Effect of Music on Reading Comprehension for Patients With Aphasia (EMRA1)

NCT02875938

April 28, 2016

Method

Human Rights Protections

Respect for Persons requires that "1) individuals should be treated as autonomous agents, and 2) that persons with diminished autonomy are entitled to additional protections" (NIH Protecting Human Research Participants p. 37). Because a patient with aphasia could be considered to have Diminished Autonomy due to injury, it was necessary to obtain consent from the participant to the extent possible, based on 1) the individual's level of capacity and 2) the complexity and risks of the study. Patients without capacity to make an informed decision about treatment were excluded from this study. In order to enroll a patient as a subject in this research, it was required that the referring physician document that the patient had capacity to make an informed decision about their treatment. The physician's determination was documented either through written communication with the researcher, including emails, or by a note to file by the researcher that documented a verbal communication from the physician.

In order to lessen any prospect of undue influence by the treating SLP for her patients to consider taking part in research, prospective subjects were initially approached by another research team member, the Director of Rehabilitation, for recruitment. Potential subjects had met certain eligibility criteria established by the principal investigator (PI). The recruiter scheduled a meeting with prospective subjects who were accompanied by a family member or person of their choosing. The recruiter then followed the outline of a script prepared by the PI to introduce prospective subjects to the study. The script emphasized that taking part in the study is purely voluntary, that refusal to take part would not result in any change in their standard treatment by the SLP, and that there is no way to know whether being in the study will benefit their treatment. The recruiter was trained by the PI on how to communicate with patients with aphasia, particularly with those who rely on responding to yes-no questions. If prospective subjects indicated a willingness to meet with the PI to discuss taking part in the study, the recruiter provided their names and contact information to the PI.

The PI then scheduled a meeting with prospective subjects who had indicated a willingness to learn about the study. The prospective subjects were asked to bring a family member or other person of their choosing to the meeting. Strategies for limited or non-readers were used including reading the consent document to prospective subjects and spending additional time to confirm their understanding. The PI reviewed the consent document section by section. At the end of the consent discussion, the PI assessed the patient's understanding of the study by asking the following questions requiring only yes-no-don't know answers based on the Aid to Capacity Evaluation (ACE) questions with or without picture/word cues to express responses:

1. Do you have trouble with language understanding, talking, reading and writing since your stroke?

2. Can therapy for reading help you get better?

3. Can music be used with reading to help you get better?

4. If you don't want to use music, can you still get regular speech therapy?

5. Is it ok to say no to the research?

6. Can you stop using the music with reading therapy if you want to?

7. Will you get paid to be in the research?

If a prospective subject answered a question incorrectly, the part of the consent document addressing the answer was reviewed. The prospective subject was then asked the questions again. If, in the judgment of the PI, the prospective subject still had significant misunderstanding of the research, the subject was not asked to enroll.

Subjects who did exhibit an understanding based on the above questions were then given a week to think about being in the research. If they were interested, another meeting was scheduled during which they would sign the consent document. The subject was asked to include a family member or other person of their choosing to accompany them and monitor the informed consent discussion. This person was asked to sign a statement on the consent document stating they were present during the consent discussion, the material in the consent form was explained to the subject, and consent by the subject was given voluntarily.

Data was not considered *sensitive* but the anonymity of participants needed to be protected (Gast & Ledford, 2014, p.43). Participant data was number-coded, so it could not be traced back to the participant's identifying information. The PI and rehabilitation manager were the only ones with access to the code, which was locked in a file cabinet in the speech therapy office. The number-code key will be shredded following completion of the research. No research data was entered into the participant's electronic medical record. Paper files of session data were also protected in the locked file cabinet. The investigation, participants, or participants' progress were not discussed with anyone not directly involved in the study, and for those involved communications were conducted in private locations, for example the speech therapy office. Any potential compromise of participant confidentiality was to be reported to the Internal Review Board (IRB) within one week; none occurred. Digital recording of every third session was used to obtain reliability and procedural fidelity data. Patient performance on the phrase completion task was videotaped every third visit/weekly using the hospital emergency iPad. The iPad was checked out in the morning before the participant's treatment session, the last treatment trial (for procedural fidelity) and the participant's pointing responses to the phrase completion task (for inter-observer reliability) recorded, and stored in the Speech Therapy office in the Rehabilitation building until viewed by the assistant in a private location and recorded on the paper reliability data forms. The digital recording was then deleted and the iPad returned. Digital videos did not show participants' faces, only hands circling the response were recorded.

Data storage included patient telephone numbers which were accessible in the participant's soft file for patient contact while the patient was undergoing speech therapy treatment, as is common practice. On discharge the soft file was stored in the Medical Records Department.

One potential problem would be a negative result, that music would decrease language/reading by overloading/interfering with semantic processing pathways used for language. Because progress was continually monitored, any negative result could be found quickly so severity and duration would be minimized, and intervention procedures adjusted to increase the potential for a

beneficial outcome. In order to assess whether the study interventions appear to be interfering with normal speech therapy for reading comprehension phrase tasks, the researcher continuously monitored the average percentage correct on multiple choice phrase completion questions for the two experimental music interventions. The researcher used two methods to decide if subjects were making progress comparable to patients who have had standard speech therapy interventions for completion of reading phrase tasks.

The first method was based on experience with a previous patient whose progress with reading phrase completion in normal speech therapy was measured in a celeration graph. For this study, normal progress was defined as an ascending celeration line over up to 17 visits on graphic presentation of the intervention data with two or fewer outliers. Outliers were defined as three standard deviations beyond the trendline. If at any time a subject had three or more outliers in either experimental music intervention, that intervention would be stopped and alternative approaches used.

The second method was to review the percentage correct in completion of phrases at each experimental condition visit. If there was no increase in percentage correct over three consecutive music intervention visits for a condition, that intervention condition would be stopped and alternative approaches used.

Individual benefit could not be guaranteed because the intervention was experimental. The benefit to the medical community and society in general was the advancement of knowledge regarding therapeutic procedures to improve healing/neuroplasticity after stroke.

Participant Recruitment

Participants were pre-selected consecutively from the accessible out-patient or home health population as they were diagnosed and referred for ST by a physician or neurologist.

1. Patient was referred by physician or neurologist with orders for ST.

2. Evaluation by the certified SLP showed a history of an isolated left MCA CVA based on electronic medical records (EMR) and CT/MRI imaging reports; if ambiguous, the SLP would verify with the neurologist.

3. Evaluation by the certified SLP showed resultant aphasia based on skilled interpretation of limited language performance, and the Boston Diagnostic Aphasia Examination (BDAE-3) test scores were below the 70th percentile, with scores on the Reading subtests at or between the 20th and 70th percentiles.

4. Visual field deficits were ruled in-out based on the EMR. Any subsequent behavioral evidence indicating possible visual field deficits would generate a referral for evaluation by a licensed ophthalmologist.

5. Pre-selected candidates for participation were then reported to the Rehabilitation Director, who interviewed the candidate and family/power of attorney (POA) for recruitment according to the established script, so that voluntary participation without coercion was ensured.

6. Interested candidates were then interviewed by the SLP outside of treatment time to conduct the informed consent process, and given at least one week to consider their participation and consult with others before signing the informed consent document.

Considerations in Participant Selection

The main neurological structures that must remain intact for the proposed treatment to work are lateralized left hemispheric, that is left amygdala and left extrastriate cortex. While it seems intuitive to choose patients with right CVA as the trial subjects, deeper analysis of the fMRI results reveals that areas on the right middle and superior temporal gyri and bilaterally the superior parietal gyrus post central were simultaneously activated. In addition, it has been shown that reading comprehension is bilaterally organized. For example, Hauk et al. revealed that words are processed by distributed neuronal networks that reflect word semantics (Hauk, Johnsrude, & Pulvermuller, 2004). Specifically, event-related fMRI showed that action words/verbs presented in passive reading differentially activated areas in the primary motor cortex involved in the actual movements. Further, the left hemispheric structures involved in re-entrant processing are not part of the primary speech/language areas, that is Broca's, conduction, or Wernicke's areas. Therefore, the population of subjects who could most benefit from the proposed treatment should be those with localized cortical aphasia.

Because the identification of the intact targeted structures, that is left amygdala to left extrastriate cortex, was the basis for inclusion in this study, signs/symptoms defining injury to those neuro-anatomical structures should also be the basis of exclusion criteria. Type of aphasia need not be a limiting factor. In single-subject research design, each subject serves as his/her own control, so severity levels need not be defining criteria, because the effect on the outcomes of both conditions would be the same. Hemianopsia or other visual field defects, unilateral hemi-neglect or inattention affecting vision for reading needed to be excluded, and in particular right homonymous hemi- or inferior quadrantanopsia which would indicate a lesion of the left parietal or temporal lobes and involvement of the left optic radiations.

Inclusion/Exclusion Criteria

Inclusion Criteria: Participants were post-hospitalization with no previous speech therapy intervention for reading above the word level determined from electronic medical records (EMR) or transmitted treatment reports; premorbid reading at the 8th grade level or higher based on the participant's/family's stated years of formal education; post-stroke aphasia/left MCA CVA based on EMR and CT/MRI imaging reports; intact left amygdala to left extrastriate cortex based on EMR and CT/MRI imaging reports, or consult with the neurologist.

Exclusion Criteria: Potential participants were excluded based on co-morbidities diagnosed and reported in the EMR or shown on CT scan or MRI including past history of stroke with residual deficits, dementia, Parkinson's disease, head injury, etc.; hemianopsia or other visual field defects affecting vision for reading, and in particular right homonymous hemi- or inferior quadrantanopsia; previous history of learning disabilities in reading/writing; significant psychiatric diagnosis; English as a Second Language or non-English language.

Reporting Participant Characteristics

Characteristics to be reported included age; sex; time post onset of left MCA CVA; type of stroke; previous level of function including educational level. Body functions/structures, activities, and participation were reported for previous level of function and changes from the stroke per the World Health Organization (WHO) International Classification of Functioning Disability and Health (ICF) (World Health Organization, 2013).

A meta-analysis of studies using music for the treatment of neurological language and speech disorders showed that although measurable improvement was reported, the methodological quality of studies was rated low (Hurkmans et al., 2011). Specific recommendations were made to report education, dominance, and musical background, all of which were included in the current study. Education was reported based on highest level of formal education attained. Dominance was determined by premorbid handedness for writing. Musical background included reporting previous formal instruction for singing/choir, playing an instrument, or reading music, and experience with performance of singing, dancing, and listening to music.

Sample Size

A minimum of four participants was required to demonstrate replication of treatment effects (Gast & Ledford, 2014, p.327). A maximum of ten participants was requested from the Internal Review Board (IRB) to account for attrition. This study reports on the first two participants of the approved trials.

Environment

The physical environment was the Out-Patient Rehabilitation clinic in the ST office which is also a private treatment room (for Participant #2), and the patient's home when served by home health (for Participant #1). The clinic was located at the end of a hallway away from the other therapies so there was little noise or commotion. The SLP and client both sat at a large table with a minimum of materials laid out. The room was well-lit with no windows. Limited use of bright color added appeal without drawing undue attention. Distractions were kept to a minimum, but included training materials and a desk-top computer behind the therapist. All other materials were kept in a closed cabinet and filing cabinet. The patients and significant others were seated in various locations at the table to best accommodate vision and hearing. Home settings were controlled for lighting, sound, and undue distractions to the maximum extent possible.

Independent Variable:

The independent variable was manipulated to contrast three conditions:

1. Polyphonic music with song lyrics presented simultaneously with silent reading and observation of pointing to written lyrics (independent processing condition);

2. Priming with polyphonic music with lyrics followed by silent reading and observation of pointing to written lyrics (shared processing condition);

3. A limited control set using listening to oral reading of the lyrics and silent reading of the written lyrics with observation of pointing to written lyrics (no music).

Dependent Variables:

The dependent variable was defined as silent reading/comprehension of targeted words/phrases by pointing to words/phrases in a phrase completion exercise without music. The dependent variable was kept constant for all three conditions.

Dependent Measures

Measurement of the dependent variable consisted of both direct systematic observational recording (DSOR) and treatment record video (for reliability/procedural fidelity assessment) of percentage correct over time (including follow-up) and trials- or sessions-to-criterion.

Research Design

An AATD was used for comparison of two separate but similar conditions when the effects may be non-reversible (see Gast & Ledford, 2014, pp. 324-332). Time between sessions was at least a day.

"Criteria:

1. Behaviors must be non-reversible—the behaviors will continue to be performed after instruction has stopped.

2. Behaviors should not be in the participants' repertoire.

3. Behaviors must be independent, meaning one behavior set/chain can be acquired without influencing performance on other behavior sets/chains.

4. Behaviors also must be functionally equivalent, meaning behaviors are likely to be influenced by the same environmental variable (e.g., the instructional strategies being studied).

5. *Behavior sets/chains must be of equal difficulty*. The dependent variable in this investigation will be phrases acquired, but three different sets will be required to fulfill the requirement that the behaviors are independent. To ensure that they are of equal difficulty, they must all be assessed for word complexity at a minimum.

Randomizing items between treatment conditions should result in a relatively equal distribution of difficulty" (Romer, Billingsley, & White (1988).

Procedures

All songs were pretested for familiarity and positive valence by questioning the participants (i.e. Do you know this song? Do you like it?). Response data was recorded. All songs perceived as not liked were eliminated and substitutions made; the new songs were then also pretested for familiarity and positive valence. Songs that were not familiar but were liked were noted but not changed.

During the baseline (A) all 25 phrases (Appendix C) were tested using the phrase completion task (Appendix D). Both correct responses and participation were verbally reinforced, as is required in an AATD. Results were analyzed by the SLP to ensure that the sets were of equal difficulty for each participant using two different procedures. First, logical analysis of the linguistic characteristics of the phrases was completed, including the number of words, the number of syllables in the words, the number of consonant blends in the words, and the number contractions. Second, experimental evaluation of the number of error responses in each set during the baseline phase was completed, and without changing the items or responses, the set assignment was manipulated to maximally equate the number of error phrases without significantly changing the equivalency of the linguistic characteristics. One phrase was shifted from set 1 to set 2 during the baseline phase. In the end, the number of error responses on baseline for set 1 was three, for set 2 was two, and for the control set was two.

During the intervention (B), each song (music with lyrics) from set 1 (B1) was played in its entirety. When the targeted verse or chorus played, the written lyrics were simultaneously shown along with the song lyrics. The SLP pointed to the words of the targeted verse/chorus which was silently observed by the participant as they were sung. This was repeated for all ten songs. The phrase completion task was then presented, including the total 25 phrases, without music. The participant silently or orally read the phrase and circled the phrase completion choice. Correct responses were randomly verbally reinforced. Error responses were followed by verbal modeling and pointing to the correct choice by the SLP. The next session, each song (music with lyrics) from set 2 (B2) was played in its entirety. After the song played, the written lyrics from the targeted verse or chorus were presented. The SLP pointed to the words of the targeted verse/chorus which was silently observed by the participant. This was repeated for all ten songs. The phrase completion task was then presented, including the total 25 phrases. The participant silently or orally read the phrase and circled the phrase completion choice. Correct responses were randomly verbally reinforced. Error responses were followed by verbal modeling and pointing to the correct choice by the SLP. This was repeated for all ten songs. For all interventions, the number of correct responses were tallied on a LaPoint Base 10 form and the percent correct entered on the graph for that day. The same procedures were repeated, alternating procedures on sets 1 and 2 on subsequent treatment visits, until a criterion of 9/10 (90%) x 3 consecutive visits was reached on both sets. It was predetermined that if the criterion was not reached on one of the intervention sets, defined as no increase in percentage correct x 3 consecutive visits, that intervention would be considered to be limited in effectiveness and discontinued. Every third session was digitally recorded for a reliability check by the trained assistant. When both of the sets (B1 and B2) had been acquired or one acquired and the other discontinued, the number of sessions to criterion were counted for each intervention.

Potential Threats to Internal Validity

Potential threats to internal validity included multiple-treatment interference in the form of sequential confounding and carryover effects. Counterbalancing conditions was necessary, and any carryover effects identified (see Gast & Ledford, 2014, chapter 5 p.101). Because AATD studies test efficiency, the goal involves demonstrating optimal performance as opposed to artificially deflated performance, so responses were reinforced during the baseline phase (see Gast & Ledford, 2014, p. 326).

Potential threats to internal validity also included possible history or maturation effects. A control set was used to rule out history and maturation effects.

Outcome Measures and Time Points

Pretest/post-test information on the ICF Body Functions/Structures, Activities, and Participation was provided (Hurkmans et al. 2011). The Boston Diagnostic Aphasia Examination (BDAE) (Goodglass, Harold, & Kaplan, 1983) was used as an outcome measure of reading acquisition. The BDAE is a classic anatomically-based diagnostic instrument. Because various components of language function may be selectively damaged by CNS lesions, results indicate the neuro-anatomical organization of language and localization of the lesion. Thus it serves as an appropriate assessment for research exploring neurological organization in the brain. Pretest reading scores on selected subtests from the BDAE full version were completed prior to baseline A and initiating intervention, and immediately following completion of intervention to assess generalization to non-trained stimuli. At four weeks follow-up the trained stimuli from all three interventions was retested to assess retention. The dependent variables, that is differential response to the two interventions (counted as the percent correct responses per day, a ratio scale which has equal intervals and an absolute zero, which allows for multiplication and division so is appropriate for averages), number of sessions required to achieve mastery at each level (simple counting but also a ratio scale as it has equal intervals and an absolute zero which allows for multiplication and division so allows for the difference between means), retention of the material (number of items remembered, another simple count which is also a ratio scale as it has equal intervals and an absolute zero and allows for the difference between means), was assessed using data analysis as described below.

Data Analysis

Reliability of the items in the task assessment tool was established using Cronbach's Alpha correlation coefficient once at the beginning of the baseline phase. Inter-observer reliability on the scoring of the tool was determined weekly using the kappa coefficient, also known as Cohen's Kappa. Reliability data was first analyzed for normality/skewness, accuracy, and completeness. Outliers were identified using the Outlier Labeling Rule based on rank and percentile. If data were incomplete, for example if the behavior could not be clearly seen on the digital recording, the corresponding score from the other researcher was eliminated from the reliability determination.

The correlation coefficient chosen is dependent on the scale of measurement of the variables being correlated. For the reliability of the task assessment tool it was used to determine if scores recorded by one researcher were strongly correlated with scores recorded by the other. Scores were recorded as correct versus incorrect, which is a nominal scale. One classic correlation coefficient used with a nominal scale is Cronbach's Alpha, used when both X and Y are dichotomous variables. The assumption that the items of the test are unidimensional cannot be met, meaning that each item measures the same ability on the same scale, because many different mental processes may be used in various combinations from item to item. Cronbach's alpha provides a solution to this problem, so was an appropriate measure of internal consistency (Cronbach, 2004, see p.8). Acceptable values of alpha range from .70 to .90 (Tavakol &

Dennick, 2011). It was predetermined that if alpha was below .70, either the task assessment tool would be revised or the assistant would be re-instructed.

Inter-observer reliability was determined by the kappa coefficient, used when there are two observers. Again scores were recorded as correct vs. incorrect, a nominal scale. Simple percent agreement cannot account for chance agreement/uncertainty. The kappa statistic was developed to control for random agreement. Interpretation was based on the kappa coefficient squared to show the accuracy in the data due to congruence between data collectors. Acceptable kappa2 values will range from (positive) .90-1.00 (McHugh, 2012). It was predetermined that if the accuracy was questionable, the cause would be determined, and may indicate the need for further instruction of the assistant(s) on data collection. If reliability was below simple percent agreement 80% or kappa2 .90 on any day, the recording would be reviewed again by both researchers in an attempt to reach consensus.

Data analysis for treatment included both visual analysis and quantitative statistical analysis as follows:

- Visual analysis included graphing average (mean) correct responses on the multiple-choice reading comprehension questions each session during each phase. Trend was described by celeration lines depicting slope in each phase with split middle lines to measure central tendency.

- Quantitative statistical analysis included the binomial test for dichotomous outcomes, and the two standard deviations band method.

- The binomial test involved counting data points above and below the split middle line and comparing the total number of points above and below with the number with fewer points to determine Spearman's Rank Correlation Coefficient, one-tailed probabilities which show whether the response pattern could have occurred by chance (< 0.05 is considered significant) (Portney & Watkins, 2009).

- The two standard deviations band method shows variability within the baseline phase based on the mean and standard deviation of data points, depicted by an area on the graph two standard deviations above and below the mean, then extending this area into the intervention phase. If at least two consecutive data points in the intervention phase fall above (in this case) the two standard deviation band, changes from baseline to intervention are considered significant (Portney & Watkins, 2009).

- The difference in number of sessions/trials to achieve mastery at each level for each intervention, and the difference in retention at three weeks follow-up were analyzed by comparison of the difference between data points for the control treatment and the two target intervention conditions. After the research design has been completed by a minimum of four subjects, data will be consolidated for group means.

Data Storage

Patient telephone numbers were accessible in the participants' soft file for patient contact while the patient was undergoing speech therapy treatment, as is common practice. On discharge the soft file was stored in the hospital Medical Records Department.

Participants' performance on the phrase completion task was videotaped every third visit using the hospital emergency iPad. The iPad was checked out in the morning before the participant's treatment session. The last treatment trial (for procedural fidelity) and the participant's responses to the phrase completion task (for inter-observer reliability) were recorded and stored in the speech therapy office in the Rehab building until viewed by the assistant in a private location and recorded on the paper reliability data forms. The digital recording was then deleted and the iPad returned. Digital videos did not show participant faces, only hands on the response task were recorded.

Role of the researcher

The primary investigator (PI) pre-selected candidates for participation according to the stated recruitment procedures (above). The rehabilitation manager recruited participants, but the PI obtained informed consent**.** Given the therapy location, the PI conducted the experimental treatments, making adjustments to the protocol based on preliminary findings as necessary, with IRB approval for any changes. The PI was also responsible for training the (blinded) assistant for reliability checks from videotapes, and interpreting and reporting the data.

Assumptions and Limitations

The role of the primary investigator presents the possibility of a threat to construct validity in the form of potential experimental bias, either by expectations of the subjects via the Hawthorne effect, or from unintentional experimenter effects. To reduce this threat, participants were blinded to the exact purpose of the experiment until all data was collected, and reliability checks of measurement data were conducted by an assistant blinded to the purpose of the study, but the study could still be limited by experimenter effects. Given a significant result of this investigation, follow-up systematic replication studies will be conducted in other locations with other skilled therapists performing the same experimental treatments.

Threats to internal validity included possible history or maturation effects. History effects are extraneous events prior to post-test that can affect the DV. Maturation effects are changes in the DV resulting from the passage of time, e.g. spontaneous recovery. In AATDs, a control set is used to rule out history and maturation effects.