Vocal music enhances memory and language recovery after stroke: pooled results from two RCTs

Aleksi J. Sihvonen, Vera Leo, Pablo Ripollés, Terhi Lehtovaara, Aki Ylönen, Pekka Rajanaro, Sari Laitinen, Ani Forsblom, Jani Saunavaara, Taina Autti, Matti Laine, Antoni Rodríguez-Fornells, Mari Tervaniemi, Seppo Soinila, Teppo Särkämö

1Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland
2Department of Neurosciences, Faculty of Medicine, University of Helsinki, Helsinki, Finland
3Department of Psychology, New York University, New York, New York
4Private Music Therapy Practitioner, Helsinki, Finland
5Department of Medical Physics, Turku University Hospital, Turku, Finland
6Department of Radiology, HUS Medical Imaging Center, Helsinki University Central Hospital, University of Helsinki, Helsinki, Finland
7Department of Psychology, Abo Akademi University, Finland
8Cognition and Brain Plasticity Group, Bellvitge Biomedical Research Institute, L’Hospitalet de Llobregat, Barcelona, Spain
9Department of Cognition, Development and Education Psychology, University of Barcelona, Barcelona, Spain
10Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain
11CICERO Learning, University of Helsinki, Helsinki, Finland
12Division of Clinical Neurosciences, Department of Neurology, Turku University Hospital and University of Turku, Turku, Finland

Correspondence
Teppo Sarkamo, Cognitive Brain Research Unit, Department of Psychology and Logopedics, Haarmanninkatu 3, P.O. Box 21, FI-00014 University of Helsinki, Finland. Tel: +358 941 29464; E-mail: teppo.sarkamo@helsinki.fi

Funding information
Financial support for the work was provided by the Academy of Finland (grants 257077, 277693, 299044), Tyks Research Funding (grant 13944), Finnish Brain Research and Rehabilitation Foundation, Finnish Brain Foundation, Finnish Cultural Foundation, National Doctoral Programme of Psychology, Ella and Georg Ehrnrooth Foundation, Signe and Ane Gyllenberg Foundation, Maire Taponen Foundation, and Jenny and Antti Wihuri Foundation.

Received: 20 March 2020; Revised: 27 August 2020; Accepted: 18 September 2020

doi: 10.1002/acn3.51217

Abstract

Objective: Previous studies suggest that daily music listening can aid stroke recovery, but little is known about the stimulus-dependent and neural mechanisms driving this effect. Building on neuroimaging evidence that vocal music engages extensive and bilateral networks in the brain, we sought to determine if it would be more effective for enhancing cognitive and language recovery and neuroplasticity than instrumental music or speech after stroke. Methods: Using data pooled from two single-blind randomized controlled trials in stroke patients (N = 83), we compared the effects of daily listening to self-selected vocal music, instrumental music, and audiobooks during the first 3 poststroke months. Outcome measures comprised neuropsychological tests of verbal memory (primary outcome), language, and attention and a mood questionnaire performed at acute, 3-month, and 6-month stages and structural and functional MRI at acute and 6-month stages. Results: Listening to vocal music enhanced verbal memory recovery more than instrumental music or audiobooks and language recovery more than audiobooks, especially in aphasic patients. Voxel-based morphometry and resting-state and task-based fMRI results showed that vocal music listening selectively increased gray matter volume in left temporal areas and functional connectivity in the default mode network. Interpretation: Vocal music listening is an effective and easily applicable tool to support cognitive recovery after stroke as well as to enhance early language recovery in aphasia. The rehabilitative effects of vocal music are driven by both structural and functional plasticity changes in temporoparietal networks crucial for emotional processing, language, and memory.
Introduction

During the last decade, there has been growing interest toward music as a neurorehabilitation tool, especially for stroke. This has been fueled by (1) the rapidly increasing prevalence of stroke and its massive socioeconomic burden and growing need for cost-effective rehabilitation tools and (2) advances in music neuroscience, uncovering the wide-spread cortical and subcortical networks underlying the auditory, motor, cognitive, and emotional processing of music and their malleability by musical training. In the rehabilitation context, music can be viewed as a form of environmental enrichment (EE) that increases activity-dependent neuroplasticity in the large-scale brain network it stimulates. In animals, EE is a powerful driver of synaptic plasticity, neurotrophin production, and neurogenesis, improving also cognitive-motor recovery. In stroke patients, EE where patients are provided additional social interaction and stimulating activities (e.g., games) is emerging as a promising way to increase physical, social, and cognitive activity.

Previously, we explored the long-term efficacy of musical EE in a three-arm randomized controlled trial (RCT) comparing daily music listening to a control intervention (audiobook listening) and standard care (SC) in stroke patients. Music listening enhanced the recovery of verbal memory and attention and reduced negative mood as well as increased gray matter volume (GMV) in spared prefrontal and limbic areas in left hemisphere-lesioned patients. Corroborating results were recently obtained in another RCT where daily music listening, alone or in combination with mindfulness training, enhanced verbal memory and attention more than audiobooks. While these results imply that music listening can be cognitively, emotionally, and neurally effective after stroke, its tailored, results suggest that (i) vocal music would be superior to instrumental music and audiobooks in enhancing cognitive and language recovery, (ii) both vocal and instrumental music would enhance mood more than audiobooks, and (iii) the rehabilitative effects of vocal music would be linked to GMV changes in temporal, frontal, and parietal regions associated with the processing of language, music, and memory and commonly induced by musical training as well as increased resting-state functional connectivity (FC), particularly in the default mode network (DMN), which has recently been linked to stroke recovery. Moreover, given previous evidence on singing-based speech rehabilitation in aphasia, we explore whether listening to vocal music can be effective for aphasia recovery.

Methods

Subjects and study design

Subjects were stroke patients pooled from two RCTs performed in Turku and Helsinki, Finland. Data pooling was done to increase sample size and statistical power, and was feasible because both studies had common (1) inclusion/exclusion criteria (MRI-verified acute unilateral stroke; right-handed; <80 years old; Finnish speaking; able to co-operate; no hearing loss, prior neurological/psychiatric disease, or substance abuse); (2) assessment time points [<3 weeks (baseline, T0), 3-month (T1), and 6-
Vocal Music Listening Enhances Stroke Recovery

A. J. Sihvonen et al.

In the Turku RCT (ClinicalTrials.gov identifier: NCT01749709, previously unreported trial), 50 stroke patients (83% of planned sample size of N = 60) were recruited during 2013–2016 from the Department of Clinical Neurosciences of Turku University Hospital. The patients were randomized to the following three groups: vocal music group (VMG, N = 17), instrumental music group (IMG, N = 17), and audiobook group (ABG, N = 16). Forty-five patients completed the trial up to T1 and 44 up to T2. In the Helsinki RCT [nonregistered (data collection began before the ICMJE trial registration recommendations), previously reported9 trial], 60 stroke patients (100% of planned sample size) were recruited during 2004–2006 from the Department of Neurology of Helsinki University Central Hospital (HUCH). The patients were randomized to the following three groups (N = 20 in each): music group (MG), ABG, and control (SC only) group. In the Helsinki RCT, 55 patients

Figure 1. Flow chart outlining the design and progress of the study.
completed the trial up to T1 and 54 up to T2. Those Helsinki MG (N = 19) and ABG (N = 19) patients who had T1 follow-up data were included in the present study. Based on the proportion of listening to vocal and instrumental music (data obtained from the listening diaries of the patients and the case notes of the music therapists), the 19 MG patients were reclassified to VMG [N = 13; vocal music: mean 82.6% (SD 11.9%)] and IMG (N = 6; instrumental music: mean 54.2% (SD 13.3%)), which differed highly significantly regarding the music type [t (17) = 6.63, P < 0.001]. In other words, the MG patients in the Helsinki study were not randomized to VMG and IMG, but were reallocated to these groups post hoc based on the type of music they had listened to during the intervention period. This was done in order to pool the data with the Turku study. The final pooled data comprised 83 patients from T0 to T1 (VMG: N = 27, IMG: N = 23, ABG: N = 33) and 81 patients from T0 to T2 (VMG: N = 26, IMG: N = 23, ABG: N = 32).

Intervention

After baseline (T0) assessments, a professional music therapist contacted the patients, informed them of the group allocation, and interviewed them about prestroke leisure activities, including music listening and reading. The therapist provided the patients with a portable player, over-ear headphones, and a collection of listening material, which was vocal music with sung lyrics in VMG, instrumental music (with no sung lyrics) in IMG, and narrated audiobooks (with no music) in ABG. All material was in a language that the patients understood well (mostly Finnish or English). Within each group, the material was selected individually to match the music/literature preferences of the patient as closely as possible. The music material comprised primarily pop, rock, and schlager music songs in the VMG and classical, jazz, and film score music in the IMG. The patients were trained in using the players and instructed to listen to the material by themselves daily (min. 1 h per day) for the following 2 months at the hospital or at home. They were also asked to keep a listening diary. During the 2-month intervention, the therapist kept regular contact with the patients to encourage listening, provide more material, and help with the equipment if needed. On average, the therapist spent around 4 h with each participant during the intervention period. After the intervention period, the patients were free to continue listening at their own will.

Behavioral outcome measures

Neuropsychological testing was performed three times (T0/T1/T2), blinded to the group allocation of the patients. The assessment battery (Table 1) used in both the Turku and Helsinki studies covered three cognitive domains (verbal memory, language skills, and focused attention) and mood (POMS). The raw test scores measuring each domain were added up and these summary scores were used in the statistical analyses to reduce the number of variables. The primary outcome was verbal memory (change from T0 to T1). The secondary outcomes were verbal memory (change from T0 to T2) as well as language and focused attention (change from T0 to T1 and T2) and POMS scales (at T1 and T2). Parallel versions of memory tests were used in different testing occasions to minimize practice effects. Testing was carried out in a quiet room reserved for neuropsychological studies, over multiple sessions if needed to avoid fatigue. The level of aphasia at T0 was assessed with the BDAE-ASRS, with score ≤ 4 indicating aphasia. Scoring was done clinically primarily based on conversational speech, drawing information also from the language tests (Table 1). All the aphasic patients had left hemisphere lesions.

Statistical analysis of behavioral data

Demographic and clinical characteristics were analyzed with univariate ANOVAs or nonparametric Kruskal–Wallis tests, and chi-squared tests. Longitudinal cognitive data were analyzed using mixed-model ANOVAs with Time as within-subject factor and Group (VMG/IMG/ABG) and Aphasia as between-subject factors. Separate mixed-model ANOVAs were performed to determine the short-term (Time: T0/T1) and long-term (Time: T0/T2) effects of the intervention. Due to the emotional lability of the patients at the acute stage, the POMS data were analyzed cross-sectionally at each time point using univariate ANOVAs. In order to control for the potential effects of the two trial sites (Turku/Helsinki) and those demographic and clinical variables, in which there were group differences (see Results), in the analyses of outcome measures, trial site, amusia, prestroke music listening, and cross-listening were included as covariates in the ANOVAs. Post hoc tests of change scores (T1-T0, T2-T0) were performed using the Bonferroni correction. All statistical analyses were performed with IBM SPSS Statistics 24. Missing values in the data were considered missing at random and were not replaced.

MRI data acquisition

Structural MRI was acquired from 75 patients at T0 and T2 using a 1.5T Siemens Vision scanner of the HUCH Department of Radiology (Helsinki, N = 33) and a 3T Siemens Verio scanner of the Medical Imaging Centre of Southwest Finland (Turku, N = 42). High-resolution T1 images (Helsinki: flip angle 15°, TR 1900 msec, TE
3.68 ms, voxel size 1.0 × 1.0 × 1.0 mm; Turku: flip angle 9°, TR 2300 msec, TE 2.98 msec, voxel size 1.0 × 1.0 × 1.0 mm) were obtained, coupled with fluid-attenuated inversion recovery (FLAIR) images used in localizing the lesions. Functional MRI was acquired from 35 Turku patients at T0 and T2 using a single-shot T2*-weighted gradient-echo EPI sequence (flip angle 80°, TR 2010 msec, TE 30 msec, voxel size 2.8 × 2.8 × 3.5 mm, slice thickness 3.5 mm, 32 slices), with a total of 280 functional volumes being collected. The session included a 5-minute eyes-open resting-state condition (rs-fMRI) and a passive listening task (task-fMRI) during which auditory stimuli were presented through MR-compatible headphones using Presentation software (Neurobehavioral Systems, version 16.3). In task-fMRI, we used a block design where the patients were presented 15-second excerpts of well-known Finnish songs with (1) sung lyrics (Vocal, six blocks) and (2) without sung lyrics (Instrumental, six blocks), (3) well-known Finnish poems (Speech, six blocks), and (4) no auditory stimuli (Rest, 18 blocks). The order of the auditory blocks was randomized across subjects and time, and the rest blocks were presented in between the auditory blocks.

**MRI data preprocessing**

MRI data were preprocessed using Statistical Parametric Mapping software (SPM8, Wellcome Department of Cognitive Neurology, UCL) under MATLAB 8.4.0. The structural T1 images of each subject were reoriented to the anterior commissure and then processed using Unified Segmentation with medium regularization. Lesioned areas were not excluded from subsequent analyses (see below), but cost function masking (CFM) was applied to achieve accurate segmentation and optimal normalization of the lesioned GM and white matter (WM) tissue, with no postregistration lesion shrinkage or out-of-brain distortion. Using MRicon (https://www.nitrc.org/projects/mricron), CFM was performed by manually depicting the lesioned areas slice-by-slice to the T1 images of each subject. The segmented GM and WM images were modulated to preserve the original signal strength and then normalized to the MNI space. After this, to reduce residual interindividual variability, GM and WM probability maps were smoothed using an isotropic spatial filter (FWHM 6 mm). For fMRI data, the functional runs were first realigned and their mean image was calculated. The T1 image and its lesion mask were then co-registered to this mean functional image. The normalization parameters were again estimated using Unified Segmentation with CFM and were applied to the whole functional run to register it to MNI space. In this registration step, data were resampled into 2.0 × 2.0 × 2.0 mm voxel size. Finally, the normalized fMRI data were smoothed using an 8-mm FWHM kernel.

**Voxel-based morphometry**

VBM analysis was performed using SPM8. The individual preprocessed GM and WM images were submitted to second-level flexible factorial analyses with Time (T0/T2) and Group (VMG/IMG/ABG) as factors, and scanner type (i.e., trial site), age, sex, and total intracranial volume as additional covariates. Thus, altogether three Group (VMG > ABG, IMG > ABG, and VMG > IMG) × Time (T2 > T0) interactions were calculated. Separate analyses were also performed within the aphasic and non-aphasic patients. All results were thresholded at an uncorrected $P < 0.005$ threshold at the voxel level, and standard SPM family-wise error rate (FWE) cluster-level correction
based on random field theory (RFT) with a $P < 0.05$ was used. Only clusters surviving FWE-corrected $P < 0.05$ at cluster level are reported. For each significant comparison, the minimum size for a cluster to be corrected for multiple corrections (FWEc) is reported. Neuroanatomical areas were identified using the Automated Anatomical Labeling Atlas provided within the xjView toolbox (http://www.alivelearn.net/xjview/).

**Functional connectivity**

Group-level spatial independent component analysis was performed using the Group ICA of fMRI Toolbox (GIFT) software (http://mialab.mrn.org/software/gift/). The ICA spatial components were extracted from the rs-fMRI and task-fMRI runs. After performing intensity normalization of the preprocessed fMRI images, data were concatenated and, following previous studies, reduced to 20 temporal dimensions using principal component analysis and then analyzed using the Infomax algorithm. From the ICA spatial components representing the different networks, the default mode network (DMN) was identified and selected for further analyses based on the pattern of VBM results (see Results section). In the rs-fMRI, to obtain whole-brain group-wise statistics, the spatial maps of the DMNs from all patients were submitted to a second-level flexible factorial analyses with Time (T0/T2) and Group (VMG/IMG/ABG) as factors. In the task-fMRI, the time course of the DMN was fitted to an SPM model that included the Vocal, Instrumental, and Speech conditions as regressors, yielding beta values representing the engagement of DMN during each condition. These were then analyzed with SPSS using mixed-model ANOVAs with Time (T0/T2) and Group (VMG/IMG/ABG) as factors. BDae-ASRS and BMQRQ scores were included as additional covariates. Statistical maps were thresholded at a voxel-level uncorrected $P < 0.005$ threshold and standard SPM FWE cluster-level correction based on RFT with a $P < 0.05$ was used. For each significant comparison, the minimum size for a cluster to be corrected for multiple corrections (FWEc) is reported. Finally, in order to determine the link between the behavioral outcome and the VBM-FC results, correlation analyses (Pearson, two-tailed, FDR-corrected) were performed between changes (T1-T0, T2-T0) in language and verbal memory and the clusters showing volume or FC changes between the groups.

**Results**

The participant flow in the two studies is presented in Figure 1. Overall, adherence was very good and similar between the trial sites, with 83 of 90 (92%) patients (Turku: 90%, Helsinki: 95%) completing the study up to the 3-month stage (T1) and 81 of 90 (90%) patients (Turku: 88%, Helsinki: 93%) up to the 6-month stage (T2). The pooled data from the Turku and Helsinki studies were analyzed to determine the short-term (from T0 to T1) and long-term (from T0 to T2) effects of the music intervention.

**Group comparability at baseline**

At T0, there were no statistically significant differences between VMG, IMG, and ABG in most demographic and clinical characteristics and prestroke leisure activities (Table 2). Prestroke music listening frequency showed a group difference [Kruskal–Wallis $H = 11.81$, $P = 0.003$], with more prestroke music listening in IMG than in VMG ($P = 0.007$) or ABG ($P = 0.010$). Also the proportion of amusics patients differed between the three groups [$\chi^2(2) = 9.29$, $P = 0.010$], with less amusics in IMG than in VMG ($P = 0.034$) or ABG ($P = 0.001$). These baseline differences were considered to be due to chance as there is no reason to assume any systematic bias between the groups. The groups were comparable in the behavioral outcome measures at T0 (Tables 3 and 4).

**Group comparability during the follow-up**

The amount of motor, speech, or cognitive rehabilitation received by the patients at T1 and T2 was comparable between the groups (Table 5). The frequency of music and audiobook listening differed highly significantly between the groups, both during the intervention and follow-up periods. Listening frequency was higher for music in VMG and IMG than in ABG, and for audiobooks in ABG than in VMG and IMG at T1 ($P < 0.001$ in all) and, to a lesser extent, at T2 ($P < 0.092$ in all). There were no significant differences in music or audiobook listening between VMG and IMG. The average amount of daily listening to the allocated material was 1.8 h (SD = 0.9), totaling around 100 h ($M = 107.9$, $SD = 55.1$) over the 2-month intervention period. Even though at group level the listening frequencies followed the study protocol, there was a significant difference between the groups in cross-listening (using own devices to listen to material not part of the protocol: music in ABG, audiobooks in VMG and IMG), indicative of treatment contamination, during the intervention period ($F = 42.64$, $P < 0.001$), with ABG showing more cross-listening than VMG and IMG ($P < 0.001$ in both).
Effects of music listening on behavioral recovery

Longitudinal results of the cognitive outcome measures are shown in Figure 2, Table 3, and Figure S1. In the primary outcome measure (verbal memory), there were significant Time × Group interactions from T0 to T1 \( [F(2,67) = 6.96, P = 0.002, \eta^2 = 0.172] \) and from T0 to T2 \( [F(2,66) = 7.10, P = 0.002, \eta^2 = 0.177] \). Post hoc testing (Bonferroni-corrected) showed that verbal memory improved more in VMG than in IMG from T0 to T1 \( (P = 0.015) \) and more in VMG than in ABG from T0 to T2 \( (P = 0.046) \). In the secondary outcome measures, language showed significant Time × Group \( [F(2,58) = 5.12, P = 0.009, \eta^2 = 0.150] \) and Time × Group × Aphasia \( [F(2,58) = 3.42, P = 0.039, \eta^2 = 0.105] \) interactions from T0 to T1. A post hoc subgroup analysis with separate mixed-model ANOVAs within aphasics and non-aphasics showed that the Time × Group interaction was significant only in the...
Aphasics $F(2,18) = 12.82, P < 0.001, n_{p}^2 = 0.588)$. Post hoc testing (Bonferroni-corrected) within the aphasics showed that language recovery was better in VMG than in ABG ($P = 0.002$). There were no significant effects in focused attention (hits and RTs) or in the POMS scales (see Table 4 and Fig. S2).
Figure 2. Significant changes in language skills and verbal memory and (mean ± SEM) from acute to 3-month (gray bars) and 6-month (white bars) poststroke stage are shown in the three groups. Upper panel shows results across all patients and lower panel within non-aphasic and aphasic patients. The significance of the Time × Group interaction is shown with asterisks. ABG = Audiobook group, IMG = Instrumental music group, T0 = baseline (acute), T1 = 3-month stage, T2 = 6-month stage, VMG = Vocal music group.
Effects of music listening on structural neuroplasticity

In the longitudinal VBM analyses (secondary outcome, see Fig. 3 and Table 6), GMV increased more in VMG than in ABG in one cluster in left temporal [superior (STG), middle (MTG), and inferior (ITG) temporal gyrus] areas across all patients from T0 to T2 (Fig. 3B). In aphasics, WMV increased more in VMG than in ABG in one cluster comprising right medial parieto-occipital [lingual gyrus (LG), cuneus, middle occipital gyrus (MOG)] areas (Fig. 3D), and the increased WMV in this cluster correlated with the improvement in language and verbal memory from T0 to T2 (r = 0.68, P = 0.005 and r = 0.56, P = 0.024). In aphasics, there was also larger GMV increase in the IMG than in the ABG in one cluster in right temporoparietal (MTG, MOG) areas (Fig. 3C), and the increased GMV in this cluster correlated with the improvement in language from T0 to T2 (r = 0.64, P = 0.040).

Figure 3. Pooled voxel-based morphometry (VBM) results from the Helsinki and Turku studies (N = 75). Panel A: Lesion overlay maps of the three groups. Panels: B-D: Significant group differences in VBM from acute (T0) to 6-month (T2) stage in (B) gray matter volume (GMV) across all patients, (C) GMV within aphasics patients, and (D) white matter volume (WMV) within aphasics patients. Data reported in the histograms are mean ± SEM. Correlations to change in language/verbal memory are shown with scatter plots. Results are at P < 0.005 (uncorrected) and only clusters surviving a FWE-corrected P < 0.05 threshold are shown and labeled. ABG = Audiobook group, IMG = Instrumental music group, T0 = baseline (acute), T1 = 3-month stage, T2 = 6-month stage, VMG = Vocal music group. Anatomical abbreviations: CUN = cuneus, ITG = inferior temporal gyrus, LG = lingual gyrus, MOG = middle occipital gyrus, MTG = middle temporal gyrus, STG = superior temporal gyrus.
Table 6. Gray and white matter volume changes (acute to 6-month stage) in VBM analyses.

| Area | MNI coordinates | Cluster size (2017) | FWEc t-value | Volume type | Group contrast | Patients
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VMG &gt; ABG</td>
<td>T1-T0</td>
<td>43.33***</td>
<td>7070</td>
<td>LAN MEM</td>
<td>6.17*</td>
<td>Aphasics VMG</td>
</tr>
<tr>
<td>T2-T0</td>
<td>42.22**</td>
<td>7070</td>
<td>LAN MEM</td>
<td>6.16*</td>
<td>87 9</td>
<td></td>
</tr>
<tr>
<td>GMV</td>
<td>T1-T0</td>
<td>6.00</td>
<td>7070</td>
<td>LAN MEM</td>
<td>6.06</td>
<td>18 18</td>
</tr>
<tr>
<td>WMV</td>
<td>T1-T0</td>
<td>6.00</td>
<td>7070</td>
<td>LAN MEM</td>
<td>6.06</td>
<td>18 18</td>
</tr>
</tbody>
</table>

All results are thresholded at a whole-brain uncorrected p < 0.005 threshold at voxel level.

FWEc is the minimum number of voxels for a cluster to be significant at the FWE-corrected p < 0.05 level, according to SPM standard cluster-level correction based on random field theory and cluster-forming threshold of p < 0.005.

ABG, Audiobook group; BA, Brodmann area; GMV, gray matter volume; IMG, Instrumental music group; LAN, language skills; MEM, verbal memory; VMG, Vocal music group; WMV, White matter volume.

Effects of music listening on functional connectivity

Given that the music-induced GMV/WMV changes were located primarily within the posterior and temporal parts of the DMN, which has been linked also to episodic or verbal memory, we sought to determine as a further secondary outcome if FC changes in the DMN could underlie the cognitive benefits and structural neuroplasticity induced by the vocal music listening (Fig. 4 and Table 7). In the rs-fMRI analysis, which focused on the spatial component of each brain network, VMG showed a larger increase in FC between left temporal (STG, MTG) areas and the rest of the DMN than ABG or IMG from T0 to T2. Moreover, VMG also showed increased FC between right temporal [STG, Heschl's gyrus (HG)] areas and the rest of the DMN than IMG from T0 to T2. In the task-fMRI, there was a significant Time × Group interaction from T0 to T2 [F (1, 25) = 3.73, P = 0.038] in whole network-level DMN engagement in the Vocal condition, with post hoc tests (Bonferroni-corrected) showing a larger increase in VMG than in ABG (P = 0.041). No significant effects were observed in the Instrumental and Speech conditions. Correlation analyses showed that in VMG patients, the increased resting-state FC between the different clusters of the DMN and the left STG/MTG correlated with the improvement in language (T0 to T1: r = 0.78, P = 0.040) and verbal memory (T0 to T2: r = 0.66, P = 0.040). Interestingly, this increase in rs-fMRI connectivity between left temporal regions and the DMN also correlated with the mean DMN engagement during the Vocal condition: the greater the DMN engagement after 6 months while listening to vocal music, the more functionally connected the left temporal lobe was with the DMN (r = 0.60, P = 0.038).

Discussion

The present study set out to verify and extend previous results on beneficial effects of daily music listening on cognitive, emotional, and neural recovery after stroke and to explore whether the vocal (sung) component of music plays a key role in its rehabilitative efficacy. Specifically, we pooled data from two single-blind parallel-group RCTs of stroke patients (total N = 83), comprising our previous trial (N = 38) and a new trial (N = 45), which both had a 6-month follow-up and utilized a combination of cognitive, emotional, and neuroimaging outcome measures. Our main findings were that compared to audiobooks vocal music listening enhanced the recovery of verbal memory and language. An exploratory post hoc subgroup analysis suggested that especially aphasic
Patients may benefit from listening to vocal music as a means to facilitate early language recovery. These positive effects of vocal music listening were coupled with increased structural neuroplasticity, indicated by GMV/WMV changes, in left temporal areas in all patients and in right medial parietal areas in aphasics and with functional neuroplasticity, indicated by resting-state and task-related FC increases in the DMN.

The specific enhancement of language and verbal memory induced by vocal music suggests that the sung (vocal...
component of music listening is crucial for its rehabilitative effect. Conceptually, songs represent an interface between speech and music, binding together lyrics and melody and providing a structured temporal scaffolding framework that facilitates their recall. The close coupling of vocal music and verbal memory is evidenced by behavioral studies of stroke patients showing that (1) verbal material (stories) is learned and recalled better when presented in sung than spoken format\textsuperscript{33} and (2) overt production of verbal material during memory encoding is more effective for later recall when done through singing than speaking.\textsuperscript{35} Even though VMG patients were not instructed to sing along to the songs, it is plausible that listening to the songs may have elicited subvocal processing, which could have covertly trained the phonological loop function of working memory, leading to an enhanced verbal memory recovery. Subvocal training could also underlie the enhancement of language skills induced by the vocal music listening, especially in aphasic patients, as singing-based rehabilitation has been found effective for speech in aphasia.\textsuperscript{21,22}

Vocal music can also enhance vigilance or arousal,\textsuperscript{34} which is likely mediated by emotional factors. Music evokes strong emotions and induces pleasure and rewarding experiences which arise from increased activation of the mesolimbic reward system\textsuperscript{4,35} in which dopamine plays a causal role.\textsuperscript{36} Given that music-induced pleasure and engagement of the limbic network are higher when listening to familiar versus unfamiliar music\textsuperscript{37} and sung versus instrumental music,\textsuperscript{15,16} and that the individual reward value of music mediates its positive effect on episodic memory,\textsuperscript{38} it is possible that the observed positive effect of vocal music on verbal memory recovery is at least partly driven by its intrinsic ability to engage motivation- and reward-related dopaminergic networks.

Although music listening has a general mood-enhancing and stress-reducing effect in daily life,\textsuperscript{39} the effects of music listening on poststroke mood, as measured by POMS, were not significant in the present study when compared to audiobook listening (as shown in Table 4, there was a slight trend for reduced Depression and Confusion in the music groups compared to the ABG at T1, but the group effect did not reach significance). Given also our previous results where the positive effects of music listening on POMS Depression and Confusion were seen only when compared to the standard care control group\textsuperscript{9} as well as the results of Baylan et al.\textsuperscript{11} where the effects of music and audiobook listening did not differ on another clinical mood scale (Hospital Anxiety and Depression Scale), the impact of daily music listening for enhancing mood after stroke is still unclear, at least when compared to another stimulating recreational activity.

Our previous exploratory VBM findings indicated that in left hemisphere-lesioned stroke patients (N = 23), music listening increased GMV in left and right superior frontal gyrus, left anterior cingulate, and right ventral striatum compared to audiobooks and standard care.\textsuperscript{10} Using a larger sample (N = 75) and more rigorous statistical criteria (FWE-correction), the present results showed that compared to audiobooks, vocal music listening specifically increased GMV in left temporal areas (STG, MTG, ITG) areas across all patients. Stronger activation in left temporal regions has been reported in previous fMRI studies of healthy subjects comparing song and speech listening.\textsuperscript{13,14} These regions also play a crucial role in perceiving the spectrotemporal structure of sounds\textsuperscript{40} as well as in the combinatorial processing of lexical, phonological, and articulatory features of speech\textsuperscript{41} and also in verbal working memory.\textsuperscript{42,43} In aphasic patients, increased GMV/WMV induced by vocal or instrumental music listening was seen in right medial parieto-occipital areas (LG, cuneus, MOG) and in posterior temporal (MTG) areas, which have been linked to music and speech perception\textsuperscript{44} and memory-related visual imagery.\textsuperscript{45} Importantly, the volume changes in these posterior temporal and parietal regions correlated with enhanced recovery of language and verbal memory, which is in line with previously reported therapy-induced changes in aphasia in

### Table 7. Functional connectivity changes (acute to 6-month stage) in fMRI analyses.

<table>
<thead>
<tr>
<th>Group contrast</th>
<th>Area</th>
<th>MNI coordinates</th>
<th>Cluster size</th>
<th>FWEc</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMG &gt; ABG</td>
<td>Left superior temporal gyrus</td>
<td>−44 to −34 4</td>
<td>342</td>
<td></td>
<td>4.52*</td>
</tr>
<tr>
<td>VMG &gt; IMG</td>
<td>Left middle temporal gyrus</td>
<td>−56 to −40 4</td>
<td>332</td>
<td>296</td>
<td>4.48*</td>
</tr>
</tbody>
</table>

All results are thresholded at a whole-brain uncorrected P < 0.005 threshold at voxel level. FWEc is the minimum number of voxels for a cluster to be significant at the FWE-corrected P < 0.005 level, according to SPM standard cluster-level correction based on random field theory and cluster-forming threshold of P < 0.005.

ABG, Audiobook group; IMG, Instrumental music group; VMG, Vocal music group.

*P < 0.05 FWE-corrected at the cluster level.
Notably, the VMG > ABG effects were partly driven by a reduction of GMV in ABG in these clusters. After stroke, spared brain regions can show both volume increase, which indicates recovery-related neuroplasticity, and volume decrease, which indicates atrophy and is associated with poor recovery and lower gains induced by rehabilitation. It is possible that the large-scale neural activation induced by music listening after stroke can have a long-term neuroprotective impact by preventing atrophy in cortical areas most strongly activated by songs.

Music listening induced also long-term FC changes in the DMN. As the DNM is linked to emotional processing, self-referential mental activity, and the recollection of prior experiences, it is strongly engaged also when listening to music, especially when it is familiar and self-preferred. In line with this, VMG showed larger increase in FC in the whole DMN during vocal music (but not instrumental music or speech) listening than ABG, indicating functional neuroplasticity specific to the type of stimulus and intervention. Importantly, also resting-state FC in the left temporal (STG/MTG) areas of the DMN increased more in VMG than in ABG or IMG, and correlated with the improved recovery of language and verbal memory. Previously, reduced DMN connectivity has been associated with verbal memory impairment in aging and after stroke and increased DMN connectivity with successful stroke recovery. Together, the VBM and FC results provide compelling evidence that the rehabilitative effect of vocal music is underpinned by both structural and functional plasticity changes in temporoparietal networks crucial for emotional processing, language, and memory.

The present study has some methodological limitations, which should be taken into account when evaluating its findings. First of all, the study was not a single RCT but a pooled analysis of two RCTs, one with randomization to three (VMG/IMG/ABG) groups (Turku) and the other with randomization to two groups (MG/ABG) and then post hoc reclassification to three (VMG/IMG/ABG) groups (Helsinki). The results may therefore include a slight self-selection bias, more so between the two music groups and less so between the music groups and the ABG. However, given that the study design of the two trials was otherwise highly similar and that all outcome measure results were covaried for trial site, we do not feel that this represents a significant bias. Regarding power, due to the combined sample size of the two studies, the pooled analysis had greater test power than each of the individual studies. Using G^2Power, we performed a post hoc calculation of the achieved power based on the effect sizes of the original Helsinki study. This showed that the pooled sample yielded 97% power for the primary outcome (verbal memory) and 76%–87% power for the secondary outcomes (language skills, focused attention, POMS Depression, POMS Confusion) to detect a significant change between groups from T0 to T1, suggesting that the study was sufficiently powered. The effect sizes in the present study for the efficacy of music listening on verbal memory and language recovery were of medium level, reaching a large level for the efficacy of vocal music listening on language recovery in aphasics. Second, owing to the relatively small sample sizes in the subgroup analyses of aphasic patients (N = 29) and in the fMRI analyses (N = 35), their results should be considered still somewhat tentative and need to be confirmed with larger studies. Especially studies of aphasic patients with more varying severity levels and aphasia subtypes are warranted. Also uncovering if and how different demographic and clinical background factors, such as pre-stroke music listening and amusia which were included as covariates in the analyses, mediate the efficacy of music listening need to be explored in a larger trial, as this could pave the way toward more individualized use of music listening in stroke rehabilitation.

Clinically, the findings address a vital issue of how the patient environment can be optimized for recovery during the first weeks after stroke when typically over 70% of daily time is spent in nontherapeutic activities even though this time-window is ideal for rehabilitation from the standpoint of neuroplasticity. Corroborating previous findings, the present study provides further evidence for the use of music listening as an effective, easily applicable, and inexpensive way to support cognitive recovery after stroke. Importantly, our results show for the first time that the vocal (sung) component of music is driving its rehabilitative effect on verbal memory and that vocal music can also speed up language recovery in aphasia during the first 3 months. The reason why the positive effect of vocal music occurs particularly at the early post-stroke stage is likely linked to the dynamic pattern of language reorganization in aphasia, where upregulation of both left and right frontotemporal regions takes place at the early (first weeks and months) recovery stage, followed by more pronounced reorganization of perilesional left regions at the chronic (6-month) stage. It is plausible that vocal music engages and stimulates the bilateral frontotemporal network more extensively than audiobooks, leading to a better language recovery in aphasia at the early stage, whereas the effects begin to level off at the chronic stage when left hemisphere mechanisms (engaged more evenly by vocal music and audiobooks) become more dominant. Although more research is still needed to verify the effects of vocal music listening on aphasia, our novel findings suggest that it could perhaps be used to supplement speech therapy, which is often difficult to
implement at the early poststroke stage due to severity of symptoms, general fatigue, and lack of rehabilitation resources.

**Acknowledgments**

Financial support for the work was provided by the Academy of Finland (grants 257077, 277693, 299044), Tyks Research Funding (grant 13944), Finnish Brain Research and Rehabilitation Foundation, Finnish Brain Foundation, Finnish Cultural Foundation, National Doctoral Programme of Psychology, Ella and Georg Ehrnrooth Foundation, Signe and Ane Gyllenberg Foundation, Maire Taponen Foundation, and Jenny and Antti Wihuri Foundation. We warmly thank the patient subjects for their participation in the study and the staff members of Tyks Department of Clinical Neurosciences, HUCH Department of Neurology, and other neurological rehabilitation units in Helsinki and Turku area for their generous help and collaboration.

**Author Contributions**


**Conflict of Interest**

Nothing to report.

**References**


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

Figure S1. Bar charts displaying the cognitive domain scores (mean ± SD) of the patients at the acute (T0), 3-month (T1), and 6-month (T2) poststroke stages.

Figure S2. Bar charts displaying the Profile of Mood States (POMS) scale scores (mean ± SD) of the patients at the acute (T0), 3-month (T1), and 6-month (T2) poststroke stages.