This article has been accepted for publication in JNNP following peer review.
The definitive copyedited, typeset version is available online at 10.1136/jnnp

Efficacy of spoken word comprehension therapy in patients with chronic aphasia: A cross-over randomised controlled trial with structural imaging

*Victoria Fleming 1
*Sonia Brownsett 3,11
Anna Krason 4
Maria Andree Maegli 5
Henry Coley-Fisher 1
Yean-Hoon Ong 1
Davide Nardo 6
Rupert Leach 1
David Howard 7
Holly Robson 8
Elizabeth Warburton 9
John Ashburner 10
Cathy J. Price 10
Jenny Crinion 2
Alexander P. Leff 12
***Victoria Fleming and Sonia Brownsett contributed equally to this manuscript**

Affiliations

1 Department of Brain Repair and Rehabilitation, UCL Queen Square Institute of Neurology, UK
2 Institute of Cognitive Neuroscience, University College London, UK
3 School of Health and Rehabilitation Sciences, The University of Queensland, Australia
4 Department of Psychology and Language Sciences, University College London, UK
5 Department of Psychology, Universidad del Valle de Guatemala, Guatemala
6 MRC Cognition and Brain Sciences Unit, University of Cambridge, UK
7 School of Education, Communication and Language Sciences, Newcastle University, UK
8 Psychology and Clinical Language Sciences, University of Reading, UK
9 Department of Medicine, University of Cambridge, UK
10 Wellcome Centre for Human Neuroimaging, University College London, UK
11 Centre of Research Excellence in Aphasia Recovery and Rehabilitation, Australia

Corresponding author

Victoria Fleming, Victoria.fleming@ucl.ac.uk
Institute of Cognitive Neuroscience, 17-19 Alexandra House, Queen Square, London WC1N 3AR

ABSTRACT

Objective

The efficacy of spoken language comprehension therapies for persons with aphasia remains equivocal. We investigated the efficacy of a self-led therapy app, 'Listen-In', and examined the relation between brain structure and therapy response.

Methods

A cross-over randomised repeated measures trial with five testing time points (12-week intervals), conducted at the university or participant's homes, captured baseline (T_1), therapy (T_2-T_4), and maintenance (T_5) effects. Participants with chronic post-stroke aphasia and spoken language comprehension impairments completed consecutive Listen-In and standard care blocks (both 12-weeks with order randomised). Repeated measures ANOVAs compared change in spoken language comprehension on two co-primary outcomes over therapy versus standard care. Three structural MRI scans (T_2-T_4) for each participant (subgroup, n=25) were analysed using cross-sectional and longitudinal voxel-based morphometry.

Results

Thirty-five participants completed, on average, 85 hours (IQR=70-100) of Listen-In (therapy first, n=18). The first study-specific co-primary outcome (Auditory Comprehension Test, ACT) showed large and significant improvements for trained spoken words over therapy versus standard care (11%, Cohen's *d*=1.12). Gains were largely maintained at 12 and 24 weeks. There were no therapy effects on the second standardised co-primary outcome (Comprehensive Aphasia Test: Spoken Words and Sentences). Change on ACT trained words was associated with: volume of pre-therapy right hemisphere white matter; and post-therapy grey matter tissue density changes in the right temporal lobe.

Conclusions

Individuals with chronic aphasia can improve spoken word comprehension many years after stroke. Results contribute to hemispheric debates implicating the right hemisphere in therapy-driven language recovery. Listen-In will soon be available on GooglePlay.

INTRODUCTION

Auditory spoken language comprehension impairments are common in individuals with aphasia [1], and for many, these remain chronic. Those with severe spoken language comprehension impairments have worse outcomes and higher drop-out rates from rehabilitation programs [2]. Developing effective evidence-based therapies for such impairments is a priority.

A recent Cochrane review, and two further randomised controlled trials (RCTs), have provided unequivocal evidence that speech and language therapy (SLT) benefits people with aphasia across multiple language domains [3–5]. However, surprisingly few investigations have focused on spoken language comprehension therapies. Several single cases and small group studies have investigated several different approaches (*e.g.,* phonological discrimination training [6,7], semantic-based therapies [8], or a mixture of both [9]) with inconclusive results. These equivocal findings are unsurprising given the overall small numbers of participants and variability of therapy tasks employed.

The majority of these studies delivered small doses of therapy (under 20 hours in total). However, reviews of speech and language intervention studies for PWA have revealed that high dose therapies are associated with superior outcomes (~100 hours) [5,10]. These single case and small group interventions have typically targeted a specific aspect of auditory comprehension (*e.g.,* phonological discrimination). As spoken language comprehension necessitates effective integration of acoustic, phonological, and semantic processing, therapies that only target specific aspects may fail to adequately capture the interactive nature of the system. RCTs of high dose spoken language comprehension therapy are needed, and the absence of such trials motivated this study.

In Phase 1 we collaborated with persons with aphasia (PWA) and game developers to create a novel gamified therapy application (‘Listen-In’) that would support self-led high

dose therapy by PWA (submitted elsewhere). The therapeutic component of Listen-In involved massed practice of spoken word-to-picture matching (SWPM) challenges that captured the spoken language comprehension system more broadly by incorporating phonological processing (*e.g.,* minimal pair foils) and semantic processing (*e.g.,* semantic foils) as well as the processing of spoken words within more naturalistic auditory scenes (*i.e.,* within phrases and sentences, and within background noise). Here, we present the results of the Phase 2 RCT of Listen-In. We hypothesised that Listen-In would improve spoken word comprehension for trained items in persons with chronic post-stroke aphasia, and that generalisation might occur for untrained items through improvements in phonological processing.

Previously, patterns of left-hemisphere brain damage have been associated with lexical-based therapy outcomes in individuals with chronic aphasia [11]; however, there have been no equivalent studies for spoken word comprehension. Given that specific predictions are clearly premature and that some models of speech processing posit a role for the right hemisphere [12], we used a whole-brain analysis with structural MRI (sMRI) to investigate whether the volume of pre-therapy grey and white matter was associated with any therapy effects.

We also wanted to investigate whether Listen-In therapy induced changes in brain structure, given that experience-dependent changes in grey and white matter structure have been reported in response to these types of repetitive training tasks in healthy individuals [13]. Whilst there have been no longitudinal structural imaging studies of spoken language comprehension therapy, two functional imaging studies of auditory therapy in PWA identified functional changes in the auditory cortex, part of the superior temporal gyrus [7,14]. Functional imaging studies in healthy individuals have identified an antero-lateral gradient of speech processing, away from the auditory cortex along the superior temporal

gyrus (STG) [15]. Given that Listen-In engages the speech perception system, we expected that if brain changes were observed they may be located in the region of the STG. In this analysis, we present a novel application of longitudinal sMRI.

MATERIALS AND METHODS

Study design

We used a cross-over randomised repeated measures design, with five evenly-spaced testing time points (T_1-T_5) at 12-week intervals (Fig. 1). Ethical approval was obtained from the National Research Ethics Service, Hampstead Committee (15/LO/0569), and the trial protocol was pre-registered on ClinicalTrials.gov (NCT02540889).

[Figure 1 – study design]

Participants

Participants were recruited between 22 March 2016 and 7 April 2017 from the Predicting Language Outcome and Recovery After Stroke (PLORAS) database [16]; a local outpatient aphasia clinic; and focus groups in Phase 1. Inclusion criteria at the screening assessment were: (i) more than six months post-stroke; (ii) English as a dominant language; (iii) scores below 26/30 for comprehension of Spoken Words and 28/32 for Spoken Sentences on the Comprehensive Aphasia Test (CAT) [17]. Exclusion criteria were: (i) a premorbid significant neurological (*e.g.,* degenerative brain disease) or psychiatric (*e.g.,* major depression) disorder (self-report at screen); (ii) unable to give informed consent. Thirty-eight individuals were enrolled over the course of the study; data are presented for 35 participants who completed both cross-over study blocks (Fig. 2). All participants gave written informed consent before commencing the study. Participants had a mean age of 61 years (*SD*=12 years), and mean time since stroke of 76 months (*SD*=59) (Supplementary Table 3).

[Figure 2 – CONSORT flow diagram)

Randomisation and blinding

A researcher blinded to participants' identities and not involved in enrolment or testing used a randomisation minimisation procedure [18] to allocate participants to therapy first (Group 1, n=18) or standard care (Group 2, n=17) to minimise potential differences between groups. This method considered severity of spoken language comprehension impairment (Comprehensive Aphasia Test, Spoken Word and Sentence subtests [17]) and time since stroke with equal weighting.

Intervention and standard care blocks

The intervention block consisted of 12 weeks of daily self-managed spoken word comprehension therapy (Listen-In) on a computer tablet with a target dose of 100 hours (~80 minutes per day). The standard care block consisted of participants' usual daily activities. The app was developed for the trial, in collaboration with software developers (SoftV) and PWA [19]. Listen-In therapy consisted of spoken word/phrase/sentence-to-picture matching challenges that users progressed through based on an adaptive algorithm (see Supplementary Material). In each challenge, the participant heard a word presented on its own, or within a phrase or sentence (*e.g. 'clock', 'an old clock, 'the man holds a clock'*) (Supplementary Tables 1 and 2). The target picture (*e.g., 'clock'*) was always flanked by two to five foils, that were a combination of semantic (*e.g.,* '*stopwatch'*), phonological (*e.g., 'lock' and 'rock'*), and unrelated (*e.g., 'shoe'*) foils (Supplementary Fig. 1). Time participants spent on therapy challenges (in seconds) was recorded by the app daily; automated shut-down after 3 minutes of inactivity ensured high fidelity data on dose.

Outcomes

The first co-primary outcome assessed spoken language comprehension on two standardised subtests of the CAT [17], Spoken Words and Spoken Sentences. Each subtest contained 15 and 16 spoken word and sentence-to-picture matching trials respectively, where the target was always flanked by three foils. Maximum scores were 30 for Spoken Words and 32 for Spoken Sentences, with each trial scored as follows: correct=2 points; correct but repeated or delayed (>5 seconds)=1 point; incorrect=0 points. Raw scores were converted to a percentage for statistical analyses, and are presented as raw scores in Supplementary Table 4.

A second study specific co-primary assessment was created (prior to the trial commencing) to measure comprehension of trained and untrained spoken words (see Supplementary Material). The Auditory Comprehension Test (ACT) contained 220 spoken word/phrase/sentence-to-picture matching trials ('items'), that always contained one target picture, and five foils (phonological, semantic, and unrelated) (Supplementary Fig. 2). The ACT consisted of two sets of psycholinguistically matched items (Set A/Set B) that formed participants' trained (n=110) and untrained (n=110) items (Supplementary Fig. 3). To reduce possible stimulus-response binding effects (identity priming), all images used in the ACT were different from those used in Listen-In therapy. Verbal stimuli in the ACT and Listen-In were also randomly selected on each trial from a small pool of male and female versions of the audio recording. Participants scored 1 point for a correct response, and this was converted to percentage correct for trained (n=110) and untrained (n=110) items.

Three secondary outcomes measured change in other auditory tasks (phoneme discrimination, environmental sound discrimination, spoken vocabulary comprehension) to assess the specificity of any therapy effects (see Supplementary Material).

Primary and secondary outcomes were collected at every time point (T_1-T_5) by three speech and language therapists and one research assistant. Testing sessions were conducted at

the Institute of Cognitive Neuroscience (UCL) or in participants' homes and were carried out over sessions of 1-2 days within the same week.

Baseline data

Baseline data are presented in Supplementary Table 3. All participants met diagnostic criteria for aphasia in accordance with the CAT for both receptive and expressive subtests. By definition, all participants performed below aphasia cut-off criterion on Spoken Word and Sentence comprehension subtests at a screen prior to T_1. Hearing levels were tested using pure-tone free field audiometry at 1000, 2000, and 4000 Hz, and participants were encouraged to wear hearing aids if prescribed.

MRI data acquisition and lesion identification

Structural MRI scans were obtained for all participants who met safety requirements for scanning (n=25) and were acquired at T_2, T_3, and T_4 on the day of testing or within a week if there were scheduling constraints (mean days between scans: T_2-T_3=85, *SD*=4; T_3-T_4=86 *SD*=6). T1-weighted images were acquired on a Siemens 1.5T Avanto (n=7) or Siemens 3T Trio (n=18) scanner with standard 20 channel head coils using standard acquisition sequences with a voxel resolution of $1mm^3$ (see Supplementary Material).

Figure 3 displays the distribution of lesions across 25 participants. All had extensive damage throughout the left perisylvian middle cerebral artery territory. Three participants had additional right hemisphere (non-aphasic) lesions (prior to their left hemisphere infarct) and one of these participants also had a lesion in the left cerebellum.

[Figure 3 – Lesion overlap map]

Statistical analysis

We conducted an a-priori power calculation, which indicated that 32 participants would enable us to detect a small effect size (Cohen's d_z=0.45) with 80% power using a paired samples t-test (one way alpha 0.05). We anticipated a dropout rate of 20%, and therefore aimed to recruit 38 participants (see Supplementary Material).

Behavioural outcomes

To test the efficacy of Listen-In, and in-line with the crossover study design, we compared change over therapy and standard care for all primary and secondary outcomes. Change was calculated as raw percentage change over study blocks (*i.e.*, Group 1 therapy [$\Delta(T_3\text{-}T_2)$], standard care [$\Delta(T_4\text{-}T_3)$]; Group 2 therapy [$\Delta(T_4\text{-}T_3)$]; standard care [$\Delta(T_3\text{-}T_2)$]). For the ACT, we used repeated-measures ANOVA with two within-subject factors: change (Δ therapy, Δ standard care) and item condition (trained/untrained). For all other outcomes, one-way repeated measures ANOVAs were conducted. All ANOVAs included Group as a between-subjects factor. Planned *post-hoc* paired samples t-tests (two-tailed) explored significant interactions and assessed change over baseline and maintenance. The alpha-level for all analyses was $p < 0.05$.

Cross-sectional voxel-based morphometry

To investigate whether the pre-therapy brain structure was associated with response to therapy, we used whole-brain voxel-based morphometry (VBM) (see Supplementary Material) [20]. This produced two tissue images (grey and white matter) for each participant at baseline (n=25) that were entered into two multiple linear regression models to identify regions where pre-therapy volume correlated with significant behavioural therapy effects. Effects of age, time since stroke, and lesion volume were included as covariates of no interest. The statistical voxel-level threshold was set at a standard level of p <0.001 with

voxel and cluster-level significance set at $p<0.05$ after family-wise error (FWE) correction for multiple comparisons across the whole search volume.

Longitudinal voxel-based morphometry

To investigate possible changes in brain structure associated with therapy, we explored whether behavioural therapy effects covaried with any changes in tissue density. First, we quantified change in tissue density (using unprocessed whole-brain scans) over therapy (Group 1 [$\Delta(T_3-T_2)$]; Group 2 [$\Delta(T_4-T_3)$]) and standard care blocks (Group 1 [$\Delta(T_4-T_3)$]; Group 2 [$\Delta(T_3-T_2)$]) for each participant. To carry out the aforementioned process, we used serial longitudinal registration in SPM12 [21] (for methodology see Supplementary Material). Resultant probabilistic change images were subtracted within-subject (therapy minus standard care) to produce one final image for each participant that represented a change in voxel density over the therapy block, over and above the period of standard care. Finally, images were normalised to MNI space and smoothed with an isotropic kernel of 6mm FWHM.

These images were entered into simple linear regression models in SPM12 to identify any regions where a change in tissue density correlated with significant behavioural therapy effects. No other regressors were modelled as participants acted as their own control. Our statistical voxel-level threshold was $p<0.001$ with voxel and cluster-level significance set at $p<0.05$ after FWE correction for multiple comparisons across the whole brain, and within a bilateral STG mask (WFU PickAtlas toolbox, voxels=10496).

RESULTS

There were no significant differences between Group 1 and Group 2 for age, time since stroke, or CAT Spoken Words and Spoken Sentences at baseline (T_1), in line with the

minimisation randomisation method (all $p \geq 0.20$). On the ACT, Group 1 scored significantly better than Group 2 at baseline ($t(33)=2.52$, $p=.02$). Between T_1 and T_2 (baseline change) performance on CAT Spoken Words and Sentences remained stable across the whole group (both $t \leq 1.75$, $p \geq 0.09$). For the ACT, there was a small significant improvement between T_1 and T_2 (M=2%, *SD*=6%, $t(34)=2.37$, $p=0.02$) (for reliability analysis see Supplementary Fig. 4).

Listen-In dose

Participants spent an average of 85 hours (interquartile range (IQR)=70-100) on Listen-In therapy over 12-weeks, completed ~27,000 individual challenges each (IQR=21,000-33,000), and of these 18% (IQR=14-21%) targeted words probed within the ACT (trained items). For each trained word, participants completed an average of 43 (IQR=35-52) challenges.

Behavioural therapy effects

Post-hoc analyses indicated this study was sensitive to small therapy effects (Cohen's $\alpha=0.05$; $1-\beta=0.80$, $d_z=0.49$, see Supplementary Material). Results for all outcomes are presented in Supplementary Table 4.

There was no significant difference in performance change between therapy (Group 1 [$\Delta(T_3-T_2)$]; Group 2 [$\Delta(T_4-T_3)$]) and standard care (Group 1 [$\Delta(T_4-T_3)$]; Group 2 [$\Delta(T_3-T_2)$]) for CAT Spoken Words ($F(1, 33)=1.87$, $p=0.18$, $\eta^2=0.05$) and Spoken Sentences ($F(1, 33)=0.52$, $p=0.48$, $\eta^2=0.02$). For the ACT, there was a significant interaction between block (Δ therapy, Δ standard care) and item condition (trained items, untrained items) ($F(1, 33)=39.16$, $p<0.001$, $\eta^2p=0.54$) (Fig. 4). This was driven by improvements for trained items (Fig. 4A) over therapy versus standard care ($t(34)=4.09$, $p<0.001$, Cohen's $d_z=1.32$). For trained items, from pre to post therapy, effect sizes were large for both unstandardised and

standardised measures: 11% (absolute change), 30% (%back2normal); and Cohen's d=1.12 (see Supplementary Material). Figure 4B shows the variability in magnitude of therapy effects across participants. There was a small but non-significant loss of therapy effects for Group 1 at 12 weeks (T_4-T_3, t(17)=-2.01, p=0.06). For the remaining maintenance periods (Group 1 T_5-T_4; Group 2 T_5-T_4, n=15) there were no significant declines in performance (both $p \geq 0.70$). All secondary outcomes showed no significant changes in performance over therapy versus standard care (one way ANOVAs all $p \geq 0.11$).

[Figure 4 – Behavioural therapy effects]

Pre-therapy brain structure and response to therapy

Two multiple linear regression models were conducted separately for grey and white matter with change on trained items as the behavioural variable (Group 1 [$\Delta(T_3$-$T_2)$]; Group 2 [$\Delta(T_4$-$T_3)$]) and age, time since stroke, and lesion volume as covariates of no interest (Fig. 5A). Figure 5B displays significant clusters of voxels where greater volume at baseline covaried with greater response to therapy. In grey matter, one significant cluster was identified in the right head of the caudate nucleus (Cluster 1). In white matter, four significant clusters were identified. In the right hemisphere, these were Cluster 3 (white matter intrinsic to the temporal lobe) and Cluster 4 (white matter extending from the fusiform cortex to subcortical structures). In bilateral hemispheres, these were Cluster 2 (right ventrolateral prefrontal cortex) and Cluster 5 (subcortical white matter, mainly in the left body of the cerebellum).

Region	Side	Coordinates (mm)			Cluster size, kE	T
		x	y	z		
Cross-sectional VBM						
Grey matter:						
Right head of caudate nucleus	R	12	12	0	435	4.94
White matter:						

Temporal lobe	R	64	-12	-26	1589	4.11
Ventrolateral prefrontal cortex	B	44	32	6	2415	5.39
Deep white matter (fusiform cortex to thalamus)	R	30	-38	-8	2050	4.90
Bilateral cerebellum	B	-6	-62	-28	822	5.54
Longitudinal VBM						
Grey matter:						
Right posterior superior temporal gyrus	R	66	-40	12	154	5.05
White matter:						
Left middle superior temporal gyrus	L	-56	-22	-4	106	4.59

Table 1. Brain regions associated with response to Listen-In therapy. Coordinates are in MNI space and represent the peak voxel in that cluster. FWE=family wise error. Voxel wise threshold set at $p<0.001$, $p<0.05$ FWE correction following small volume correction with a bilateral superior temporal gyrus mask (voxels=10496). B=bilateral; L=left; R=right; kE=cluster size (voxels); T=t-value.

[Figure 5 – Cross-sectional VBM results]

Change in brain structure induced by therapy

No clusters or peaks survived family-wise error correction ($p<0.05$) across the whole brain. Figure 6 shows significant cluster level results from white (Fig. 6A) and grey (Fig. 6B) matter VBM analyses, following small volume correction with a bilateral STG mask and correction for multiple comparisons (Table 1). Two clusters were identified: grey matter in the right posterior STG, and white matter in the left middle STG bordering the superior temporal sulcus. The scatterplots show individual participant results for change in performance on trained items by the change in relative tissue density. (For overall density change over time (T_1-T_4) see Supplementary Fig. 5).

[Figure 6 – Longitudinal VBM results]

DISCUSSION

Behavioural therapy effects

We present the first cross-over RCT of spoken language comprehension therapy, using a novel tablet-based therapy app, Listen-In. We observed large, item-specific improvements in spoken word comprehension when comparing the therapy block with an equivalent period of no therapy. Participants achieved a large average dose of therapy, much larger than previous speech comprehension studies [6–9], and much closer to the large doses observed in positive intervention studies [10]. The effect size compares favourably with other lexical based interventions in reading [22] and naming [23]. Importantly, considering the amount of time and effort participants contributed, improvements for trained items were maintained up to 24 weeks after the therapy blocks were completed. Participants varied in the severity of their spoken language comprehension impairments, suggesting Listen-In therapy is suitable for a range of PWA.

Improvements did not generalise to untrained items, consistent with item-specific effects frequently observed in lexical based therapies [24]. However, while effects were item specific, they were not exemplar specific, as images and verbal stimuli varied between therapy and the assessment task (ACT), such that participants were able to generalise their improvements to novel exemplars of that word. This RCT demonstrates that item-specificity should be expected for some interventions, and necessitates that clinicians tailor 'to be trained' items to each individual.

Role of pre-therapy brain structure

We investigated whether variation in response to therapy could be partly accounted for by differences in pre-therapy brain structure. With the exception of one cluster (head of the caudate), the significant clusters were located in white matter underlying frontal and

temporal cortices, as well as subcortical structures, suggesting that greater integrity of white matter networks facilitated participants' response to therapy.

It is not possible to distinguish whether variability in volume is due to normal pre-morbid individual differences in white matter volume, neurobiological sequelae of stroke, or, most likely, a mixture of both. Pre-morbid possibilities include differences in cognitively active lifestyles, white matter asymmetries [25], and declines in volume due to normal ageing [26]; post-stroke possibilities include accelerated stroke-related atrophy [27] and Wallerian (dying back or retrograde degeneration of white matter connections) and transcallosal (trans-synaptic, anterograde degeneration of white matter connections) degeneration. It is also possible that post-stroke reorganisation may play some role, as increased right hemisphere grey matter volume has been reported in chronic aphasic participants [28]. Regardless of the underlying cause of variability, these findings demonstrate a link between the neuroanatomical regions identified in our data and therapy-related language recovery.

The role of the right hemisphere in language recovery is a topic of much debate [29]. A common finding is increased right hemisphere activity in regions homologous to the lesioned left hemisphere [30]. However, the nature of this activity is widely debated. Three prominent accounts are adaptive compensation [31], upregulation of domain-general neural networks rather than language-specific reorganisation [32], and maladaptive interference due to release of transcallosal disinhibition [33]. The majority of significant clusters in our data were located in the right hemisphere, in frontal and temporal regions homotopic to key left hemisphere language nodes. Given that participants with greater volume in these regions made greater gains in spoken word comprehension, our interpretation is that greater volume reflects a better functioning system in accordance with a facilitatory account of right hemisphere involvement (*e.g.*, upregulation of support mechanisms, or reorganisation).

We also identified subcortical structures known to support and monitor language-specific cortical activity. These include the head of the caudate nucleus, a region with extensive cortical connectivity [34] and implicated in cognitive control of language in the dominant hemisphere [35]; deep white matter which may relate to language monitoring functions that have been ascribed to some of the basal ganglia-thalamocortical circuits [36]; and the cerebellum, which is involved in a range of motor, cognitive and language tasks [37].

Therapy related changes in tissue density

We identified two clusters where change in tissue density covaried with a change in trained item performance. In the left hemisphere, we identified a white matter cluster underlying the middle STG. As the scatterplot illustrates (Fig. 6C), many participants show a lesion at this location, and the remaining correlation is being driven by two outlying participants who show striking declines in density relative to the group. For these reasons, any interpretation is highly speculative. However, this cluster is located in close proximity to part of the cortex implicated in phonetic speech processing and learning new words in healthy individuals [15,38,39], a plausible region to predict neuroplasticity given the demands of Listen-In.

In the right hemisphere, we observed a grey matter cluster in the posterior STG (Wernicke's area homologue), a region proposed to be part of bilateral and parallel early speech processing streams [12,38]. When the data from this region was plotted (Fig. 6D), two types of participants were identified: those with a relative increase in tissue density during therapy and who responded well (upper right quadrant); and those with a relative decrease in tissue density and who responded poorly (lower quadrants). The relative changes in tissue density within the right hemisphere are of a small magnitude and are on a background of the global, and somewhat accelerated, atrophy seen in individuals with stroke (~1% per year)

[27]. Given this background of likely atrophy over time, maintenance of tissue density, or lesser decline, may well be regarded as a demonstration of experience-dependent plasticity in this particular population [40]. Further studies are needed to specifically address this mechanism. Nevertheless, we have demonstrated more broadly, for the first time, a relationship between the structural integrity of this right hemisphere region and therapy-induced improvements in understanding spoken language after stroke.

Limitations and future directions

Whilst some may consider lack of generalisation (to untrained words) to diminish the clinical utility of Listen-In, it is important to note that participants showed generalisation across different exemplars of trained items, and improvements were maintained as long as 24 weeks later. As relatively little research has been reported for spoken language comprehension therapies, improvements, whether item specific or otherwise, are an important step forward for aphasia rehabilitation. Future work is now required to investigate possible mechanisms of generalisation, to maximise benefits for PWA . We have a manuscript in preparation that will investigate one aspect of this.

It is also unclear how task-based item-specific improvements may relate to improved spoken word comprehension in everyday environments. This is a methodological challenge given the often hidden nature of comprehension. Future research may investigate how improvements generalise across tasks and settings.

We aimed to investigate the efficacy of high dose (hours) therapy, as well as its' feasibility. One limitation is that early phase dosing investigations were not conducted prior to the start of the trial. These may have helped to identify a different, perhaps more suitable, target dose. Listen-In would benefit from employing a systematic therapy development pipeline to specify the suitable parameters for delivery.

This efficacy study showed improvements in the context of a well-controlled RCT with a highly specific population. The effectiveness of Listen-In now needs to be evaluated in the real world setting, and a pragmatic trial will be soon be underway.

CONCLUSION

We report the first RCT of high-dose spoken language comprehension therapy in individuals with chronic aphasia using a novel, self-led tablet-based spoken word comprehension therapy 'Listen-In'. These results demonstrate that individuals with chronic aphasia can significantly improve their comprehension of spoken words many years after stroke, given a sufficient dose. Future work is needed to investigate optimal dose, intensity, and distribution, as well as generalisation of improvements to communication environments. Structural imaging results add to the topical debate of hemispheric contributions to language recovery, by providing novel indirect evidence that right-hemisphere regions support the recovery of language in PWA. Researchers, clinicians, and individuals with aphasia can access the public release of Listen-In soon, with a free trial period: www.ucl.ac.uk/ion/Listen-In.

FIGURE LEGENDS

Figure 1. Group randomised cross-over study design.

Figure 2. Consolidated Standards of Reporting Trials (CONSORT) flow diagram.

Figure 3. Lesion overlap map for 25 participants in MNI space. *Numbers represent MNI coordinates of brain slices along the x-axis. The colour bar represents the number of participants with a lesion at that location, from 1 to 25.*

Figure 4. Therapy effects on spoken word comprehension ability. *(A) Performance on the ACT, for trained and untrained items, by group. G1=Group 1; G2=Group 2. Error bars are within-subject standard error of the mean (for G2, two participants data were excluded from all standard error calculations due to no* T_5 *data). (B) Pre and post-therapy performance on ACT trained items, for all participants (n=35).*

Figure 5. Pre-therapy brain regions associated with therapy effects (VBM). *(A) Multiple linear regression design matrix in SPM12; (B) Average grey matter slices from 25 participants, showing grey (blue) and white (yellow) matter clusters where volume at baseline positively covaried with change on trained items (%) from pre- to post-therapy. Voxel-wise threshold set at p<0.001. All clusters survived FWE correction at p<0.05. Numbered regions are as follows: (1) right caudate nucleus; (2) right ventrolateral prefrontal cortex, genu of the corpus callosum, anterior corona radiata; left: paracingulate cortex; (3) temporal lobe; (4) cluster extending from the right fusiform cortex to thalamus; (5) bilateral cerebellum.*

Figure 6. Changes in brain structure associated with therapy effects (VBM). *Average grey matter slices from 25 participants showing regions of (A) white and (B) grey matter where a change in tissue probability over therapy significantly covaried with change on trained items (%) from pre- to post-therapy. Voxel-wise threshold set at p<0.001, with FWE correction at p<0.05, following small volume correction with a bilateral superior temporal gyrus mask. Below, scatterplots show percentage change on trained items from pre- to post-therapy plotted against change in tissue density for (C) white and (D) grey matter clusters. Tissue density values taken as an average of a 5mm sphere around the peak voxel in each cluster. Unfilled circles are participants with a lesion at the peak voxel in that cluster. Note: on the y-axis, a negative value reflects a relative decrease in density, while a positive value reflects a relative increase (where relative change is within-subject change in tissue density over therapy more than standard care).*

ACKNOWLEDGEMENTS

We would like to thank all of the participants who helped complete this study. We would also like to acknowledge SoftV who developed the Listen-In software.

FUNDING

This study is funded by the National Institute for Health Research (NIHR) i4i stream (Grant Reference Number II-LB-0813-20004). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Brain imaging was supported at the Wellcome Centre for Human Neuroimaging (203147/Z/16/Z) with additional funding to C.J.P. (205103/Z/16/Z).

COMPETING INTERESTS

The authors report no competing interests.

DATA AVAILABILITY

Data is available by email request to the Principal Investigator (a.leff@ucl.ac.uk)

REFERENCES

1 Breese EL, Hillis AE. Auditory comprehension: Is multiple choice really good enough? *Brain Lang* 2004;**89**:3–8. doi:10.1016/S0093-934X(03)00412-7

2 Paolucci S, Matano A, Bragoni M, *et al.* Rehabilitation of Left {Brain-Damaged} Ischemic Stroke Patients: The Role of Comprehension Language Deficits. 2005;**20**:400–6. doi:10.1159/000088671

3 Breitenstein C, Grewe T, Flöel A, *et al.* Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *Lancet* 2017;**389**:1528–38. doi:10.1016/S0140-6736(17)30067-3

4 Palmer R, Dimairo M, Cooper C, *et al.* Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (Big CACTUS): a multicentre, single-blinded, randomised controlled trial. *Lancet Neurol* 2019;**18**:821–33. doi:10.1016/S1474-4422(19)30192-9

5 Brady MC, Kelly H, Godwin J, *et al.* Speech and language therapy for aphasia following stroke. *Cochrane Database Syst Rev* Published Online First: 1 June 2016. doi:10.1002/14651858.CD000425.pub4

6 Maneta A, Marshall J, Lindsay J. Direct and indirect therapy for word sound deafness. *Int J Lang Commun Disord* 2001;**36**:91–106.

7 Woodhead ZV, Crinion J, Teki S, *et al.* Auditory training changes temporal lobe connectivity in 'Wernicke's aphasia': a randomised trial. *J Neurol Neurosurg Psychiatry* 2017;**88**:586–94. doi:10.1136/jnnp-2016-314621

8 Morris J, Franklin S. Investigating the effect of a semantic therapy on comprehension in aphasia. *Aphasiology* 2012;**26**:1461–80. doi:10.1080/02687038.2012.702885

9 Woolf C, Panton A, Rosen S, *et al.* Therapy for auditory processing impairment in aphasia: An evaluation of two approaches. *Aphasiology* 2014;**28**:1481–505. doi:10.1080/02687038.2014.931921

10 Bhogal SK, Teasell R, Speechley M. Intensity of Aphasia Therapy, Impact on Recovery. *Stroke* 2003;**34**:987–93. doi:10.1161/01.STR.0000062343.64383.D0

11 Aguilar OM, Kerry SJ, Ong Y-H, *et al.* Lesion-site-dependent responses to therapy after aphasic stroke. *J Neurol Neurosurg Psychiatry* 2018;**89**:1352–4. doi:10.1136/jnnp-2017-317446

12 Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci* 2007;**8**:393–402. doi:10.1038/nrn2113

13 Lövdén M, Wenger E, Mårtensson J, *et al.* Structural brain plasticity in adult learning and development. *Neurosci Biobehav Rev* 2013;**37**:2296–310. doi:10.1016/j.neubiorev.2013.02.014

14 Musso M, Weiller C, Kiebel S, *et al.* Training-induced brain plasticity in aphasia. *Brain* 1999;**122**:1781–90. doi:10.1093/brain/122.9.1781

15 DeWitt I, Rauschecker JP. Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci* 2012;**109**:E505–14. doi:10.1073/pnas.1113427109

16 Seghier ML, Patel E, Prejawa S, *et al.* The PLORAS Database: A data repository for Predicting Language Outcome and Recovery After Stroke. *Neuroimage* 2016;**124**:1208–12. doi:10.1016/j.neuroimage.2015.03.083

17 Swinburn K, Howard D, Porter G. CAT : comprehensive aphasia test. 2004.

18 Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ* 2005;**330**:843. doi:10.1136/bmj.330.7495.843

19 Fleming V, Brownsett S, Crinion J, *et al.* Listen-In: The development and testing of a tablet-based therapy application for patients with impaired speech comprehension

caused by stroke: Phase 1 Development and Testing. *Stem- , Spraak- en Taalpathologie 16th Int Sci Aphasia Conf* 2015;**20**:45–7.
20 Ashburner J, Friston KJ. Voxel-based morphometry - The methods. *Neuroimage* 2000;**11**:805–21. doi:10.1006/nimg.2000.0582
21 Ashburner J. Symmetric diffeomorphic modeling of longitudinal structural MRI. *Front Neurosci* 2013;**6**:197. doi:10.3389/fnins.2012.00197
22 Woodhead Z, Kerry SJ, Aguilar OM, *et al.* Randomized trial of iReadMore word reading training and brain stimulation in central alexia. *Brain* 2018;**141**:2127–41. doi:10.1093/brain/awy138
23 Best W, Greenwood A, Grassly J, *et al.* Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex* 2013;**49**:2345–57. doi:10.1016/j.cortex.2013.01.005
24 Wisenburn B, Mahoney K. A meta-analysis of word-finding treatments for aphasia. *Aphasiology* 2009;**23**:1338–52. doi:10.1080/02687030902732745
25 Catani M, Allin, Matthew PG, Husain M, *et al.* Symmetries in human brain language pathways correlate with verbal recall. *Proc Natl Acad Sci* 2007;**104**:17163–8.
26 Draganski B, Ashburner J, Hutton C, *et al.* Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). *Neuroimage* 2011;**55**:1423–34. doi:10.1016/j.neuroimage.2011.01.052
27 Seghier ML, Ramsden S, Lim L, *et al.* Gradual lesion expansion and brain shrinkage years after stroke. *Stroke* 2014;**45**:877–9. doi:10.1161/STROKEAHA.113.003587
28 Hope TMH, Leff AP, Prejawa S, *et al.* Right hemisphere structural adaptation and changing language skills years after left hemisphere stroke. *Brain* 2017;**140**:1718–28. doi:10.1093/brain/awx086
29 Hartwigsen G, Saur D. Neuroimaging of stroke recovery from aphasia – Insights into plasticity of the human language network. *Neuroimage* 2019;**190**:14–31. doi:10.1016/j.neuroimage.2017.11.056
30 Turkeltaub PE, Messing S, Norise C, *et al.* Are networks for residual language function and recovery consistent across aphasic patients? *Neurology* 2011;**76**:1726–34.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed10&NEWS=N&AN=2011288592
31 Menke R, Meinzer M, Kugel H, *et al.* Imaging short- and long-term training success in chronic aphasia. *BMC Neurosci* 2009;**10**:118. doi:10.1186/1471-2202-10-118
32 Brownsett SLE, Warren JE, Geranmayeh F, *et al.* Cognitive control and its impact on recovery from aphasic stroke. *Brain* 2014;**137**:242–54. doi:10.1093/brain/awt289
33 Breier JI, Juranek J, Maher LM, *et al.* Behavioral and Neurophysiologic Response to Therapy for Chronic Aphasia. 2009;**90**:2026–33. doi:10.1016/j.apmr.2009.08.144
34 Draganski B, Kherif F, Klöppel S, *et al.* Evidence for Segregated and Integrative Connectivity Patterns in the Human Basal Ganglia. 2008;**28**:7143–52. doi:10.1523/JNEUROSCI.1486-08.2008
35 Robles GS, Gatignol P, Capelle L, *et al.* The role of dominant striatum in language: a study using intraoperative electrical stimulations. *J Neurol Neurosurg Psychiatry* 2005;**76**:940–6. doi:10.1136/jnnp.2004.045948
36 Klostermann F, Krugel LK, Ehlen F. Functional roles of the thalamus for language for language capacities. *Front Syst Neurosci* 2013;**7**:1–8. doi:10.3389/fnsys.2013.00032
37 Stoodley CJ, Schmahmann JD. Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *Neuroimage* 2009;**44**:489–501. doi:10.1016/j.neuroimage.2008.08.039
38 Saur D, Schelter B, Schnell S, *et al.* Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *Neuroimage*

2010;**49**:3187–97. doi:10.1016/j.neuroimage.2009.11.009
39 Mårtensson J, Eriksson J, Bodammer NC, *et al.* Growth of language-related brain areas after foreign language learning. *Neuroimage* 2012;**63**:240–4. doi:10.1016/j.neuroimage.2012.06.043
40 Lövdén M, Schaefer S, Noack H, *et al.* Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiol Aging* 2012;**33**:620.e9-620.e22. doi:10.1016/j.neurobiolaging.2011.02.013