



Published in final edited form as:

Stroke. 2009 March ; 40(3): 853–858. doi:10.1161/STROKEAHA.108.532499.

Treating visual speech perception to improve speech production in non-fluent aphasia

Julius Fridriksson, Ph.D.^a, Julie M. Baker, M.A.^a, Janet Whiteside, Ph.D.^c, David Eoute, M.S.P.^a, Dana Moser, Ph.D.^a, Roumen Vesselinov, Ph.D.^b, and Chris Rorden, Ph.D.^a

a Department of Communication Sciences & Disorders, University of South Carolina

b Department of Statistics, University of South Carolina

c Department of Communication Sciences & Disorders, University of Central Florida

Abstract

Background and Purpose—Several recent studies have revealed modulation of the left frontal lobe speech areas not only during speech production, but also for speech perception. Crucially, the frontal lobe areas highlighted in these studies are the same ones that are involved in non-fluent aphasia. Based on these findings, this study examined the utility of targeting visual speech perception to improve speech production in non-fluent aphasia.

Methods—Ten patients with chronic non-fluent aphasia underwent computerized language treatment utilizing picture-word matching. To examine the effect of visual speech perception upon picture naming, two treatment phases were compared – one which included matching pictures to heard words and another where pictures were matched to heard words accompanied by a video of the speaker's mouth presented on the computer screen.

Results—The results revealed significantly improved picture naming of both trained and untrained items following treatment when it included a visual speech component (i.e. seeing the speaker's mouth). In contrast, the treatment phase where pictures were only matched to heard words did not result in statistically significant improvement of picture naming.

Conclusions—The findings suggest that focusing on visual speech perception can significantly improve speech production in non-fluent aphasia and may provide an alternative approach to treat a disorder where speech production seldom improves much in the chronic phase of stroke.

Keywords

Aphasia; speech disorders; stroke recovery; therapy; treatment

Introduction

Patients with chronic non-fluent aphasia experience minimal recovery of speech production,¹ even in the case of language treatment.^{2,3} Behavioral treatment of non-fluent aphasia has traditionally focused on speech production, a behavior that is inherently difficult for this population. Thus, patients get minimal practice at producing speech and, perhaps, experience more negative rather than positive feedback due to repeated failures in treatment.

Impaired speech production has been associated with damage to both Broca's area and the left anterior insula.^{4,5} Several neuroimaging studies have revealed that the same areas implicated in impaired speech production are also recruited during speech perception.^{6–8} Furthermore, Skipper and colleagues⁹ showed increased modulation of the frontal speech areas when participants were able to listen to and see the speaker compared to only hearing the speaker's voice. This evidence suggests an intimate neuroanatomical connection between speech perception and speech production by demonstrating that the perception of auditory and visual speech is associated with increased activity in the cortical speech-motor regions. These findings have potential implications for improving impaired motor function following stroke, particularly in relation to the treatment of speech production in individuals with non-fluent aphasia. That is, it may be possible that training audio-visual speech perception, a task which modulates the frontal motor speech areas in normal participants, will improve speech production in persons with non-fluent aphasia by stimulating the residual speech-motor network. Moreover, by focusing on speech perception, a skill which is relatively spared compared to speech production in patients with non-fluent aphasia,¹⁰ success during treatment could be enhanced and repeated failures at speech production could be minimized.

Using computer-based treatment administration, we examined the effect of using audiovisual speech stimuli compared to audio-only speech stimuli to improve overt picture naming in ten persons with non-fluent aphasia. We hypothesized that persons with non-fluent aphasia would show greater improvement in overt naming when treatment included perceptual matching of pictures paired with audio-visual speech stimuli as compared to when the pictures were paired with audio-only speech stimuli.

Methods

Participants

The inclusion criteria for this study were as follows: 1) non-fluent aphasia classification; 2) single left hemisphere stroke; 3) at least one year post-stroke; 4) pre-morbidly right-handed; and 5) native speaker of English. Ten persons (eight males) who met the above criteria participated in this study. The mean age was 59.2 years (SD = 12.1) with a range of 41 years (Table 1). Aphasia assessment employing the *Western Aphasia Battery* (WAB)¹¹ revealed that all ten participants had a language impairment most consistent with Broca's aphasia – the most common type of chronic non-fluent aphasia.¹² In addition to the WAB, the *Boston Naming Test* (BNT)¹³ and Subtest 6 (Inventory of Articulation Characteristics) of the *Apraxia Battery for Adults-Second Edition* (ABA-2)¹⁴ were administered to further address anomia and motor speech impairment, respectively. Subtest 6 of the ABA-2 is a rating scale where speech characteristics are evaluated on fifteen different items (e.g. the patient exhibits: marked difficulty initiating speech; highly inconsistent errors; or visible/audible searching). The range of scores on Subtest 6 is 0–15, where a score above 5 is thought to signify the presence of apraxia of speech (AOS); higher scores indicate more severe AOS. All participants included in this study were rated above the 5-point threshold, suggesting that AOS was present in all cases, although to varying degrees (Table 1).

Treatment

The treatment task was chosen with three factors in mind: 1) It was important that the task be structured so that the impact of auditory-only speech vs. auditory-visual speech on treatment outcome could be compared; 2) the treatment program needed to be as simple as possible so that participants could readily understand the task and run it by themselves with minimal or no assistance (treatment was self-administered on a lap-top computer at the participant's place of residence); and 3) given that our previous aphasia treatment research mainly has focused on object naming as the dependent factor,^{15,16} we chose to target naming in the current study so

as to draw upon previous experience with regard to post-testing and statistical treatment of the results.

With the above factors in mind, the current study examined the effect of audio-visual speech stimuli (Treatment AV) compared to audio-only speech stimuli (Treatment AO) on overt picture naming. To compare the outcome of the two treatment phases within participants, an AB/BA ‘within subject design’ was utilized. Each participant was provided with a lap-top computer installed with the treatment program, a pair of headphones, and a set of large green and red response buttons. The computer-based program consisted of two separate treatment phases (Treatment AO and Treatment AV), each including eighteen color pictures depicting high-frequency nouns¹⁷ presented in a three level hierarchy: Level 1 included six pictures, each randomly presented twelve times; Level 2 included six novel pictures, in addition to the six previously treated nouns, for a total of twelve nouns, each randomly presented eight times; and Level 3 included six novel nouns, in addition to the twelve previously treated nouns, for a total of eighteen nouns, each randomly presented five times during a single treatment session. The two lists of eighteen nouns targeted in each treatment phase were controlled for word frequency, phonological complexity, semantic content, and word length.

During Treatment AO, a picture was presented on the lap-top computer screen for two seconds. Once the picture disappeared, the screen displayed a fixation point (black crosshair on a white background), and an audio-recorded spoken noun was presented via headphones. Half of the picture/word pairs matched, while the other half of the trials were non-matched pairs, in which the target picture was randomly paired with an audio presentation of another treatment target. During Treatment AV, a picture was presented on the lap-top computer screen for two seconds and was followed by video of a male’s mouth saying a noun. Audio of the male producing the noun was presented in synchrony with the video via headphones. That is, participants both heard and “saw” the speech presentation. As with Treatment AO, half of all picture displays were followed by an audio-visual presentation of a matching noun, while the other half included a random audio-visual presentation of the remaining nouns. Treatment AO and Treatment AV were identical with the exception of AV including audio-visual presentation of nouns instead of audio-only. Half of the participants (P1–P5) began treatment with Treatment AO and then proceeded to Treatment AV, while the other half (P6–P10) received the opposite treatment order. Each treatment session lasted approximately 30 minutes.

Prior to the initiation of treatment, each participant and caregiver were provided with explicit directions regarding the task, which was to determine if the displayed picture matched the subsequent spoken word. If the pair represented a match, the participant would press the green button, and in the case of a non-match, the participant would press the red button. The next picture would not be displayed until a response was recorded. During both treatment phases, immediate feedback was provided following a response in the form of a “smiley face” for correct answers and a “frowny face” for incorrect answers. Additionally, following the completion of a treatment session, a data file of the participant’s responses was automatically saved, and the accuracy score from that session was displayed on the computer screen.

Participants completed one treatment session per day, five days per week. Each of the three levels was treated for a minimum of five sessions. If the participant reached at least 90% accuracy during at least one session, treatment proceeded to the next level of the hierarchy, in which additional nouns were added to the corpus of treatment items. Within each phase, treatment continued until 90% accuracy was achieved for all three levels, for a minimum of fifteen sessions per phase and thirty sessions in total. If a participant did not reach 90% accuracy during a given level within fifteen sessions, he or she was automatically moved up to the next level. This scenario only occurred once during treatment, in which P8 was unable to complete the first level of Treatment AV within fifteen sessions.

Outcome measures

To determine whether the participant's ability to name the trained items improved over the course of treatment, a naming task consisting of the 36 trained nouns (eighteen from each treatment phase) was administered. Color pictures were presented for three seconds on a computer screen, and participants were asked to name the items. Following each picture, a fixation point was presented on the screen for eight seconds to prepare the participant for the upcoming picture. Participants were allowed to respond during presentation of the fixation point.

To determine generalization from trained to untrained items, the *Philadelphia Naming Test* (PNT)¹⁸ was administered. The PNT consists of 175 low-, medium-, and high-frequency nouns. A picture representing each noun was displayed on a computer screen, and participants were asked to overtly name the picture as soon as the picture was displayed. Trials ended following a response or after twenty-seconds elapsed, in which the administrator said the correct picture name in order to discourage perseveration on subsequent trials.

The 36-item naming task and the PNT were administered twice each during two consecutive days prior to the start of treatment (baseline), between treatment phases, and after both treatment phases were completed, for a total of six administrations. Administrations of the 36-item naming task and the PNT were videotaped and later transcribed and scored by two speech-language pathologists for naming accuracy.

Statistics

Regression models for the number of correctly named items at baseline testing and following each of the two treatments were generated. The models' estimation presented two problems: 1) Rather than utilizing continuous variables, this study included discrete independent (type of treatment) and dependent (number of correctly named items) variables; and 2) since the outcome following each of the two treatments (Treatment AV vs. Treatment AO) constituted repeated measures, the dependent measures (number of correctly named items) were correlated within each participant. If this correlation is not taken into account, the standard errors of the regression coefficients' estimates will not be valid and results will be non-replicable. In order to solve these problems, the regression analyses utilized the Generalized Linear Models (GLM) approach with Generalized Estimation Equations (GEE) method for the parameter estimation.¹⁹ To implement these procedures, we utilized the SAS PROC GENMOD with binomial logit link function and repeated measures with an exchangeable correlation matrix.

Results

Treated Items

All ten participants completed both treatment phases and testing (Figure 1). On average, participants named 9.2 pictures (SD = 7.97) correctly across all six testing sessions for the 36 treated items. The average number of correctly named pictures at baseline was 7.5 (SD = 6.84). The average number of correctly named pictures following Treatment AV was 11.32 (SD = 9.76) and following Treatment AO it was 9.0 (SD = 7.00) (Figure 2).

The statistical analyses revealed that more items were named correctly following Treatment AV compared to baseline, $Z = 4.02$, $p < 0.0001$. Crucially, participants were able to name more items following Treatment AV compared to Treatment AO, which did not include the visual speech perception component, $Z = 3.43$, $p = 0.0006$. Treatment AO improved naming of treated items compared to baseline; however, this change in performance was not statistically significant, $Z = 1.05$, $p = 0.295$.

Treatment generalization

On average, 42.5 items (SD = 33.138) were named correctly across all six testing sessions on the 175-item PNT. At baseline, participants named an average of 36.05 pictures (SD = 31.12). Following Treatment AV, the average number of correctly named pictures was 49.32 (SD = 35.60); similarly, participants named an average of 42.26 items (SD = 32.981) following Treatment AO (Figure 2).

The statistical analyses revealed significant improvement in naming of untrained items following administration of Treatment AV compared to baseline performance, $Z = 4.42$, $p < 0.0001$. Treatment AV had a slightly better treatment effect than Treatment AO; however, this difference was not statistically significant, $Z = 1.54$, $p = 0.1244$. Compared to baseline, participants were able to name more untrained items following Treatment AO, although this difference was not statistically significant, $Z = 1.08$, $p = 0.0714$.

Discussion

The present findings suggest that incorporating a visual speech perception component in the treatment of non-fluent aphasia can improve speech production. Although this study did not address the neural mechanism which supports such recovery, it is possible that including the visual speech component modulated the residual frontal speech network and, thereby, promoted improved speech production. Such an assumption would rely on previous evidence from normal participants, suggesting that perception of audio-visual speech is associated with increased activity in the frontal speech-motor regions.^{6,9,20,21} Focusing on perception rather than production in patients whose speech output is already very limited allows for increased practice compared to more traditional treatments of non-fluent aphasia.^{22–24} That is, in addition to modulating the cortical speech network, a perceptual motor speech task can yield relatively high success compared to a speech production task and, as a result, is both motivating and therapeutic.

Although all but one of our participants was able to reach the 90% accuracy at each level of the treatment hierarchy, our data do not suggest that visual perception of others' motor speech movements is within normal limits in non-fluent aphasia. Schmid and Ziegler²⁵ found that compared to normal participants, patients with aphasia were impaired when visually discriminating both speech and non-speech oral movements. Furthermore, the study by Schmid and Ziegler revealed that the presence of AOS, an impairment of motor speech planning, was a predictor of task success regardless of the type of oral movements. This is not surprising given that critical brain damage associated with non-fluent speech includes the same frontal areas modulated by visual speech perception in normal participants. The present findings suggest that treatment of speech production in non-fluent aphasia can capitalize on motor speech perception, even though this process may be impaired.

It is pertinent to note that the positive outcome associated with using our simple computerized treatment task does not reduce the importance of clinician-based aphasia treatment administration. The treatment was designed with one goal in mind: To demonstrate that visual perception of others' speech-motor movements can be taxed as a medium to improve speech production in non-fluent aphasia. It is quite possible that more rigorous and patient-specific perceptual speech treatment could have elicited even greater success. By the same token, it is possible that manipulation of the treatment task could have improved success for those who struggled with the task. Based on our clinical experience, we speculate that better treatment outcome could be obtained by tailoring the treatment program to better fit individual patients. This could be accomplished by manipulating factors such as: 1) overall word frequency (e.g. higher word frequency for more severe aphasia; lower word frequency for milder aphasia); 2) the number of words included in the treatment phase (e.g. fewer words for more severe aphasia

to increase initial task success); 3) time interval between the picture and onset of the spoken word (e.g. lengthening the time for patients with slower reaction time); and 4) session length (longer sessions for patients with better endurance). It is also worth mentioning that participants' feedback regarding the computer-based treatment was overwhelmingly positive. We think that this positive response most likely can be attributed to the participants' feeling of greater control over their own treatment, as opposed to when all aspects of aphasia therapy (e.g. treatment pace, treatment duration, time of administration) are dictated by a clinician.

The current study did not include assessment of brain activity before and after treatment. It is possible that neuroimaging can reveal what brain areas support improved speech production following visual speech perception treatment. Thereby, it would be possible to determine whether the left perilesional frontal lobe areas or the right homologues of the classical language areas are recruited to support this kind of recovery. Our lab is currently working on this issue using functional magnetic resonance imaging (fMRI). It is also important to understand whether greater manipulation of task difficulty could improve treatment outcome. The current treatment proceeded from low- to high-task difficulty by increasing the number of treated words in a stepwise manner. This is somewhat similar to errorless treatment methods that proceed from lower to higher levels of difficulty to minimize errors and, consequently, capitalize on minimal reinforcement of incorrect responses.²⁶ An alternative approach would be to administer treatment utilizing targets of greater complexity (e.g. using mostly multisyllabic and/or low frequency words), as a number of studies have found that treatment of more complex structures can facilitate greater generalization to less complex stimuli.^{27,28}

All of the participants in the current study had Broca's aphasia. Yet, the overall theme of this paper focused on non-fluent aphasia. Based on the WAB, non-fluent aphasia would also include trans-cortical motor aphasia and global aphasia. Thus, it is pertinent to consider whether patients with other kinds of non-fluent aphasia than Broca's aphasia would also benefit from the kind of treatment presented here. In this regard, it is important to note that although all of the participants in the current study were classified as having Broca's aphasia based on their performance on the WAB, their language profiles varied considerably with regard to, for example, speech content and repetition scores. Treatment success also varied widely, whereas P1 and P10 responded very poorly while P2–P6 and P9 experienced a much more favorable outcome. In addition, more severe aphasia also tended to be associated with less or no improvement in naming. Based on this observation, it seems possible that patients with global aphasia, where auditory comprehension is more severely affected than in Broca's aphasia, would be less likely to benefit from the current treatment compared to those with less severe aphasia (e.g. Broca's or trans-cortical motor aphasia). Conversely, treatment success did not appear to be related to the severity of AOS, suggesting that impaired speech production is not necessarily a predictor of treatment success. Clearly, further research is needed to examine the relation between communication impairment and success using audio-visual treatment to target speech perception in aphasia.

In closing, our data suggest that behavioral treatment that targets visual speech perception to improve speech production can improve overt picture naming in patients with non-fluent aphasia. This hypothesis was initially derived from evidence from normal participants, which suggests that the speech areas in the left frontal lobe are modulated not only during speech production but also for visual perception of others' speech-motor movements. As is almost always the case with aphasia treatment, there was a wide range of treatment outcomes among the current participants. Nevertheless, our data suggest that treating speech perception may be a viable approach to improve speech production in patients with non-fluent aphasia, a population where speech output seldom improves significantly in the chronic phase of the disorder.

Acknowledgements

This work was supported by grants to Julius Fridriksson from the NIDCD (DC005915 & DC008355) and a grant to Chris Rorden from the NINDS (NS054266). We would like to thank Adair Kopani and Jerry Wilcoxon for their assistance in data collection.

References

1. Marshall RC, Phillips DS. Prognosis for improved verbal communication in aphasic stroke patients. *Arch Phys Med Rehabil* 1983;64:597–600. [PubMed: 6197949]
2. Fridriksson J, Morrow L, Moser D, Fridriksson A, Baylis G. Neural correlates of anomia recovery in aphasia. *Neuroimage* 2006;32:1403–1412. [PubMed: 16766207]
3. Pickersgill MJ, Lincoln NB. Prognostic indicators and the pattern of recovery of communication in aphasic stroke patients. *J Neurol Neurosurg Psychiatry* 1983;46:130–139. [PubMed: 6842216]
4. Dronkers NF. A new brain region for speech: The insula and articulatory planning. *Nature* 1996;384:159–161. [PubMed: 8906789]
5. Hillis AE, Work M, Barker PB, Jacobs MA, Breese EL, Maurer K. Re-examining the brain regions crucial for orchestrating speech articulation. *Brain* 2004;127:1479–1487. [PubMed: 15090478]
6. Calvert GA, Campbell R. Reading speech from still and moving faces: The neural substrates of visible speech. *J Cogn Neurosci* 2003;15:57–70. [PubMed: 12590843]
7. Fridriksson J, Moss J, Davis B, Baylis G, Rorden C. Motor speech perception modulates the cortical language areas. *Neuroimage*. in press
8. Hall DA, Fussell C, Summerfield AQ. Reading fluent speech from talking faces: typical brain networks and individual differences. *J Cogn Neurosci* 2005;17:939–953. [PubMed: 15969911]
9. Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. *Neuroimage* 2005;25:76–89. [PubMed: 15734345]
10. Goodglass, H.; Kaplan, E. The assessment of aphasia and related disorders and the Boston Diagnostic Aphasia Examination. Philadelphia: Lea and Febiger; 1983.
11. Kertesz, A. Western Aphasia Battery. New York, NY: Harcourt Brace Jovanovich; 1982.
12. Pedersen PM, Vinterb K, Olsen TS. Aphasia after stroke: type, severity, and prognosis. The Copenhagen Aphasia Study. *Cerebrovasc Dis* 2004;17:35–43.
13. Kaplan, E.; Goodglass, H.; Weintraub, S. Boston Naming Test. Vol. 2. Boston: Lippincott Williams & Wilkins; 2001.
14. Dabul, BL. Apraxia Battery for Adults. Vol. 2. Tigard, OR: CC Publications, Inc; 2000.
15. Fridriksson J, Morrow-Odom L, Moser D, Fridriksson A, Baylis G. Neural recruitment associated with anomia treatment in aphasia. *Neuroimage* 2006;32:1403–1412. [PubMed: 16766207]
16. Fridriksson J, Moser D, Bonilha L, Morrow-Odom KL, Shaw H, Fridriksson A, Baylis GC, Rorden C. Neural correlated of phonological- and semantic-based anomia treatment in aphasia. *Neuropsychologia* 2007;45:1812–1822. [PubMed: 17292928]
17. Francis, WN.; Kucera, H. Frequency Analysis of English Usage: Lexicon and Grammar. Boston, MD: Houghton Mifflin Company; 1982.
18. Roach A, Schwartz MF, Martin N, Grewal RS, Brecher A. The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology* 1996;24:121–133.
19. Agresti, A. Categorical Data Analysis. Vol. 2. Hoboken, NJ: John Wiley & Sons, Inc; 2002.
20. Campbell R, MacSweeney M, Surguladze S, Calvert GA, McGuire PK, Brammer MJ, David AS, Suckling J. Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain Res* 2001;12:233–243.
21. Ojanen V, Möttönen R, Pekkola J, Jääskeläinen IP, Joensuu R, Autti T, Sams M. Processing of audiovisual speech in Broca's area. *Neuroimage* 2005;25:333–338. [PubMed: 15784412]
22. Chapey, R. Language Intervention Strategies in Adult Aphasia. Vol. 3. Baltimore, MD: Williams & Wilkins; 1994.

23. Weinrich M, Shelton JR, McCall D, Cox DM. Generalization from single sentence to multisentence production in severely aphasic patients. *Brain Lang* 1997;58:327–352. [PubMed: 9182751]
24. Raymer A, Kohen F. Word-retrieval treatment in aphasia: Effects of sentence context. *J Rehabil Res Dev* 2006;43:367–378. [PubMed: 17041822]
25. Schmid G, Ziegler W. Audio-visual matching of speech and non-speech oral gestures in patients with aphasia and apraxia of speech. *Neuropsychologia* 2006;44:546–55. [PubMed: 16129459]
26. Fillingham JK, Sage K, Lambon Ralph MA. The treatment of anomia using errorless learning. *Neuropsychol Rehab* 2006;16:129–154.
27. Thompson CK, Shapiro LP. Complexity in treatment of syntactic deficits. *Am J Speech-Lang Path* 2007;16:30–42. [PubMed: 17329673]
28. Kiran S, Thompson CK. The role of semantic complexity in treatment of naming deficits: training semantic categories in fluent aphasia by controlling exemplar typicality. *J Speech Lang Hear Res* 2003;46:773–787. [PubMed: 12959459]

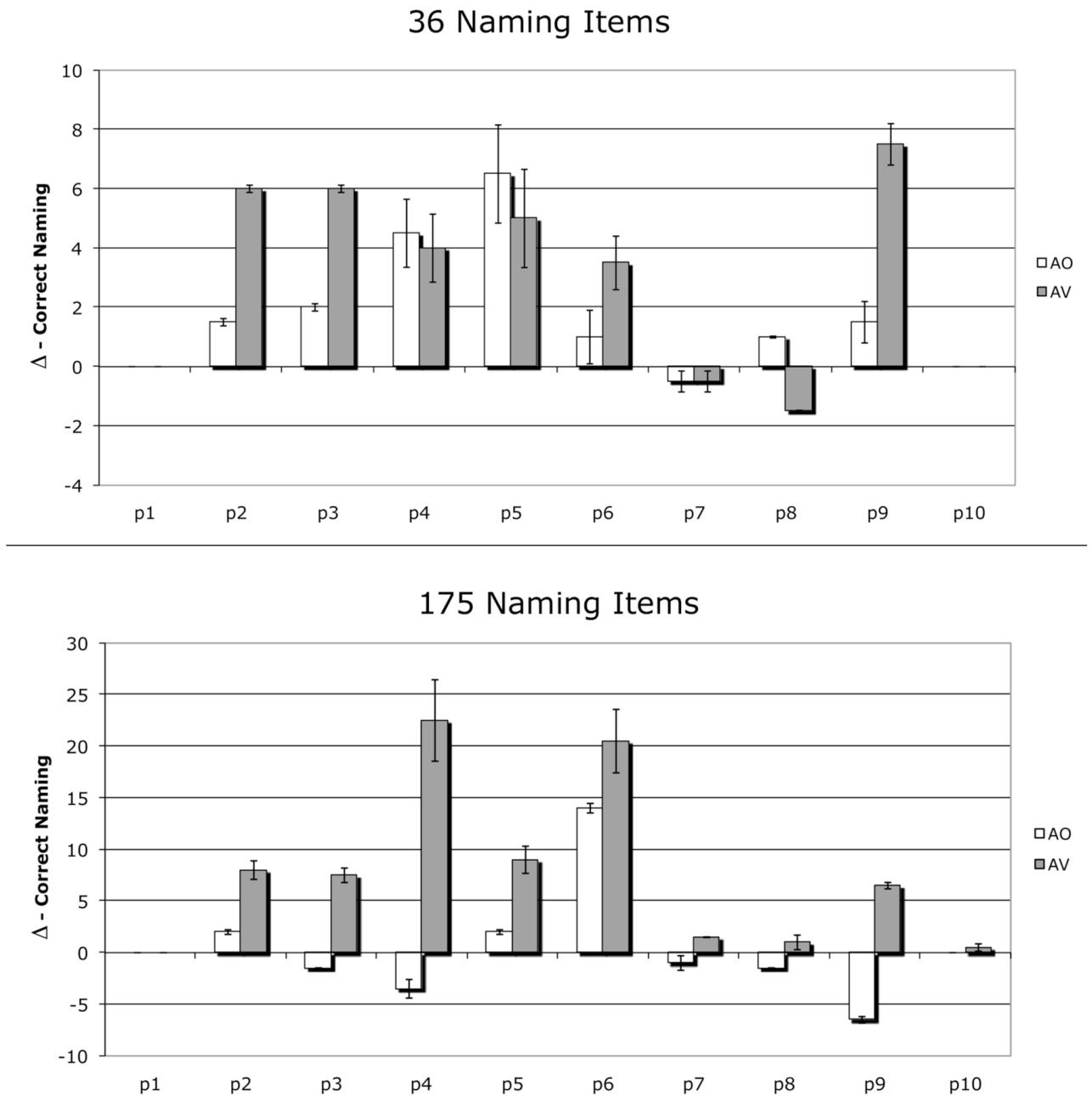


Figure 1. Change (Δ) in correct naming by each of the participants following the audio-only (AO) and audio-visual (AV) treatment phases. The top graph shows results for 36-item naming task (treated items); the bottom graphs shows data for the 175-item PNT (treatment generalization).

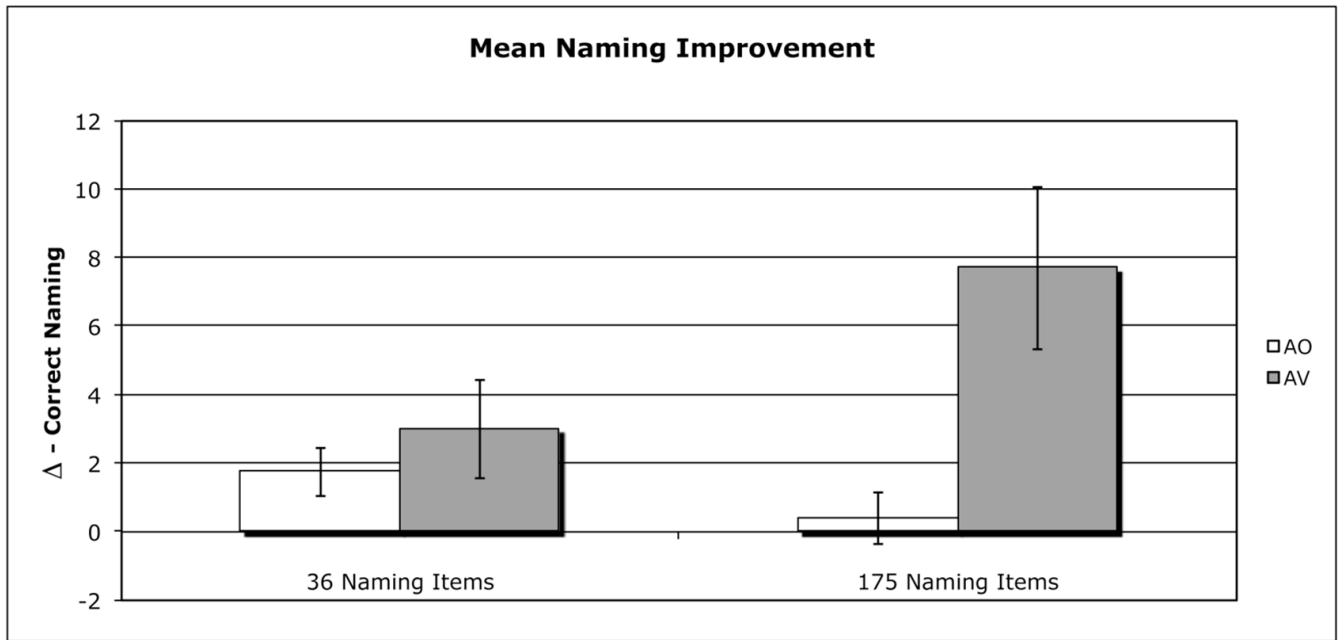


Figure 2. Mean change (Δ) in correct naming following the audio-only (AO) and audiovisual (AV) treatment phases across all participants.

Table 1

Participant's biographical information and diagnostic testing information

P	Biographical Information			WAB						BNT	ABA-2
	Sex	Age	Post-onset*	Content [†]	Fluency [†]	Auditory Comp [†]	Repetition [†]	Naming [†]	AQ‡		
1	M	71	80	2	1	5.60	0	0	17.2	0	10
2	M	63	88	8	4	7.95	5.6	6.4	63.9	24	7
3	F	43	49	8	4	8.05	2.5	2.8	50.7	13	11
4	M	59	204	8	4	6.20	2.7	6.2	54.2	12	9
5	M	74	32	8	4	7.05	7.5	5.2	63.5	20	10
6	M	58	43	9	4	9.70	6.4	6.7	71.6	8	8
7	M	70	216	3	1	4.35	3.0	1.7	26.1	3	10
8	M	63	101	3	4	5.85	6.4	3.3	45.1	3	6
9	M	60	23	5	3	5.65	3.2	5.3	44.3	7	9
10	F	33	17	2	2	6.90	2.4	2.6	31.8	0	11

*The time post-onset of stroke is measured in months

†Maximum score of 10

‡Maximum score of 100

§Maximum score of 60

||Range of scores = 0-15