which outliers, that would falsely inflate KC's probe scores and, ultimately, affect the outcomes of this study, were removed: 1) When two-out-of-three or three-out-of-four raters' data resulted in scores that were in close proximity, (within 10 percentage points), those scores were averaged and the mean score was applied to interpret the data, with the outlying rater's data (greater than 10 percentage points from all other scores) being discarded; 2) When all three raters' data resulted in scores that all differed by greater than 10 percentage points, data from all three raters were used to create an average score and the mean was applied to the interpretation; 3) When only two raters collected data per treatment session, both data sets were scored and averaged, and the mean score was utilized in the interpretation.

Interrater reliability agreement

Due to the complexity of the scoring system, all raw scores were submitted for interrater reliability analyses prior to within-treatment data analyses. As mentioned previously, eight raters, undergraduate students in communicative disorders, were carefully trained by the first author to view pre-recorded treatment sessions and to report and tally specific observed behaviors that were then applied to the scoring system.

Eight treatment sessions and one follow-up session from Mayer et al.'s (2014) study, and eight treatment sessions and one follow-up session from the current study (M-STIM) were submitted to interrater reliability analyses. The raters' tallies (raw scores) were converted to the M-STIM scoring system for each trained phrases' probe performance. Then the percent correct scores were added, and the combined scores per rater per treatment session were analyzed using the intraclass correlation coefficient (ICC) that was derived using Excel's Anova: Two Factor without Replication data analysis and then calculated from the output (see Figure 5).

$$\boxed{\frac{var(\beta)}{var(\alpha) + var(\beta) + var(\varepsilon)}}$$

Figure 5. Formula for deriving intraclass correlation coefficient.

Notes: *var(β)*=the variability due to alterations in observed behavior from one treatment session to the next; *var(α)*=the variability of differences in the rating levels of observed behaviors; *var(ε)*= variability due to the differences in ratings of the observed behaviors by the raters.

The ICC statistic was chosen based on a number of factors that Hallgren (2012) reported as significant in choosing a statistic to determine interrater reliability, with the main reason being that the ICC incorporateed the magnitude of disagreement among raters to compute interrater reliability estimates. Hallgren's recommendations for choosing an interrater reliability measure based on the observational data suggested using the ICC for ordinal data, such as a scaled scoring system. Since the M-STIM scoring system was developed to increase detection of clinically relevant changes in KC's performance, it appeared more appropriate to use a statistic that would account for the magnitude of disagreement as opposed to absolute disagreement. The larger-magnitude disagreements result in lower ICC values than smaller-magnitude disagreements, with ICC values ranging from 0, representing no agreement, to 1, representing perfect agreement. The raters' scores across eighteen sessions revealed an ICC of 0.977, signifying an "excellent" correlation with a high level of interrater reliability agreement (Hallgren, 2012, p. 9). With a 95% confident interval (alpha=0.05), the ICC values ranged from 0.955 to 0.993 (See Appendix B). This high level of interrater reliability agreement also suggested that the observer drift variable had little effect on interrater reliability agreement.

A scatterplot containing the observers' ratings of probed behaviors per analyzed session was created to visually inspect the correlation coefficient, with Rater 1's combined scores appearing in blue and Rater 2's scores appearing in red (Figure 6). It should be noted that Rater 1 and Rater 2 were not the same individuals across all sessions, as eight raters were assigned 4-to-5 randomly ordered treatment sessions, with a total of two raters assigned to each session. As shown in Figure 6, the majority of raters' scores converged, with only four points in which the raters' scores differed by more than three points.

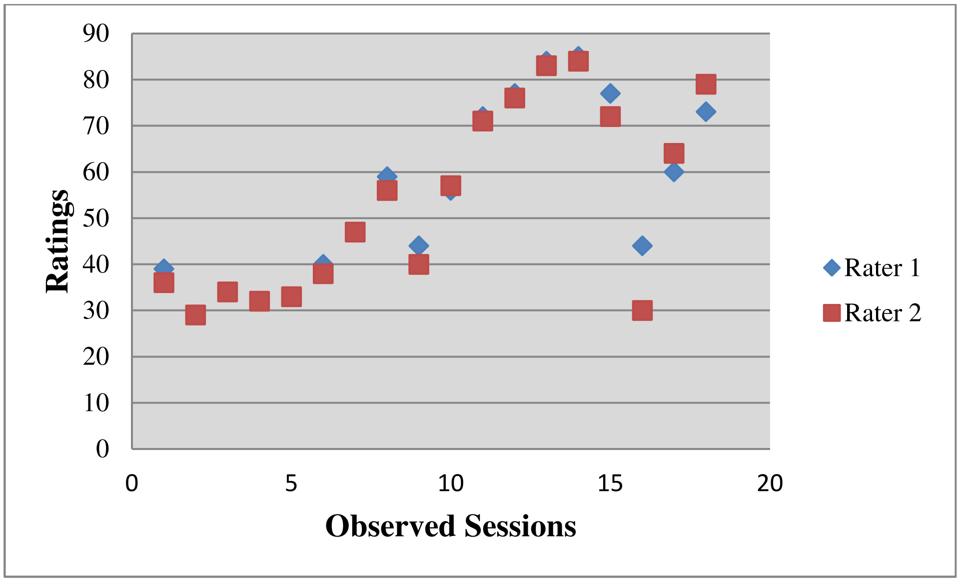
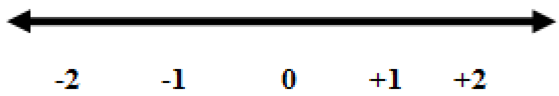


Figure 6. Interrater reliability agreement across 18 treatment sessions

Caregiver Questionnaire

After conclusion of the M-STIM Phase, KC's family completed a questionnaire developed by the first author (see Appendix A) asking them to rate both the frequency and the appropriateness of which they heard KC use trained phrases during everyday life. KC's caregivers were asked to rate KC's frequency of use of trained phrases in relationship to how often KC used these words/phrases before treatment to directly compare the differences of functional use of these words from pre-treatment to post-treatment (see Figure 7). Whether KC used the word/phrase in the appropriate context was determined through a separate caregiver response, in which the caregivers were asked to rate KC's appropriate use by circling *yes, no*, or *sometimes.*



-2= uses these words a lot less since treatment
-1=uses these words somewhat less since treatment
0= uses these words about the same as before treatment
+1= uses these words somewhat more since treatment
+2= uses these words a lot more since treatment

Figure 7. Rating scale for reporting frequency of use of trained phrases in everyday life

Scoring system comparisons

The comparison of two scoring systems, the one used in this study (M-STIM scoring system, see p.44) and the traditional MIT scoring system, was completed to determine which scoring system more accurately represented KC's clinically significant progress over time to better inform future treatment protocols and outcome measures for patients with severe aphasia. Raw data collected during within-treatment probes from one-third of trained phrase responses (two phrases per modified approach) across 19 Mayer et al. (2014) and M-STIM Phase treatment sessions and 2 follow-up sessions were applied to both the M-STIM scoring system and to the traditional MIT scoring system used in Level III, Step 3 (participant's response to a question) of the MIT hierarchy (Sparks, 2008).

The traditional MIT scoring system allots a therapy participant to earn a total of two points for completion of Step III of Level III, in which the participant is required to provide an appropriate and immediate response using a trained phrase to answer the clinician's prompt. If it is necessary for the clinician to repeat the prompt but the participant responds appropriately the second time, then the participant earned a score of one; however, if the participant failed to respond appropriately the second time, he or she earned a score of 0 (see Table 6). In order to compare both scoring systems equitably, the raw scores were converted to percent correct for each response, with a score of 9/9 in the M-STIM scoring system receiving 100% correct and a score of 2/2 in the MIT scoring system receiving 100% correct.

Table 6

Comparison of the M-STIM scoring system to the traditional MIT scoring system

Range of Scores	M-STIM Scoring System	Associated Percent Correct	Step III, Level III MIT Scoring	Associated Percent Correct
Appropriate Response	8-9	89-100%	2	100%
Prompt Repeated Once with Appropriate Response	7	78%	1	50%
Responded with Further Cueing	1-6	11-67%	0	0%
No Response	0	0%	0	0%

4. RESULTS AND DISCUSSION

Outcome Measures

Post-Treatment Assessment Performance: Western Aphasia Battery

The Western Aphasia Battery (Kertesz, 2007) provides an overall Aphasia Quotient, or AQ, with a maximum point value of 100. The AQ is based on scores of the following subtests: 1) spontaneous speech (e.g., answering conversational questions and describing a picture), 2) auditory verbal comprehension (e.g., responding to yes/no questions, auditory word recognition, and sequential commands), 3) naming and word finding (e.g., object naming, word fluency, sentence completion, responsive speech), and 4) repetition (e.g., repeating single and compound words/numbers, high and low probability sentences, and sentences of increasing length and complexity).

It is well known that individuals with neurogenic cognitive-communicative disorders tend to demonstrate substantial, day-to-day and even moment-to-moment variability of performance; this phenomenon of intra-individual variability has been hypothesized to reflect deficits in working memory and/or attention (Bleiberg, Garmoe, Halpern, Reeves, & Nadler, 1997; Stuss, Pogue, Buckle, & Bondar, 1994). Accordingly, it has been difficult to establish what constitutes a clinically significant change on tests such as the Western Aphasia Battery (WAB; Kertesz, 2007), which are designed for and thus standardized on individuals with aphasia. Kertesz (2006) provided qualitative data regarding characterization of aphasia severity from the WAB Aphasia Quotient (AQ), with an approximately 25-point range constituting each severity category (i.e., an

AQ of 0-25 is considered "very severe," 26-50 is "severe," 51-75 is "moderate," and 76+ is "mild"). Thus a change in severity category from pre- to post- treatment could be considered a clinically significant change.

As seen in Figure 8, KC's AQ increased by 21.2 points, from an initial assessment score of 13.4 in January 2014, to a follow-up-testing score of 34.6 in December 2015. KC's change in AQ from 24.1 to 34.6 (i.e., T3 to T4) constituted a qualitative change in her aphasia severity from "very severe" (AQ = 0-25) to "severe" (AQ = 26-50). Notably, this change occurred from post-MSTIM to follow-up; thus it does not speak to the magnitude of change from pre- to post-Mayer et al. (2014) nor from pre- to post-MSTIM.

A quantitative method for establishing magnitude of change using standardized test protocols is calculating a Reliable Change Index (Jacobson & Truax, 1991), which is based on standard deviation of scores from test standardization samples in conjunction with test-retest reliability data. An RCI of 1.96 between pre- and post-test scores (i.e., a standardized difference between two scores that exceeds the 95% confidence interval) is considered statistically significant. KC's WAB scores were subjected to RCI analysis, with results displayed in Table 7.

Table 7

RCI analysis of WAB AQ

Scores based on treatment phase	AQ change	S_{DIFF}*	RCI
Pre- to post-Mayer et al. (2014) AQ: T1 (13.4) to T2 (21.1)	T2 – T1 = 7.7	1.2982	5.93**
Pre- to post-M-STIM AQ: T2 (21.1) to T3 (24.1)	T3 – T2 = 3.0	1.2982	2.31**
Post-M-STIM to follow up: T3 (24.1) to T4 (34.6)	T4 – T3 = 10.5	1.2982	8.09**

Notes: *SDIFF was calculated based on the Standard Error of Measurement (SEM) for the WAB standardization sample of individuals with Global aphasia; where SEM is equal to the sample's standard deviation multiplied by the square root of 1-r, where r = the test-retest Pearson correlation coefficient.**Significant at p<.05

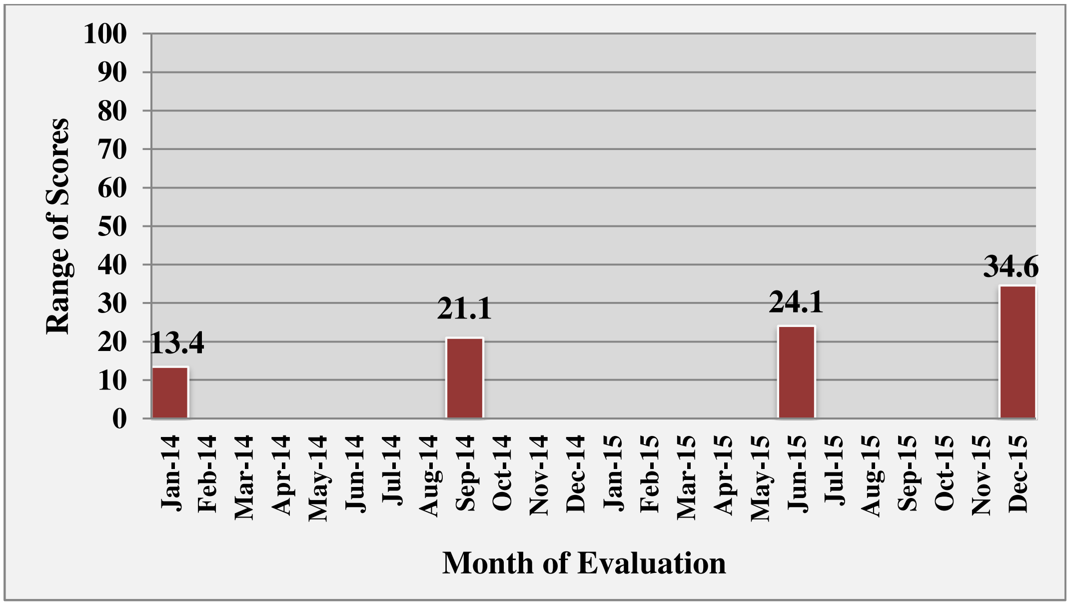


Figure 8. WAB Aphasia Quotient (AQ) from initial pre-testing to six-month follow-up testing

Notes: Jan-14= initial assessment, Sept-14= assessment after 7-months of Mayer et al. (2014) phase, Jun-15= assessment upon completion of the M-STIM Phase, Dec-15= assessment at six-month post-treatment follow-up.

KC's Western Aphasia Battery subtest scores (Figure 9) revealed steady increases in performance in the 23 months from her initial evaluation prior to the start of Mayer et al.'s (2014) study (January 2014) to six-month follow-up-testing completed following M-STIM (December 2015). Prediction 1 (see p.5) hypothesized that KC's auditory comprehension scores would increase as a result of M-STIM, despite a stable baseline score of 3.45 (out of 10) at both pre-testing prior to Mayer et al.'s (2014) study in January 2014 and at post-Mayer et al. testing in November 2014 (see Table 3). Following M-STIM, KC's auditory comprehension score increased by 1.2 points and continued to increase another 0.55 points during 6-month follow-up testing in December 2015 (see Figure 9).

In support of Prediction 2, KC's repetition score and naming/word finding score increased gradually from KC's initial assessment in Mayer et al.'s (2014) study (January 2014) and at each of the three subsequent testing intervals, with the largest increases in scores occurring from post-Mayer et al. testing (September 2014) to post-M-STIM testing (June 2015). KC's spontaneous speech score, however, dropped from 4 to 1 during this same interval of testing, but increased to 7 following six-month follow-up testing (December 2015).

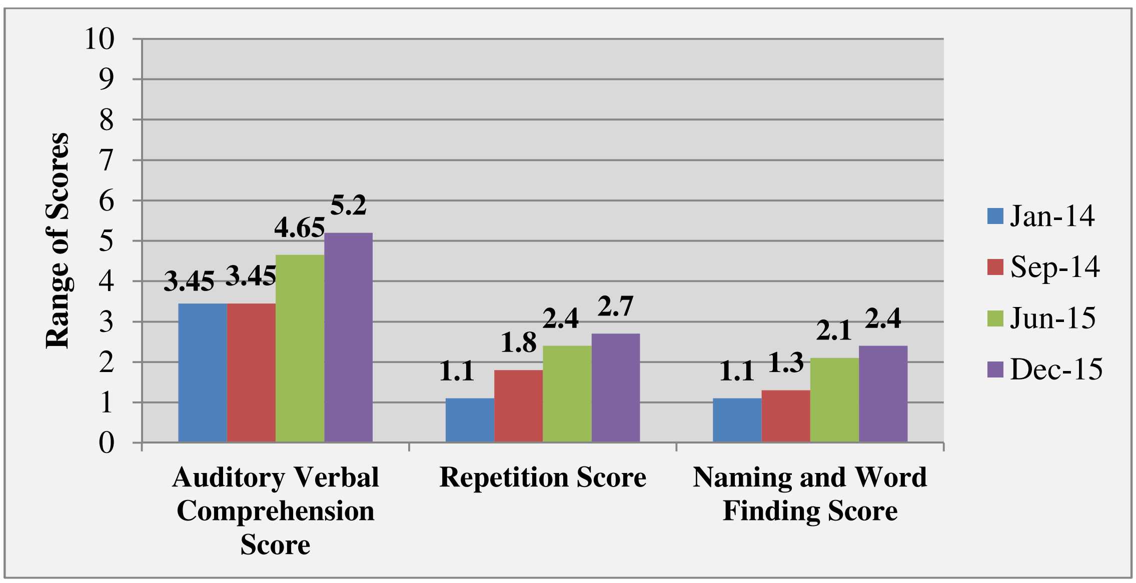


Figure 9. WAB subtest scores from initial pre-testing to six-month follow-up testing

Notes: Jan-14= initial assessment prior to Mayer et al.'s (2014) intervention, Sept-14= assessment after 7-months of the Mayer et al.'s intervention, Jun-15= assessment upon completion of M-STIM, Dec-15= assessment at six-month post-M-STIM follow-up.

Within-treatment Performance: Probes

As a refresher, within-treatment probes consisted of a functional communication question for each trained phrase, with the phrases originally trained in Mayer et al.'s (2014) study being divided into three main subcategories based on the M-STIM protocol designed to maximize phrase acquisition and generalization: 1) increased melodic uniqueness; 2) increased salience; 3) multimodal stimulation (refer to pp. 39-41 for further descriptions). That is, the trained phrases introduced in Mayer et al. were modified (based on the subcategories listed above) in M-STIM, and KC's performance of within-treatment probes from both treatment phases were analyzed to determine the effectiveness of the modifications on KC's performance.

Initial results were derived by averaging within-treatment probe performance scores across Mayer et al.'s (2014) study and the current study (M-STIM) for each trained phrase; these scores were then converted to a percentage, with the numerator being the average number of points earned using the M-STIM scoring system, and the denominator being the total number of points KC could possibly earn for each probe. For example, three phrases were trained to increase salience. The scores for each of the phrases were averaged and then divided by 27, which is the maximum amount of points KC could earn per that set of phrases (9 points possible per probed phrase). The mean and standard deviation of these percentages (see Table 8) were calculated for each modified category at Mayer et al.'s baseline and treatment sessions and for M-STIM baseline, treatment, and follow–up sessions to compare the differences in the mean and standard deviations between modified approaches and across phases of treatment.

Visual inspection of the data revealed a mean difference for all modified approaches, with the largest mean difference (37.5) occurring for phrases trained with increased salience.

Phrases trained with multimodal stimulation demonstrated the second largest mean difference (19), and phrases trained with increased melodic uniqueness showed the least mean difference (12.55). These initial data suggest that increasing salience produced the most change in KC's performance from Mayer et al.'s (2014) study to M-STIM, providing a preliminary answer to Research question #2 (see p. 5).

Table 8

Mean and standard deviation of within-treatment probe scores from Mayer et al. (2014) baseline to M-STIM follow-up

	Melodic Uniqueness		**Increased Salience**		**Multimodal Stimulation**	
	Mean	**Standard Deviation**	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**
Mayer et al. (2014) Baseline	23.66666667	5.686240703	21.66666667	8.504900548	20.5	4.924428901
Mayer et al. (2014) Treatment	26.2	7.014271167	22	4.183300133	30	11.87434209
M-STIM Baseline	27.66666667	9.291573243	22.66666667	0.577350269	38.83333333	6.370504951
M-STIM Treatment	38.75	7.49761867	59.5	22.09718793	49	6.777062153
M-STIM Follow-Up	34	8.485281374	75	1.414213562	42	0.707106781

Notes: Mayer et al. (2014) baseline=three data collection points; Mayer et al. treatment=five treatment sessions; M-STIM baseline=last two Mayer et al. treatment sessions and one Mayer et al. follow-up session the week prior to beginning M-STIM; M-STIM treatment=eight treatment sessions; M-STIM follow-up=two sessions three-weeks post-M-STIM and five-weeks post-M-STIM treatment

When comparing average percent correct of the within-treatment probes from Mayer et al. (2014) baseline to M-STIM follow-up (Figure 10), the following increases were noted: a 13% increase in percent correct of probes in phrases with increased melodic uniqueness, a 41%

increase in percent correct of probes in phrases with increased salience, and a 21% increase in percent correct of probes in phrases with added multimodal stimulation.

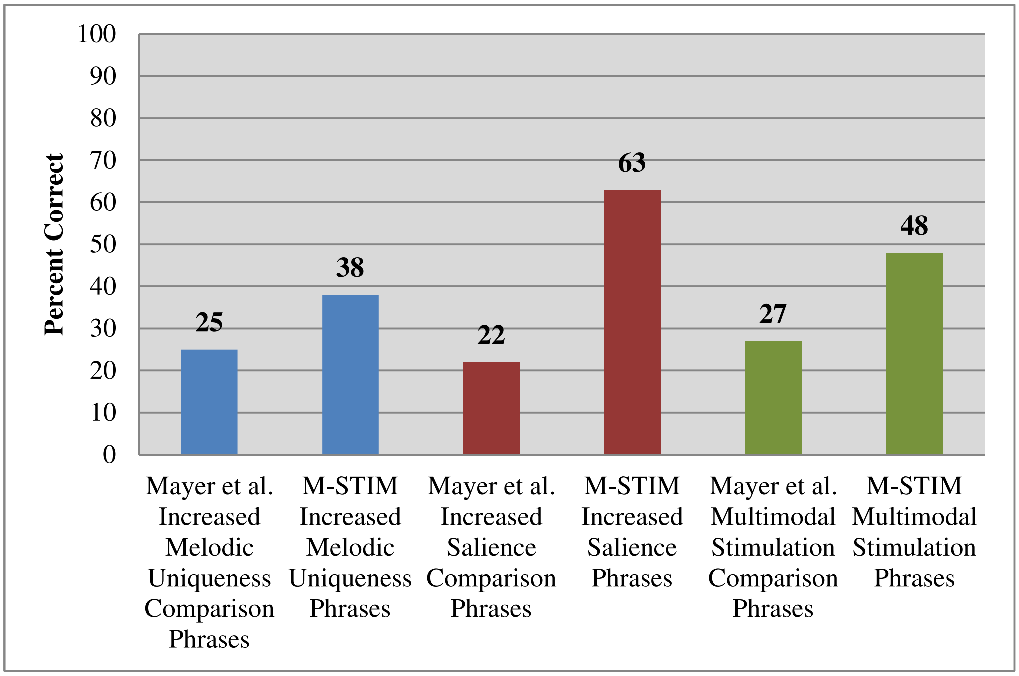


Figure 10. Average percent correct of within-treatment probes across treatment and follow-up sessions from Mayer et al. (2014) to M-STIM phases

Notes: Percent correct data from within-treatment probes were averaged from nine Mayer et al. (2014) treatment sessions and ten M-STIM treatment sessions for phrases modified in M-STIM to increase melodic uniqueness, increase salience, and/or add multimodal stimulation to compare differences in KC's performance with and without these added components/ modifications.

Statistical Analyses

Baseline data

Statistical analyses of the within-treatment probes for each M-STIM Phase modification (increased melodic uniqueness, increased salience to everyday life, and increased multimodal stimulation) were conducted to quantify the results. A time-series analysis using the *C* statistic (Tryon, 1982, 1984) was conducted to address Prediction 3(a) (see p. 5) in order to quantify trends, i.e., "any systematic departures from random variation," in KC's performance over time

(Tryon, 1982, p. 425). First, baseline stability of each modified approach was established across three consecutive treatment sessions through determining whether a trend occurred during baseline measurements. The critical value (*z*) for use with the *C* statistic at the .05 level (*p*) of significance was established as 1.64; therefore, a value (*z*) below 1.64 is not significant; conversely, a value (*z*) above 1.64 is significant (Tryon, 1982).

Since within-treatment probes were introduced during Mayer et al.'s (2014) study, the M-STIM scoring system was applied to the raw data of KC's performance, starting from the first session using within-treatment probes (September 2015), in order to obtain initial baseline measurements. Using preliminary *C*-statistic analyses, no significant trend was found for any approach during Mayer et al.'s baseline sessions (x3), and analyses provided the following statistics for each approach: Increased melodic uniqueness: C=0.45, z=0.569, p=0.742; increased salience: C=0.475, z=0.601, p=0.742; increased multimodal stimulation: C=0.3825, z=0.484, p=0.7088). The last three sessions of Mayer et al.'s study were also analyzed using the *C* statistic to determine whether a trend existed immediately preceding M-STIM. *C* statistic analyses for these sessions also revealed no significant trend for any approach and provided the following statistics: Increased melodic uniqueness: C=0.605, z=0.765, p=0.222076; increased salience: C=0.455, z=*0.576*, p=0.282308; increased multimodal stimulation: C=0.798, z=1.009, p=0.156559.

Probe data

Since we did not find any significant trends in either baseline measurement, we were able to calculate the C statistic for each modification introduced in the M-STIM Phase treatment sessions (Tryon, 1982). The trends of the M-STIM treatment sessions (x8) plus the three baseline sessions immediately preceding the M-STIM Phase yielded statistically significant results, as follows: increased melodic uniqueness: C=3.655, z=11.105, $p < 0.00001$; increased salience: C=5.05, z=15.343, $p < 0.00001$; increased multimodal stimulation: C=4.665, z=14.1429, $p < 0.00001$. A significant trend was also noted for combined Mayer et al. (2014) and M-STIM treatment sessions across all approaches, as follows: Increased melodic uniqueness: C=4.795, z=18.045, $p < 0.00001$; increased salience: C=6.165, z=23.204, $p < 0.00001$; increased multimodal stimulation: C=5.703, z=21.463, $p < 0.00001$. See Figures 11, 12, and 13 for a visual representation of the slope of all three modifications.

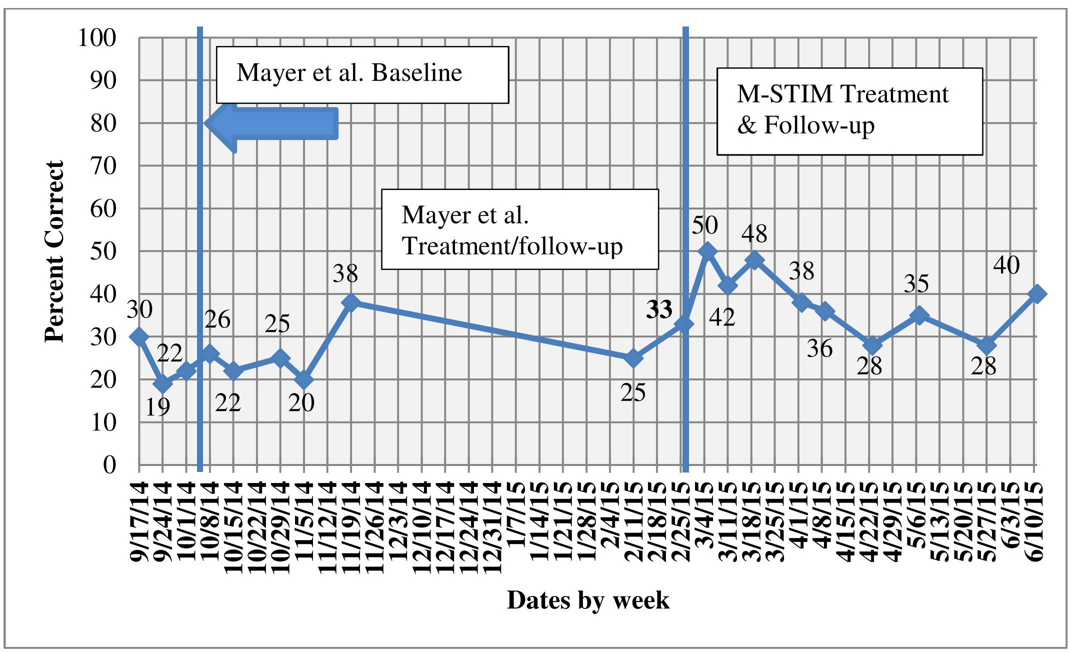


Figure 11. Slope of within-treatment probe scores for increased melodic uniqueness phrases from Mayer et al.'s baseline to M-STIM follow-up

Notes: Gridlines on the horizontal axis depict weekly increments to demonstrate course of treatment and the amount of time between treatment sessions.

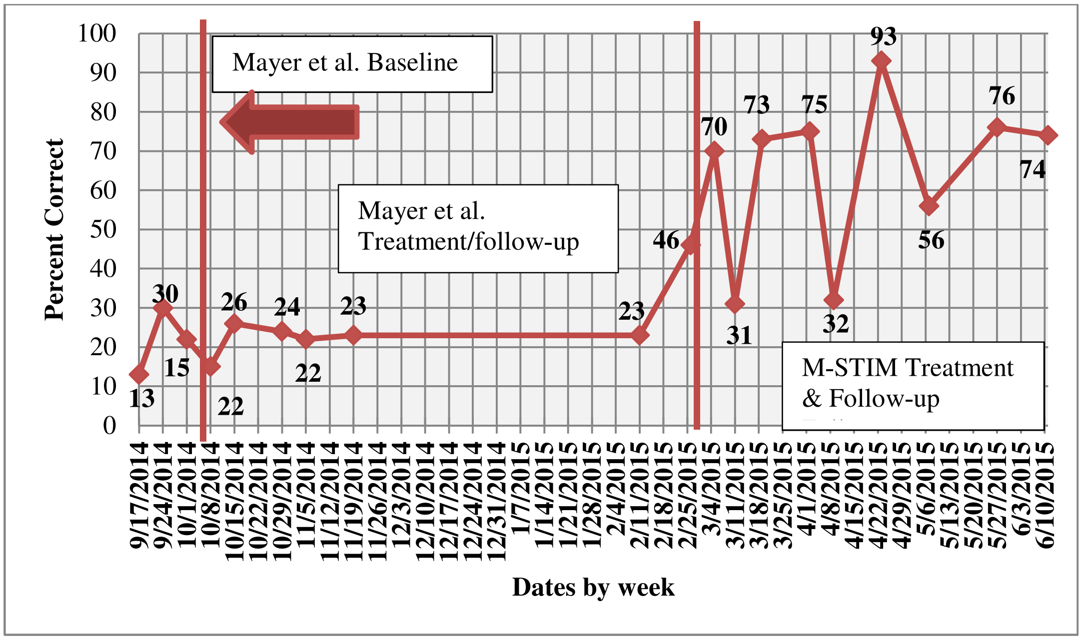


Figure 12. Slope of within-treatment probes for increased salience phrases from Mayer et al. (2014) baseline to M-STIM follow-up.

Notes: Gridlines on the horizontal axis depict weekly increments to demonstrate course of treatment and the amount of time between treatment sessions.

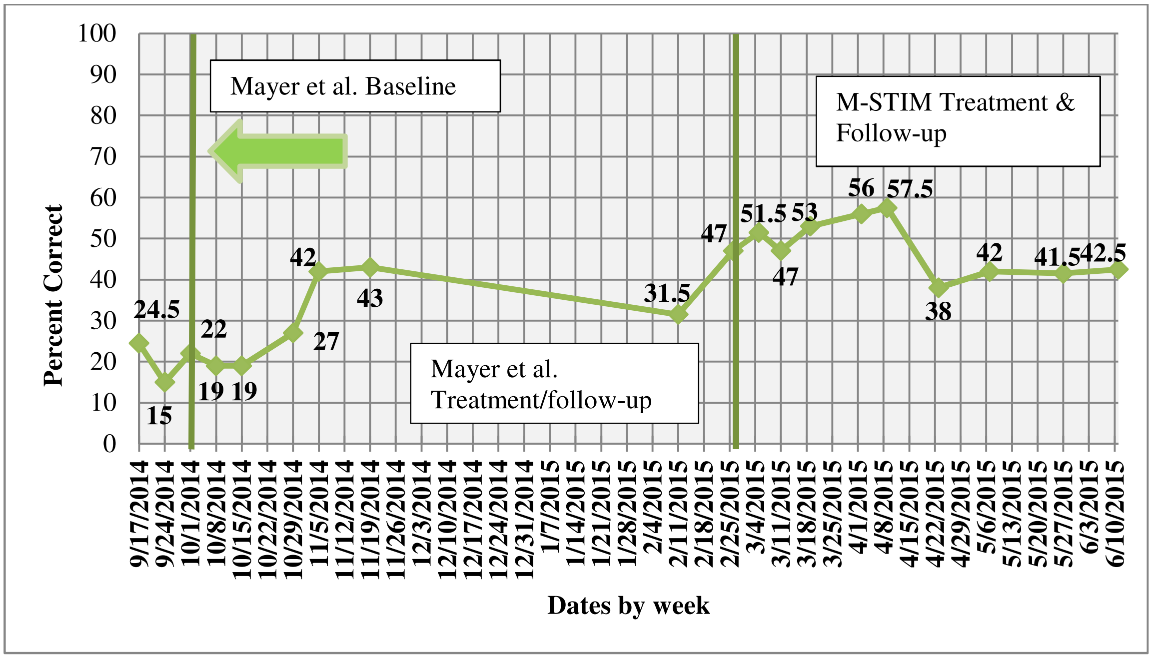


Figure 13. Slope of within-treatment probe scores for multimodal stimulation phrases from Mayer et al. (2014) baseline to M-STIM follow-up

Notes: Gridlines on the horizontal axis depict weekly increments to demonstrate course of treatment and the amount of time between treatment sessions.

Both Mayer et al.'s (2014) and M-STIM Phases' *C* statistical analyses exposed significance in the upward trend from baseline to treatment, showing a positive progression during intervention. An important question, however, was the magnitude of this change; in addition, the significant *C* statistic results lead the authors to ask which of the three phrase modifications – increased melodic uniqueness, increased salience, or multi-modal stimulation – was driving these changes, or how the contributions of each modification could be quantified over time. For this purpose, within-treatment data were further examined by calculating effect size (Beeson & Robey, 2006).

Effect size (*d*) was calculated for each Mayer et al. (2014)-versus-M-STIM trained phrase modification (see above) by dividing the difference between the Mayer et al. average (eight data points) and the M-STIM average (eight data points) by the standard deviation of the Mayer et al. study, resulting in the following values of *d*: Increased melodic uniqueness=2.2; increased salience=6.9, increased multimodal stimulation=2.1. According to the effect size norms derived from single-subject aphasia research as detailed in Beeson and Robey (2006), the effect sizes of increased melodic uniqueness and increased multimodal stimulation were less than a "small-sized" effect. The effect size of 6.9, however, for increased salience demonstrated a "large-sized" effect.

It should be noted that KC's performance on the treatment stimuli from Mayer et al.'s (2014) study, that were subsequently modified in the M-STIM Phase, did not appear to change significantly over the duration of Mayer et al.'s: That is, effect sizes *(d*) from Mayer et al.'s baseline (3 data points) and Mayer et al.'s follow-up (1 data point) were as follows: Increased melodic uniqueness=0.18; increased salience=0.12, increased multimodal stimulation=2.2. On the other hand, comparisons between the M-STIM Phase baseline (3 data points) and the M-STIM Phase follow-up (2 data points) revealed the following effect sizes (*d*): increased melodic uniqueness=1.3; increased salience=88.3, increased multimodal stimulation=0.58. Thus, the only significant effect size occurred in the increased salience approach with a greater than "large" effect size of 88.3.

To further inspect the magnitude of change during the M-STIM Phase, the multimodal stimulation category was subdivided into: 1) physical objects, 2) tactile cues, and 3) YouTube Videos; and the effect size for each category was calculated using the M-STIM baseline and follow-up data points. Results of these calculations revealed the following effect sizes (*d*): physical objects=2.6; tactile cues=0.13; YouTube videos=1.4. Whereas the multimodal stimulation category as a whole did not demonstrate significant effect size, the subcategory of physical objects (i.e., presenting a 3-D object, such as a toothbrush or coffee mug, in front of KC during within-treatment probes) demonstrated a small but significant effect size (Beeson & Robey, 2006).

When effect size was measured between Mayer et al.'s (2014) baseline for phrases subsequently modified in the M-STIM Phase (3 data points) and the M-STIM Phase follow-up, an increase in the magnitude of change (*d*) was noted as follows: increased melodic uniqueness= 2.8 (small effect size); increased salience=6.1 (large effect size); and increased multimodal stimulation= 4.5 (medium effect size).

The differences in effect size among the modified approaches suggest that just one of the three modified approaches - increased salience - may have been most effective in improving KC's functional performance on trained phrases, compared to the others (melodic uniqueness, multimodal stimulation). Visual inspection (see Table 8) was consistent with these data; the increased salience approach also showed the greatest positive slope (albeit large inter-session variability in performance) compared to the other modifications.

Functional Use of Trained Phrases Performance: Caregiver Questionnaire

Two caregivers, KC's husband and her mother, completed the Caregiver Questionnaire (see Appendix A). Their responses for both frequency of use (ratings from -2 to +2, see Figure 5) were combined, and a numerical value was substituted for their ratings of appropriateness of use (*yes*=2, *sometimes*=1, and *no*=0) to create a total range of rated responses for each phrase from -4 to +8, with a maximum use and appropriateness rating being an 8 and the least being a -4.

Combined ratings from both caregivers (Figure 14) revealed that KC used 56% of trained phrases in activities of daily living outside of therapy more frequently and appropriately post-treatment. An additional 16% of phrases were rated as being used more frequently (but not necessarily in an appropriate context) following the M-STIM treatment protocol.

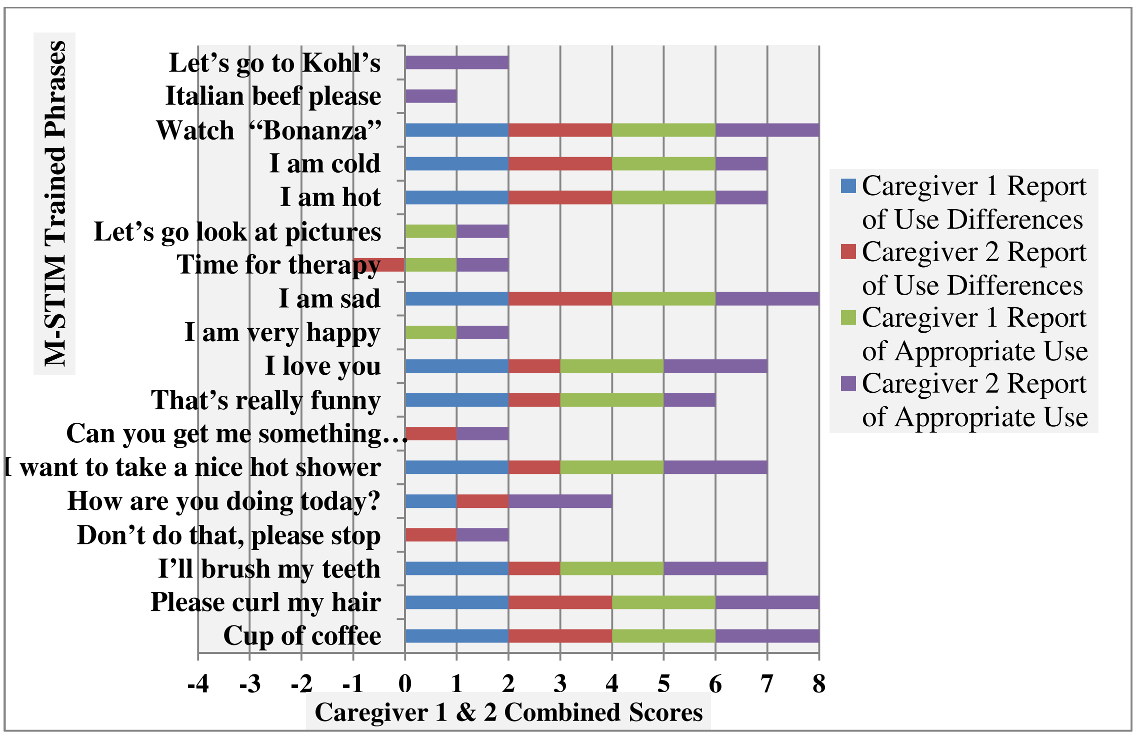


Figure 14. Results of Caregiver Questionnaire showing ratings of frequency and appropriate use of trained phrases in everyday life

Notes: This stacked figure displays both caregivers' responses in order to demonstrate increased functional and appropriate use of trained phrases in everyday living, as well as compare/contrast caregiver responses to obtain the most valid results.

M-STIM Scoring System Performance

Comparisons of the M-STIM Scoring System to the scoring system utilized in Level III, Step 3 of traditional MIT from Sparks (2008) (see Table 5) were conducted to address Prediction 4 (i.e., that the M-STIM scoring system would more accurately capture clinically significant progress over time compared to the traditional MIT scoring system). Both the M-STIM scoring system and the MIT scoring system were applied to the raw data from six probed phrases at 31 observation points during both phases, which encompassed two phrases from each of the three M-STIM Phase modifications (increased melodic uniqueness, increased salience, and multimodal stimulation).

At each observation point, the six probed phrase scores from both scoring systems were converted to percent correct and then averaged in order to equitably compare the two systems (see Figure 15). Calculations of the mean percent correct of all 31 observation points revealed a mean of 43.69% for the M-STIM scoring system and a mean of 23.66% for the MIT scoring system: i.e., a 20.03% difference.

To further explore these differences, we examined standard deviation of mean scores from each scoring system, with results as follows: M-STIM system SD = 20.81%; MIT scoring system SD = 24.92% (i.e., a 4.11% difference). This was consistent with the observation that utilizing the MIT scoring system to quantify KC's performance resulted in a wider range of scores, from 0-66.67%, while the M-STIM scoring system resulted in a score range from 12.83-70.5%.

That is, because the original MIT scoring system reflects a restricted raw score range (i.e., 0-2; see Table 6), it engendered more fluctuation of scores, with less gradation of performance outcomes, consistent with the increased SD and range of scores noted from this system compared to the M-STIM scoring system (recall that the latter was designed specifically to capture small differences in KC's performance over time, given the severity of her deficits). The higher average percent correct KC obtained when her performance was quantified using the M-STIM scoring system, therefore, was consistent with its intent of providing more opportunities to demonstrate increased performance of trained phrase responses. As seen in Figure 15, however, the close match in slope between the two scoring systems demonstrates concurrent validity of the M-STIM system compared to more traditional MIT scoring methods.

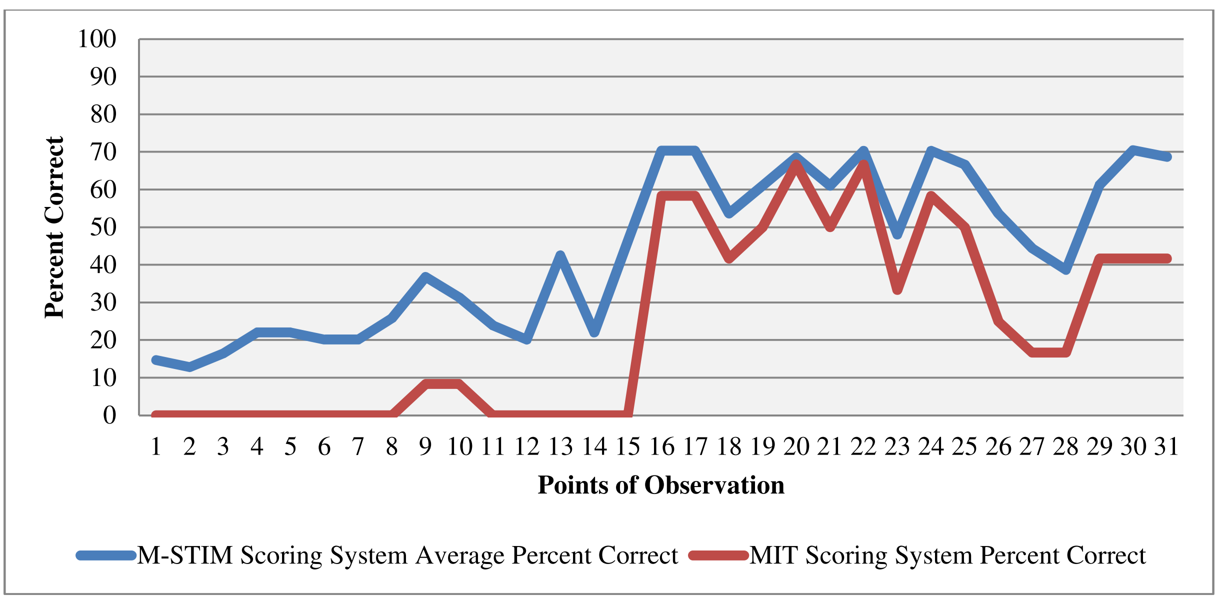


Figure 15. Scoring system comparisons for 31 observation points during Mayer et al.'s (2014) study and M-STIM Phase

Cognitive Testing: Test of Nonverbal Intelligence-3

Although we did not predict treatment-related changes in cognitive function given the language-specific nature of both phases of our treatment approach, scores on the Test of Nonverbal Intelligence: Third Edition (TONI-3; Brown, Sherbenou, & Johnsen, 1997), a norm-referenced nonlinguistic problem solving ability assessment, increased significantly from pre-Mayer et al. (2014) testing (January 2014) to post-M-STIM Phase testing (June 2015) with a change in percentile rank from 3rd to 26th (see Table 3). While these scores cannot be directly attributed to KC's intervention, it does demonstrate a positive increase in her performance that may have been related to or affected increases in other areas.

Discussion

The purpose of the current study (M-STIM) was to increase the treatment efficacy of Mayer et al.'s (2014) protocol for global aphasia by focusing on receptive language skills, in addition to the verbal expression skills that were the focus of Mayer et al.'s study, utilizing the three M-STIM strategies: 1) increased melodic uniqueness, 2) increased salience, and 3) multimodal stimulation. As mentioned previously, Mayer et al. (2014) found that phrases with increased melodic uniqueness (i.e., Modified Melodic Intonation Therapy, MMIT) were learned more quickly and had more durability over time compared to the two-toned intonation pattern of MIT; despite this success, persistent deficits in auditory comprehension, consistent with severe/global aphasia, appeared to prevent the research subject, KC, from generalizing all of the trained phrases to new contexts. The premise of M-STIM was to utilize these additional strategies with the hopes that they would lead to increased generalization outcomes of trained targets; therefore, improving KC's ability to transfer performance outcomes to learned behaviors. In addition, the M-STIM Phase took into consideration specific generalization variables, and their implications to generalization in terms of KC's performance will be further discussed.

Furthermore, following Hurkmans et al.'s (2012) recommendations for inclusion criteria in studies, the research questions and predictions of the current study were designed to address some of the holes found in the aphasia literature, especially in terms of reporting standardized pre- and post-testing measures and functional outcomes. In the current study, four outcome measures were utilized to explore statistical and functional implications of utilizing M-STIM to augment MMIT for global aphasia: 1) standardized, pre- to post-testing scores for both auditory verbal comprehension and expressive language (i.e., Western Aphasia Battery; Kertesz, 2007), 2) trends and magnitude of changes for the three M-STIM Phase modifications (increased melodic uniqueness, increased salience, and multimodal stimulation), as measured by within-treatment probes, 3) scoring comparisons of the traditional MIT scoring system and the M-STIM scoring system to explore gradation of scoring as a way to document smaller increments of change in performance outcomes, and 4) frequency and appropriateness of trained phrase uses in varied contexts and environments outside of therapy, as qualitatively measured by two caregiver questionnaires.

One of our main goals of M-STIM was to discover the most effective strategies for increasing KC's auditory comprehension, with the hopes that these would lead to increased verbal expression and carryover of trained phrases to everyday life situations. Fulfilling Research Question #1, Prediction 1, KC's auditory verbal comprehension scores on the WAB remained stable for almost an entire year (from the initial assessment in January 2014 to the post-Mayer et al. (2014) testing in September 2014) and then increased following the M-STIM treatment phase, giving us reason to speculate that her increased score was a direct result of the M-STIM intervention protocol. Even more promising was that her auditory verbal comprehension score continued to increase during testing conducted six-months post-M-STIM, suggesting the possibility that the tools we used and skills we trained during treatment may have had some generalizing effects. It is also notable that these increased auditory comprehension scores were achieved over a very short period (eight treatment sessions). These data also suggest the importance of focusing separately on comprehension versus the more mirror-neuron-focused approaches of increasing verbal expression (e.g., MIT) for language improvement and carryover.

In fulfillment of Research question #1, Prediction 2, KC's standardized expressive language scores improved from post-Mayer et al. (2014) to post-M-STIM. KC demonstrated increases in her expressive language scores, both in repetition and in naming/word finding, with the largest gains occurring in the naming/word finding subtest score. From KC's initial (pre-Mayer et al. treatment) assessment (January 2014) to post-Mayer et al. testing, this score increased by a mere 0.2 (i.e., from 1.1 to 1.3, out of 10). Following M-STIM, KC's subscore increased by 0.8 (i.e., from 1.3 to 2.1). Additionally, KC achieved this increased subtest score in a much shorter time, i.e., in eight sessions over two-and-a-half months, compared to 24 sessions over 10 months. Furthermore, although our primary goal of therapy was for KC to use spontaneously and functionally the specific phrases trained using M-STIM, her spontaneous speech score (i.e., expressing words/thoughts that were not trained during therapy) on the WAB did increase, especially from post-M-STIM to six month follow-up testing.

Visual inspection and statistical analyses of KC's performance during within-treatment probes for both direction and magnitude of change (to answer Research Question #2, Prediction 3, Parts (a) and (b)) presented mixed results, revealing that some M-STIM modifications demonstrated higher performance outcomes than others. Whereas *C* statistic analyses revealed an upward significant trend for all M-STIM modifications, visual inspection demonstrated a negative slope beginning with the last three to four M-STIM treatment sessions and continuing into follow-up sessions. While it is helpful to know that overall improvement was made during the course of treatment, it is important to ask why the upward trend did not continue throughout the full course of treatment. There are several possibilities that could account for this unexpected result; each is explored below.

First, the M-STIM scoring system was developed to track small units of progress; however, for KC, it appeared that for many probe responses, she demonstrated a bimodal trend; that is, she either scored a 9 (meaning she expressed the response immediately) or she required a phonemic cue (which lowered her score to 3), with few scores falling in between these two extremes. This variability in performance is most evident upon visual inspection of the "increased salience" data (see Figure 12). That is, although the greatest effect size was generated through this approach, KC fluctuated between performing an immediate response to the probed question and requiring a phonemic cue. Therefore, the decrease in her scores demonstrated the level of cueing necessary for KC to achieve an appropriate response.

Interestingly, an aphasia treatment protocol, called First Sound Practice (FSP), has been developed that focuses specifically on this aspect (the rehearsal of the first phoneme of a phrase) (Leonard, Rochon, & Laird, 2008). FSP has also been shown to increase word-finding ability of an entire phrase in a modified MIT protocol through the focused-practice of finding the initial phoneme (Chiou & Freuchte, 2015). Due to Chiou and Freuchte's report of preliminary success with applying FSP to a modified MIT protocol, it is plausible that applying the FSP technique while training phrases in the current study could have altered the bimodal trend in KC's performance and would be an important technique to explore in future research.

The fluctuations in the type and number of cues required for KC to produce a trained response, as well as her variable performance on standardized measures (e.g., KC's spontaneous speech score, see Table 3), led us to question the role of intra-individual variability. Hula and McNeil (2008) found that individuals post-stroke have innate variability in performance as a consequence of their brain injury. Intra-individual variability has been well-documented in traumatic brain injury literature and is thought to be linked to fluctuations in attention over time (Bleiberg, Garmoe, Halpern, Reeves, & Nadler, 1997; Stuss, Pogue, Buckle, & Bondar, 1994).

It is also possible that the lack of consistency with which KC attended therapy sessions attributed to the variations documented in her performance outcome measures. At the beginning of the M-STIM Phase, we saw KC once weekly for four consecutive weeks, during which time positive trends in probe scores were noted. Then, the next session was separated by two weeks with the following session separated by just one week, but for the rest of the course of treatment, the sessions were two to three weeks apart, due to cancelations and lack of caregiver availability for transportation. Thus, these negative probe trends, occurring at the point in which treatment became less consistent, suggest that for someone with severe deficits, like KC, consistency of treatment plays a significant role in positive outcomes and in maintaining carryover from one session to the next. It is noteworthy that the traditional MIT protocol (Sparks, 2008) calls for a high frequency of treatment sessions (five times per week).

In fact, many studies point to the intensity of treatment as an important factor in successful outcomes. According to Fox et al. (2002), the intensity of treatment has the largest effect on increasing neural plasticity, implying generalization of the target production. Fox et al. stated that several factors play a role in the intensiveness of therapy: frequency of treatment, number of repetitions, amount of effort, and accuracy of the target production. In KC's case, one can speculate that the frequency of treatment influenced the overall trajectory of her performance outcomes and contributed to the negative slope towards the end of the M-STIM protocol.

Statistical analyses for change in magnitude demonstrated similar results to the directional analyses in that the increased salience phrase modifications revealed the largest significant effect size compared to the other M-STIM modifications. It is not surprising that increased salience enhanced KC's performance outcomes of target productions, as it is well-supported in aphasia literature. Van der Meulen et al. (2014) stated that MIT has always, since its development, supported use of personally relevant treatment targets. However, by definition "personally relevant" varies from one individual to the next; and what clinicians may consider to fulfil that criterion may not actually do so. For example, one of KC's favorite past-times was watching television; thus the phrase "change the channel" appeared to be "personally relevant." Due to her comprehension deficits, however, "change the channel" did not have the level of personal meaning required for her to truly understand it. The phrase "watch Bonanza" was more specific than "change the channel" and likewise more personally relevant; as well as more concrete. Although increasing the salience or specificity of the phrase necessarily limited KC to utilizing it in only a single context, or in this case, for a single TV show (in theory), this drawback was counterbalanced by KC's immediate ability to comprehend the treatment target in this case.

While the multimodal stimulation component of M-STIM did not show a significant effect, one of the three subcategories, physical objects, did demonstrate a small, but significant, effect size. It is well known that the presence of physical objects can provide clarity for an individual with aphasia and concomitant comprehension deficits (e.g., Kertesz, 2007), due to the decreased level of abstraction required for processing (Hoff, 2014; Paul & Norbury, 2012).

The melodic uniqueness component of M-STIM also did not show a significant effect size or positive slope. One possibility is that since the Mayer et al. (2014) study began with some phrases containing unique melodies, M-STIM's introduction of more unique melodies did not create the same change in effect since a melodic component was already present during the MMIT Phase.

Relatedly, Thompson et al. (2003) showed that teaching more syntactically complex sentence structures to two individuals with Broca's aphasia and agrammatism led to generalization of behavioral gains to simpler sentence structures; whereas the two individuals in the control group, who were trained using less complex syntactic structures, did not demonstrate generalization to more complex structures. Based on Thompson et al.'s predictions, one could speculate a causal relationship between increasing melodic uniqueness of treatment targets (which ultimately resulted in more complex grammatical structures) and its potentially generalizing effects on less complex productions. For example, KC's ability to repeat nine-syllable phrases, such as “I want to take a nice hot shower,” could have further expanded her ability to repeat other less complex productions.

Furthermore, KC’s increased score on the WAB repetition subtest from post-MMIT to post-M-STIM testing demonstrated carryover of increased repetition skills to a new context, suggesting generalization of the ability to repeat. One question remains: What underlying mechanism(s) facilitated this? Considering that KC also demonstrated marked improvements on the TONI-3 (from her initial pre-MMIT assessment to post-M-STIM assessment), one can postulate that in order for KC to have improved her performance on the cognitive problem solving tasks, she may also have had an increased capacity to attend to the tasks, process the visual stimuli, and temporarily store a significant amount of information in order to correctly solve the problems. This finding suggests that training more complex material not only resulted in carryover to less complex structures, but that it also may have expanded cognitive functioning through increasing focused attention and working memory capacity However, while KC demonstrated increased performance on the TONI-3 during post-M-STIM testing, it is plausible that her attention continued to fluctuate from session to session due to intra- individual variability, as seen on the variability of within-treatment probes (Bleiberg, Garmoe, Halpern, Reeves, & Nadler, 1997; Stuss, Pogue, Buckle, & Bondar, 1994).

In support of Research Question #3, Prediction 4, the M-STIM scoring system more accurately depicted KC’s performance during within-treatment probes by capturing smaller units of clinically significant progress over time, compared to the traditional MIT scoring system. The increased standard deviation and range of KC’s performance, as measured by the MIT scoring system, demonstrated a greater fluctuation of scores, which related to the raw score range of 0-2, with less gradation of performance outcomes. The higher average percent correct of the M-STIM scoring system demonstrated that it provided more opportunities for KC to show increased performance of trained phrase responses over time. Visual representation (Figure 15) of the two scoring systems also demonstrated these differences and provided further evidence in support of Prediction 4.

Overall, the M-STIM scoring system provided increased specificity in terms of level of performance and displayed increased intricacy of raw score possibilities (from 0-9) of performance outcomes. This specificity and increased gradation of performance outcomes in the M-STIM scoring system are best demonstrated at observation points 8 through 15, wherein KC's percent correct varied between 25% and 35% using the M-STIM scoring system and varied between 0% and 9% with the MIT scoring system.

When using the traditional MIT scoring system, it is likely that a therapist would discontinue this course of therapy after the first eight sessions because the scoring system showed no progress, with a range of performance outcomes from 0 to 0%. The M-STIM scoring system, however, showed a range from 13 to 22% (a 9% increase in performance), which, while still a small increase, displayed progress and further revealed the ability of the M-STIM scoring system to identify smaller but nonetheless functionally significant gains compared to the MIT system. Scoring systems based on very small units of change are especially needed for individuals like KC, who have more severe deficits requiring iterative approaches and for whom very small units of progress may occur over time.

In support of Research Question #4, Prediction 5, KC's two caregivers reported functional, positive changes in the frequency and appropriateness of trained phrases utilized outside of therapy. These qualitative measures importantly suggest that 56% of trained phrases generalized, supporting positive treatment-induced gains.

Our goal for KC in terms of generalization was that the training we provided would produce stimulus generalization, meaning that she would apply the trained phrases to multiple contexts/ environments outside of the therapy room as a result of intervention. Response generalization, meaning expanding trained words/phrases to other words/phrases of a similar category (for example, a response generalization of the phrase "Watch Bonanza" to "Watch the Andy Griffith show") was not expected due to KC's severity (Merret, Peretz, & Wilson, 2014).

Potentially, many of the above-mentioned factors played a role in the generalization (or lack thereof) of KC's trained phrases to new environments. A possible barrier to generalization is the presence of severe cognitive deficits, which may preclude developing an internal cueing system (Fox et al., 2006). Fox et al. stated that when such deficits are present, external cues are often needed to produce target behaviors and reliance is placed on the caregiver to provide those cues. During KC's initial assessment, she placed in the 3rd percentile rank on the TONI-3, which is several standard deviations below the lowest average range, suggesting severe cognitive deficits. While KC's scores improved significantly during the M-STIM Phase post-testing (from a percentile rank of 3 to 26), 74% of the individuals her age in the norming data performed the same or better than she did, which suggests that persistent cognitive deficits could have decreased her ability to generalize treatment gains; therefore, requiring external cues to produce trained responses.

While KC's cognitive skills may not have increased enough to aid in generalizing speech beyond some of the trained responses, her increases in cognitive skills, as demonstrated by the TONI-3, further support existing literature showing increased attention and working memory as a result of music-based intervention. For example, Hurkmans et al. (2015) reported that one of their participants demonstrated increases in both attention and working memory following Speech-Music Therapy for Aphasia (SMTA), and Ren da Fontoura et al.'s (2014) likewise noted that their single participant in a modified MIT protocol demonstrated increases in working memory, executive function, and attention following intervention. In KC's case, her increased cognitive skills may have developed through the increased focused attention and working memory required to produce lengthy MMIT stimuli. In addition, through caregiver reports, KC made gains in participation in activities of daily living, including cooking TV dinners in the microwave and showering independently, which also signaled improved cognitive skills.

Study Limitations

Blank et al.'s (2002) findings demonstrated the increased complexity of brain activity for producing propositional speech, compared to more automated speech tasks, such as those trained in MIT. For KC, propositional speech was unlikely to be recovered due to the severity of her injury; hence, we attempted to automatize trained, functional phrases for application to everyday situations, i.e., facilitating stimulus generalization.

However, the current literature is unclear as to how to make the transition in therapy from non-propositional to propositional speech training for those, like KC, who appear to rely on rhythmic and/or singing components of a treatment protocol, like MIT, to produce words/phrases. Stahl, Hensler, Turner, Geyer, and Kotz (2013) found that in a 15-participant study in which participants were equally divided into three treatment groups (singing therapy, rhythmic therapy, or standard speech therapy), the participants assigned to both the singing and rhythmic therapies experienced more improvements in formulaic speech, whereas the participants in the standard speech group demonstrated more improvements in non-formulaic (i.e., propositional) speech. In fact, the singing and rhythmic groups demonstrated no significant progress in productions of non-formulaic speech, which further supported the theory that singing and rhythm therapies may lead to automatized/formulaic responses but do not directly result in changes in non-formulaic/ propositional speech.

Many factors had the potential to affect the outcomes of this study, but it appears that frequency and consistency of treatment, along with individual participant factors, may have had the most influence on study outcomes. Individual participant factors that may have limited treatment outcomes included KC's severity, level of education, internal motivation, level of family support, and her home environment. Any or all of these factors may have played a role in her ability to maintain and continue to make positive gains in outcome measures. These factors may also have affected her ability to carryover the trained phrases to everyday life, as mentioned in detail above.

Lack of consistency and sub-optimal frequency of KC's therapy attendance may have also influenced outcomes. Unfortunately, KC relied on her caregivers for transportation and the therapy clinic was far away from her home. An alternative to aid in carryover in the absence of frequent therapy sessions could be a home practice program to provide increased stimulation of the trained phrases on a more frequent basis. A home practice program was developed for KC.

Initially, during Mayer et al.'s studies, all of the trained phrases were recorded on her iPhone so that she could play each phrase back as many times as she wanted. After several sessions, the clinicians realized that KC was not able to operate her phone without the aid of a caregiver. Due to her living situation, she was home by herself much of the day and did not have someone there to assist her in the home practice program. In the M-STIM Phase, this issue was addressed by recording all of the trained phrases on one Voice Memo so that she could learn to press one button and listen to them all in a row. We asked KC to bring her device to the clinic to practice how to access the Voice Memo but she forgot it several times and to our knowledge, she practiced the phrases at home infrequently, just as in the previous Mayer et al. study.

Coppens and Patterson (2015) stated that in order for a home practice program to be effective, "it must contain well-designed, functional applications and be carefully monitored" (p.3). Coppens and Patterson also suggested having clients sign a home practice agreement, which helps to stress the importance of increasing the frequency of practice and increases the likelihood that the client will complete his/her home practice. Chiou and Freuchte (2015) also demonstrated the benefits of home practice in aphasia treatment, specifically in a modified Melodic Intonation Therapy protocol. They found that the implementation of a home program in which the caregiver facilitated the client by providing multiple prompts to a trained two-syllable target response led to increased accuracy and consistency effects of target productions across time.

Another therapy protocol for treatment of neurogenic disorders, whose success relies heavily on high compliance and home practice, is Lee Silverman Voice Treatment (LSVT). Literature regarding the effectiveness of LSVT indicated generalization occurring through a high frequency of treatment (four times a week for one month), along with an intensive home practice program (the participant completed homework once daily on therapy days and twice daily on non-therapy days). This is considered so integral to progress that the developers of LSVT also stated that if a client does not complete the home practice program, it can be cause for discharge (Fox et al., 2006; Narayana et al., 2010; Sapir, Ramig, & Fox 2011).

Since the MIT protocol (Sparks, 2008) recommended a high intensity and frequency of treatment (twice daily for 5 times per week), it can be hypothesized that KC would have also benefited from this treatment dosage, along with an intensive home program. However, due to the nature of healthcare delivery in the United States, it would appear unlikely that anyone in the chronic stage of recover, like KC, would have access to the intensity and frequency of treatment necessitated in the MIT protocol. Therefore, in future studies, it is hypothetically more probable and realistic to focus on an intensive home program to provide the intensity and frequency of exposure to trained stimuli necessary to produce a highly beneficial therapeutic effect at the impairment, activities, and participation level.

A final shortcoming related to the lack of experimental controls with which to compare KC's performance on within-treatment probes. It would have been beneficial to have a set of experimental control phrases from which to confirm that any possible treatment effect directly related to the trained items (Beeson & Robey, 2006).

Recommendations for Future Studies

Since increased salience demonstrated the largest effect size, its application in treatment of global aphasia should be further explored in future studies. Although increasing the salience, or specificity of a phrase, necessarily limits its application, using salience as a means to expand application of its use could be a point at which to start intervention for individuals with severe global deficits, such as KC. This is consistent with treatment protocols designed for individuals with severe apraxia as well. Modifications of two treatment approaches for severe apraxia, Integral Stimulation (Holtzapple & Marshall, 1977) and Sound Production Treatment Heirarchy (Wambaugh, Martinez, McNeil, & Rogers, 2010), focus on sequentially and systematically training a core set of functional words or phrases used in daily communication, under the premise that generalization to other aspects of language is not likely to occur. Whereas treatment for severe neurogenic disorders may be limited in terms of formulaic approaches, exploration of expanding comprehension of a target word to new contexts would be beneficial to future studies. For example, intervention could begin with a salient target and then expand the context in which it is used with a systematic approach focused on increasing application and generalization of the target word or phrase.

One such clinical tool that already exists in the literature aims at expanding school-aged children's vocabulary through a strategy called Word Chaining (Paul & Norbury, 2012). Word Chaining focused on the development of contextual understandings of words, i.e., meaningful interpretations of words in order to facilitate application to a variety of contexts (Spencer & Guillaume, 2006). For example, with the goal of expanding KC's use of a target word to multiple contexts, we could use the phrase "Watch Bonanza," not as a final target phrase, but as a "chain" or bridge to increase comprehension of the word "watch." For example, through initially pairing the word "watch" with "Bonanza," we would start with a known context for use of that word and then gradually expand the context by training "Watch the Andy Griffith show."

To generalize the word “watch” to a broader, but familiar, context we could train “Watch television.” Then, we could further expand the meaning of the word “watch” to a new context, such as “Watch the stove” or “Watch out!” In this way, it is possible that treating expressive language in severe, global aphasia may not necessarily be limited to a set of formulaic phrases; thereby, breaking the divide between propositional and formulaic speech training to increase the potential for generalization of trained productions to new contexts and environments.

Another area future research should explore is training of cueing strategies (e.g., directly teaching KC to self-cue for the first phoneme of a phrase, as in FSP), to increase her independence in verbal output. Perhaps focusing on internalizing such a cueing strategy would augment carryover of the trained target to novel words and phrases.

Finally, what remains unanswered in the aphasia literature is whether and how protocols could successfully transition from singing and/or rhythmic speech components to non-formulaic/ propositional speech for individuals with severe aphasia (Al-Janabi, et al., 2014; Blank, Scott, Murphy, Warburton, & Wise, 2002). Thus far, the research does not definitively support that training non-propositional speech (e.g., MIT) leads to such outcomes.

5. SUMMARY AND CONCLUSIONS

We developed a novel treatment protocol, M-STIM, to augment verbal expression for individuals with severe aphasia. The data collected and analyzed through a variety of outcome measures provided meaningful answers to the four research questions and five predictions developed for the continuation of this Phase 1 study (following Robey's [2004] Five-Phase Model for Outcome Research) and supported the practical, theoretical and statistical significance of using a modified MIT protocol (MMIT), alongside strategies to increase auditory comprehension (M-STIM), for an individual (KC) with severe, global aphasia.

It is noteworthy that M-STIM was implemented in conjunction with a scoring system designed to provide clinicians with a tool to more accurately measure smaller units of progress over time. This tool is especially helpful for individuals, like KC, whose severity influences the ability to make the same increments of improvement that someone with less severe aphasia may make.

Our findings also served to reiterate the importance of treatment intensity (in this case, frequency and consistency) for carryover of trained targets from one session to the next. In all three M-STIM modifications, KC's performance either decreased or stayed relatively the same when gaps between treatment sessions were two weeks or more, potentially affecting her overall performance outcomes during the course of therapy.

Surprisingly, despite negative trends in performance outcomes of two of the three M-STIM modifications in the latter two-thirds of treatment, KC demonstrated increases in both expressive and receptive language abilities, as measured by the WAB. Caregivers also noted appropriate and more frequent use of over half of the trained words/ phrases to everyday life situations, suggesting generalization of some of these trained skills.

In sum, M-STIM, in continuation of a Phase 1 study following Robey's (2004) Five-Phase Model for Outcome Research, achieved Robey's Phase 1 goal of successfully measuring therapeutic effect through standardized, statistical, and clinical measures. These positive therapeutic effects, most notably in the M-STIM -increased salience- subcategory, should be further explored in a Phase 2 study to measure potential treatment efficacy for individuals with severe, global aphasia. Results of the M-STIM protocol suggest that this next phase should focus on expanding target utterances to new contexts through Word Chaining (Paul & Norbury, 2012; Spencer & Guillaume, 2006) using the increased salience modification of M-STIM, alongside self-cueing strategies, such as the First Sound Production (FSP) strategy, for training the first phoneme of the target word or phrase(Leonard, Rochon, & Laird, 2008).

REFERENCES

Albert, M. L., Sparks, R., & Helm, N. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology, 29*, 130-131.

Alexander, M., & Loverso, F. (1992). A specific treatment for global aphasia. *Clinical Aphasiology, 21*, 277-289.

Al-Janabi, S., Nickels, L., Sowman, P., Burianovà, H., Merrett, D., & Thompson, W. (2014). Augmenting melodic intonation therapy with non-invasive brain stimulation to treat impaired left-hemisphere function: two case studies. *Frontiers in Psychology, 5, 37*, doi: 10.3389/fpsyg.2014.00037.

Baker, F. (2000). Modifying the melodic intonation therapy program for adults with severe non-fluent aphasia. *Music Therapy, 18*, 110-114.

Beeson, P., & Robey, R. (2006). Evaluating single-subject treatment research: Lessons learned from aphasia literature. *Neuropsychology Review, 16*, 161-169.

Belin, P., Eckhout, P., Zilbovicius, M., Remy, P., Francois, C., & Guillaume, S. (1996). Recovery from nonfluent aphasia after melodic intonation therapy: A PET study. *Neurology, 47*, 1504-1511.

Benson, F. D.-E. (1994). Assessment: Melodic intonation therapy. *Neurology, 44*, 566-568.

Blank, S., Scott, S., Murphy, K., Warburton, E., & Wise, R. (2002). Speech production: Wernicke, Broca, and beyond. *Brain, 125*, 1829-1838.

Bleiberg, J., Garmoe, W., Halpern, E., Reeves, D., & Nadler, J. (1997). Consistency of within-day and across-day performance after mild brain injury. *Cognitive and Behavioral Neurology, 10(4)*, 247-253.

Blomert, L. (1992). *The Amsterdam -- Nijmegen Everyday Language Test (ANELT).* Berlin: Springer-Verlag.

Blomert, L., Koster, C., & Kean, M. (1995). *Amsterdam-Nijmegen Everyday Language Test. Dutch version.* Lisse: Swets & Zeitlinger.

Breier, J., Randle, S., Maher, L., & Papanicolaou, A. (2010). Changes in maps of language activity activation following melodic intonation therapy using magnetoencephalography: Two case studies. *Journal of Clinical and Experimental Neuropsychology, 32(3)*, 309-314.

Brown, L., Sherbenou, R., & Johnsen, S. (1997). *Manual of Test of Nonverbal Intelligence: Third Edition.* Austin, TX: PRO-ED.

Chiou, H., & Freuchte, T. (2015, November). The power of homework practice: A modified Melodic Intonation Therapy in severe Broca's aphasia. *American Speech-Language Hearing Association Convention.* Denver, CO.

Conklyn, D., Novak, E., Boissy, A., Bethoux, F., & Chemali, K. (2012). The effects of modified melodic intonation therapy on confluent aphasia: A pilot study. *Journal of Speech, Language, and Hearing Research, 55*, 1463-1471.

Coppen, P., & Patterson, J. (2015, November). How to maximize generalization in aphasia rehabilitation. *American Speech-Language Hearing Association.* Denver, CO.

Cubelli, R., Trentini, P., & Montagna, C. (1991). Re-education of gestural communication in a case of chronic global aphasia and limb apraxia. *Cognitive Neuropsychology, 8(5)*, 369-380.

De la Plata, C. M., Hart, T., Hammond, F. M., Frol, A., Hudak, A., Harper, C. R., & ...Diaz-Arrastia, R. (2008). Impact of age on long-term recovery from traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 89(5)*, 896-903. http://doi.org/10.1016/j.apmr.2007.12.030.

Dietz, A., Thiessen, A., Griffith, J., Peterson, A., Sawyer, E., & McKelveys, M. (2013). The renegotiation of social roles in chronic aphasia: Finding a voice through AAC. *Aphasiology, 27(3)*, 309-325.

Feiken, J., & Jonkers, R. (2012). *Diagnostic Instrument for apraxia of speech.* Houten: Bohn, Stafleu en Van Loghum.

Fonseca, R., de Salles, J., & Parente, M. (2008). Development and content validity of the Brazilian Brief Neuropsychological Assessment Battery Neupsilin. *Psychology & Neuroscience, 1(1)*, 55.

Fox, C., Morrison, C., Ramig, L., & Sapir, S. (2002). Current perspectives on the Lee Silverman Voice Treatment (LSVT). *American Journal of Speech-Language Pathology, 11*, 111-123.

Fox, C., Ramig, L., Ciucci, M., Sapir, S., McFarland, D., & Farley, B. (2006). Science and practice of LSVT/LOUD: Neural plasticity-principled approach to treating individuals with Parkinson's disease and other neurological disorders. *Seminars in Speech and Language, 27*, 283-299.

Frattali, C., Thompson, C., Holland, A., Wohl, C., & Ferketic, M. (1995). The FACS of life ASHA facs--a functional outcome measure for adults. *Journal of the American Speech-Language Hearing Association, 37(4)*, 40-46.

Graetz, P., de Blesser, R., & Willmes, K. (1991). *Aachen Aphasia Test. Dutch version.* Lisse: Swets & Zeitlinger.

Hallgren, K. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8(1)*, 23-34.

Helm-Estabrooks, N., Albert, M., & Nicholas, M. (2014). Melodic Intonation Therapy. In N. Helm-Estabrooks, M. Albert, & M. Nicholas, *Manual of Aphasia and Aphasia Therapy* (pp. 279-292). Austin, TX: PRO-ED, INC.

Helm-Estabrooks, N., Fitzpatrick, P. M., & Barresi, B. (1982). Visual action therapy for global aphasia. *Journal of Speech and Hearing Disorders 47(4)*, 385-389.

Hoff, E. (2014). *Language Development, Fifth Edition.* Belmont, CA: Wadsworth, Cengage Learning.

Holtzapple, P., & Marshall, N. (1977). The application of multiphonemic articulation therapy with apraxic patients. *In Clinical Aphasiology Conference: Clinical Aphasiology Conference (1977 : 7th: Amelia Island, FL : May 17-20, 1977)* (pp. 46-58). Amelia Island FL: BRK Publishers.

Hough, M. (2010). Melodic intonation therapy and aphasia: Another variation on a theme. *Aphasiology, 24*, 775-786.

Hough, M., & Johnson, R. K. (2009). Use of AAC to enhance linguistic communication skills in an adult with chronic severe aphasia. *Aphasiology 23(7-8)*, 965-976.

Hula, W., McNeil, M., & Sung, J. (2007). Is there an impairment of language-specific attentional processing in aphasi? *Brain and Language, 103(1)*, 240-241.

Hurkmans, J., Bruijn, M., Boonstra, A., Jonkers, R., Bastiaanse, R., Arendzen, H., & Reinders-Messelink, H. (2012). Music in the treatment of neurological language and speech disorders: A systematic review. *Aphasiology, 26(1)*, 1-19.

Hurkmans, J., Jonkers, R., de Bruijn, M., Boonstra, A., Hartman, P., Arendzen, H., & Reinders-Messelink, H. (2015). The effectiveness of Speech-Music Therapy for Aphasia (SMTA) in five speakers with Apraxia of Speech and aphasia. *Aphasiology, 29(8)*, 939-964.

Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59(1)*, 12-19.

Johannsen-Horbach, H., Cegla, B., Mager, U., Schempp, B., & Wallesch, C. (1985). Treatment of chronic global aphasia with a nonverbal communication system. *Brain and Language, 24(1)*, 74-82.

Jungblut, M. (2009). SIPARI (registered trademark): A music therapy intervention for patients suffering with chronic, nonfluent aphasia. *Music and Medicine*, doi: 10.1177/1943862109345130.

Jungblut, M., & Aldridge, D. (2004). The music therapy intervention SIPARI (registered trademark) with chronic aphasics-research findings. *Neurological Rehabilitation, 10(2)*, 69-78.

Jungblut, M., Huber, W., Mais, C., & Schnitker, R. (2014). Paving the way for speech: Voice-training-induced plasticity in chronic aphasia and apraxia of speech--Three single cases. *Neural Plasticity*, 1-14.

Jungblut, M., Suchanek, M., & Gerhard, H. (2009). Long-term recovery from chronic global aphasia: A case report. *Music and Medicine, 1(1)*, 61-69.

Kalinyak-Fliszar, M., Kohen, F., & Martin, N. (2011). Remediation of language processing in aphasia: Improving activation and maintenance of linguistic representations in (verbal) short-term memory. *Aphasiology 25(10)*, 1095-1131.

Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology, 10(2)*, 159-180.

Kazdin, A. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford University.

Kertesz, A. (2007). *Western Aphasia Battery-Revised.* San Antonio, TX: Harcourt Assessment, Inc.

Lasker, J., & Garrett, K. L. (2006). Using the Multimodal Communication Screening Test for Persons with Aphasia (MCST-A) to guide the selection of alternative communication strategies for people with aphasia. *Aphasiology 20(2/3/4)*, 217-232.

Leonard, C., Rochon, E., & Laird, L. (2008). Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. *Aphasiology, 22(9)*, 923-947.

Manasco, M. (2014). *Introduction to Neurogenic Communication Disorders.* Burlington, MA: Jones & Bartlett Learning, LLC.

Mayer, J., Jeppson, T., & Henning, D. (2014, February). Modifying MIT for severe aphasia: More music, more language? *[Poster presented at:] Illinois Speech-Language-Hearing Association*.

Merret, D., Peretz, I., & Wilson, S. (2014). Neurobiological, cognitive, and emotional mechanisms in melodic intonation therapy. *Frontiers in Human Neuroscience, 8, 401*, doi: 10.3389/fnhum.2014.00401.

Narayana, S., Fox, P., Zhang, W., Franklin, C., Robin, D., Vogel, D., & Ramig, L. (2010). Neural correlates of efficacy of voice therapy in Parkinson's disease identified by performance-correlation analysis. *Human Brain Mapping, 31*, 222-236.

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy: How it is done and why it might work. *Annals of New York Academy of Sciences, 1169*, 431-436.

Paul, D., Frattali, C., Holland, A., Thompson, C., Caperton, C., & Slater, S. (2004). The American Speech-Language-Hearing Association Quality of Communication Life Scale (QCL): Manual. *American Speech-Language-Hearing Association*. Rockville, MD.

Paul, R., & Norbury, C. (2012). *Language Disorders from Infancy to Adolescence, Fourth Edition.* St. Louis, MO: Elsevier Mosby.

Peretz, I., Gagnon, L., Hebert, S., & Macoir, J. (2004). Singing in the brain: Insights from Cognitive Neuropsychology. *Music Perception, 21(3)*, 373-390.

Porch, B. (1967). *Porch Index of Communicative Ability: Theory and Development.* Palo Alto: Consulting Psychologists Press.

Ren da Fontoura, D., de Carvalho Rodrigues, J., Brandão, L., Monção, A., & Fumagalli de Salles, J. (2014). Efficacy of the adapted melodic intonation therapy: A case study of a Broca's aphasia patient. *Distúrbios Comun. São Paulo, 26(4)*, 641-655.

Robey, R. (2004). A five-phase model of clinical-outcome research. *Journal of Communication Disorders, 37*, 401-411.

Sapir, S., Ramig, L., & Fox, C. (2011). Intensive voice treatment in Parkinson's disease: Lee Silverman Voice Treatment (LSVT). *Expert Reviews in Neurotherapeutics, 11(6)*, 815-830.

Schlaug, G., Marchina, S., & Norton, A. (2008). From singing to speaking: Why patients with Broca's aphasia can sing and how that may lead to recovery of expressive langauge functions. *Music Perception, 25*, 315-323.

Schlaug, G., Marchina, S., & Norton, A. (2009). Evidence for plasticity in white-matter tracts of patients with chronic Broca's aphasia undergoing intense intonation-based speech therapy. *Annals of New York Academy of Sciences, 1169*, 385-394.

Sparks, R. (2008). Melodic Intonation Therapy. In R. Chapey, *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders, Fifth Edition* (pp. 703-717). Baltimore, MD: Lippincott Wilkins & Williams.

Sparks, R., Helm, N., & Albert, M. (1974). Aphasia rehabilitation resulting from melodic intonation therapy. *Cortex, 10*, 303-316.

Spencer, B., & Guillaume, A. (2006). Integrating curriculum through the learning cycle: content-based reading and vocabulary instruction. *The Reading Teacher, 60(3)*, 206-219.

Stahl, B., Henseler, I., Turner, R., Geyer, S., & Kotz, S. (2013). How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery. *Frontiers in Human Neuroscience*, doi: 10.3389/fnhum.2013.00035.

Stuss, D. T., Pogue, J., Buckle, L., & Bondar, J. (1994). Characterization of stability of performance in patients with traumatic brain injury: Variablity and consistency on reaction time tests. *Neuropsychology, 8(3)*, 316-324.

Thompson, C., Shapiro, L., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The Complexity of Treatment Efficacy (CATE). *Journal of Speech, Language, and Hearing Research: JSLHR, 46(3)*, 591-607.

Titze, I., & Verdolini Abbott, K. (2012). *Vocology: The Science and Practice of Voice Habilitation.* Salt Lake City, Utah: National Center for Voice and Speech.

Tryon, W. (1982). A simplified time-series analysis for evaluating treatment interventions. *Journal of Applied Behavioral Analysis, 15(3)*, 423-429.

van der Meulen, I., van de Sandt-Koenderman, W., Heijenbrok-Kal, M., Visch-Brink, E., & Ribbers, G. (2014). The efficacy and timing of Melodic Intonation Therapy in subacute aphasia. *Neurorehabilitation and Neural Repair, 28(6)*, 536-544.

van Eeckhout, P. (1982). *Sabadel: histoires insolites pour faire parler.* Paris: Medsi.

Wambaugh, J., Martinez, A., McNeil, M., & Rogers, M. (2010). Sound production treatment for apraxia of speech: Overgeneralization and maintenance effect. *Aphasiology, 13(9-11)*, 821-837.

Wan, C., Rüber, T., Hohmann, A., & Schlaug, G. (2010). The therapeutic effects of singing in neurological disorders. *Music Perception, 27(4)*, 287-295.

Wilson, S., Parsons, K., & Reutens, D. (2006). Preserved singing in aphasia: A case study of the efficacy of melodic intonation therapy. *Music Perception, 24(1)*, 23-26.

Yamaguchi, S., Akanuma, K., Hatayama, Y., Otera, M., & Meguro, K. (2012). Singing therapy can be effective for a patient with severe nonfluent aphasia. *International Journal of Rehabilitation Research, 35*, 78-81.

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

APPENDIX A

CAREGIVER QUESTIONNAIRE

Caregiver Questionnaire

1. Please describe any changes you've noticed in KC's communication since the start of treatment:
    a. Any changes in speech?
    b. New words or phrases?
    c. Any changes in understanding other's speech?
        i. To a familiar conversational partner?
        ii. To an unfamiliar conversational partner?

2. Did you notice any differences or changes after the first course of treatment (from February 2014-November 2014) compared to the second course of treatment (from March 2015-May 2015)?

3. Did you notice any changes in KC's independence in daily activities during or following treatment? For example, I know Gary mentioned that KC began to shower independently around January 2015 and started making her own meals in the microwave around the same time. Anything else?

4. Have you heard KC use any of words or phrases trained in therapy at home or while you're out and about (i.e., outside of sessions at NIU)? To answer this question, please complete the chart below by circling the best response only for the words/ phrases you've heard her say. If she never heard her say a specific word/phrase, leave it blank.

Word/ Phrase:	Frequency of Use:					Appropriateness of Use:		
If you've only heard a word or words from the phrase and not the whole phrase, please circle the word/s you heard within the phrase. For example, in the phrase "Let's go shopping," if KC said "shopping," then you would circle the word "shopping." If you never heard her say the word/phrase, leave it blank and do not fill out the "Frequency of Use" or "Appropriateness of Use" for that phrase.	CIRCLE the number that best corresponds with KC's use of each phrase before and after treatment. If you didn't hear her use the words at all, then leave that response blank **-2** **-1** **0** **+1** **+2** **-2**= uses these words a lot less since treatment **-1**=uses these words somewhat less since treatment **0**= uses these words about the same as before treatment **+1**= uses these words somewhat more since treatment **+2**= uses these words a lot more since treatment					If you heard KC use any of the words/phrases, were those words/ phrases used appropriately, i.e., did the word/phrase make sense in context? For example, if KC said, "I'll brush my teeth," was she brushing her teeth or going to brush her teeth? <u>CIRCLE:</u> Yes, No, or sometimes		
Let's go shopping	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Let's go to Kohls	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Time for therapy	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Change the channel	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Watch Bonanza	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Can we order food	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Italian Beef please	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Can we take a walk	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Let's go outside	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes

Can you get my shoes	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Cup of coffee	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I want to take a nice hot shower	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Have a great day Gary	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Let's go look at pictures	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I want that shirt	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Please curl my hair	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Good morning	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
How are you	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I want my bed	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Good night Gary	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I am tired	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Go to the bathroom	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Put on my jammas	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
See you in the morning	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I'll brush my teeth	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Clean my glasses	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I am cold	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I am hot	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I am very happy	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Can you help me	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Can you get me something please	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I love you	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes

I am feeling frustrated	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
I am sad	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
That's really funny	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes
Don't do that, please stop	**-2**	**-1**	**0**	**+1**	**+2**	Yes	No	Sometimes

5. What aspect/s of therapy do you think benefited KC most?

6. What do you think could have made therapy even better for KC?

APPENDIX B

RAW DATA USED TO CALCULATE INTRACLASS CORRELATION

Raw data used to calculate intraclass correlation

Sessions observed by raters in chronological order	Rater 1	Rater 2
1	39	36
2	29	29
3	34	34
4	32	32
5	33	33
6	40	38
7	47	47
8	59	56
9	44	40
10	56	57
11	72	71
12	77	76
13	84	83
14	85	84
15	77	72
16	44	30
17	60	64
18	73	79