Treatting Visual Speech Perception to Improve Speech Production in Nonfluent Aphasia

Julius Fridriksson, PhD; Julie M. Baker, MA; Janet Whiteside, PhD; David Eoute, Jr, MSP; Dana Moser, PhD; Roumen Vesselinov, PhD; Chris Rorden, PhD

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 nonfluent aphasia. Based on these findings, this study examined the utility of targeting visual speech perception to improve speech production in nonfluent aphasia.

Methods—Ten patients with chronic nonfluent aphasia underwent computerized language treatment utilizing picture–word matching. To examine the effect of visual speech perception on picture naming, 2 treatment phases were compared—one that included matching pictures to heard words and another in which 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 after treatment when it included a visual speech component (ie, seeing the speaker’s mouth). In contrast, the treatment phase in which 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 nonfluent aphasia and may provide an alternative approach to treat a disorder in which speech production seldom improves much in the chronic phase of stroke. (Stroke. 2009;40:853-858.)

Key Words: aphasia ■ speech disorders ■ stroke recovery ■ therapy ■ treatment

Patients with chronic nonfluent aphasia experience minimal recovery of speech production,1 even in the case of language treatment.2,3 Behavioral treatment of nonfluent 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 because of 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 et al9 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 after stroke, particularly in relation to the treatment of speech production in individuals with nonfluent aphasia. That is, it may be possible that training audio-visual speech perception, a task that modulates the frontal motor speech areas in normal participants, will improve speech production in persons with nonfluent aphasia by stimulating the residual speech-motor network. Moreover, by focusing on speech perception, a skill that is relatively spared compared to speech production in patients with nonfluent 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 audio-visual speech stimuli compared to audio-only speech stimuli to improve overt picture naming in 10 persons with nonfluent aphasia. We hypothesized that persons with nonfluent 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.

Subjects and Methods

Participants

The inclusion criteria for this study were as follows: (1) nonfluent aphasia classification; (2) single left hemisphere stroke; (3) at least 1

Received July 23, 2008; accepted August 29, 2008.

From the Department of Communication Sciences and Disorders (J.F., J.M.B., D.E., D.M., C.R.), Department of Statistics (R.V.), University of South Carolina, Columbia, SC; Department of Communication Sciences & Disorders (J.W.), University of Central Florida, Orlando, Fla. 
Correspondence to Julius Fridriksson, PhD, Associate Professor, Dept. of Communication Sciences & Disorders, University of South Carolina, Columbia, SC 29208. E-mail jfridrik@sc.edu

© 2009 American Heart Association, Inc.

Stroke is available at http://stroke.ahajournals.org

DOI: 10.1161/STROKEAHA.108.532499
year after stroke; (4) premorbidly right-hand-dominant; and (5) native speaker of English. Ten persons (8 males) who met these criteria participated in this study. The mean age was 59.2 years (SD, 12.1); with a range of 41 years (Table). Aphasia assessment using the Western Aphasia Battery\(^1\) revealed that all 10 participants had a language impairment most consistent with Broca’s aphasia—the most common type of chronic nonfluent aphasia.\(^3\) In addition to the Western Aphasia Battery, the Boston Naming Test\(^4\) and Subtest 6 (Inventory of Articulation Characteristics) of the Apraxia Battery for Adults–Second Edition\(^5\) were administered to further address anomia and motor speech impairment, respectively. Subtest 6 of the Apraxia Battery for Adults–Second Edition is a rating scale in which speech characteristics are evaluated on 15 different items (eg, the patient exhibits marked difficulty initiating speech, highly inconsistent errors, or visible/audible searching). The range of scores on Subtest 6 is 0 to 15, where a score >5 is thought to signify the presence of apraxia of speech; higher scores indicate more severe apraxia of speech. All participants included in this study were rated as above the 5-point threshold, suggesting that apraxia of speech was present in all cases, although to varying degrees (Table).

### Treatment

The treatment task was chosen with 3 factors in mind: (1) it was important that the task be structured so that the impact of audio-only speech vs audio-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 perform it by themselves with minimal or no assistance (treatment was self-administered on a laptop computer at the participant’s place of residence); and (3) given that our previous aphasia treatment research has mainly focused on object naming as the dependent factor,\(^2\)\(^3\) we chose to target naming in the current study to draw on previous experience with regard to post-testing and statistical treatment of the results.

With these 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 2 treatment phases within participants, an AB/BA “within subject design” was utilized. Each participant was provided with a laptop 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 2 separate treatment phases (treatment AO and treatment AV), each including 18 color pictures depicting high-frequency nouns\(^16\) presented in a 3-level hierarchy. Level 1 included 6 pictures, each randomly presented 12 times; level 2 included 6 novel pictures, in addition to the 6 previously treated nouns, for a total of 12 nouns, each randomly presented 8 times; and level 3 included 6 novel nouns, in addition to the 12 previously treated nouns, for a total of 18 nouns, each randomly presented 5 times during a single treatment session. The 2 lists of 18 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 laptop computer screen for 2 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, whereas the other half of the trials were nonmatched pairs, in which the target noun was presented in synchrony with the video via headphones. As with treatment AO, half of all picture displays were followed by an audio-visual presentation of the remaining noun on the screen, with the video of a male’s mouth saying a noun. Audio of the male producing the noun was presented 8 times; and level 3 included 6 novel pictures, in addition to the 6 previously treated nouns, for a total of 12 nouns, each randomly presented 5 times during a single treatment session. The 2 lists of 18 nouns targeted in each treatment phase were controlled for word frequency, phonological complexity, semantic content, and word length.

During treatment AV, a picture was presented on the laptop computer screen for 2 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, whereas the other half of the trials were nonmatched 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 laptop computer screen for 2 seconds and was followed by a 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, whereas 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 (participants 1–5) began treatment with treatment AO and then proceeded to treatment AV, whereas the other half (P6–P10) received the opposite treatment order. Each treatment session lasted \(~30\) minutes.

Before the initiation of treatment, each participant and caregiver was 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 nonmatch, 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 after a response in the form of a “smiley face” for correct answers and a “frown face” for incorrect answers. Additionally, after the comple-
tion 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 1 treatment session per day, 5 days per week. Each of the 3 levels was treated for a minimum of 5 sessions. If the participant reached at least 90% accuracy during at least 1 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 3 levels, for a minimum of 15 sessions per phase and 30 sessions in total. If a participant did not reach 90% accuracy during a given level within 15 sessions, he or she was automatically moved up to the next level. This scenario only occurred once during treatment, in which participant 8 was unable to complete the first level of treatment AV within 15 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 (18 from each treatment phase) was administered. Color pictures were presented for 2 seconds on a computer screen, and participants were asked to name the items. After each picture, a fixation point was presented on the screen for 8 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 was administered. The Philadelphia Naming Test 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 after a response or after 20 seconds elapsed, in which the administrator said the correct picture name to discourage perseveration on subsequent trials.

The 36-item naming task and the Philadelphia Naming Test were administered twice each during 2 consecutive days before the start of treatment (baseline), between treatment phases, and after both treatment phases were completed, for a total of 6 administrations. Administrations of the 36-item naming task and the Philadelphia Naming Test were videotaped and later transcribed and scored by 2 speech-language pathologists for naming accuracy.

Statistics
Regression models for the number of correctly named items at baseline testing and after each of the 2 treatments were generated. The models’ estimation presented 2 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) because the outcome after each of the 2 treatments (treatment AV vs treatment AO) constituted repeated measures, the dependent measures 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 nonreplicable. To solve these problems, the regression analyses utilized the Generalized Linear Models approach with Generalized Estimation Equations method for the parameter estimation. 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 10 participants completed both treatment phases and testing (Figure 1). On average, participants named 9.2 pictures (SD, 7.97) correctly across all 6 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 after treatment AV was 11.32 (SD, 9.76), and after treatment AO it was 9.0 (SD, 7.00; Figure 2).

The statistical analyses revealed that more items were named correctly after treatment AV compared to baseline \(Z=4.02; P<0.0001\). Crucially, participants were able to name more items after 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, 33.14) were named correctly across all 6 testing sessions on the 175-item Philadelphia Naming Test. At baseline, participants named an average of 36.05 pictures (SD, 31.12). After 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, 31.12) after treatment AO (Figure 2).

The statistical analyses revealed significant improvement in naming of untrained items after 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 after 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 nonfluent aphasia can improve speech production. Although this study did not address the neural mechanism that 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 \(6,9,19,20\). 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 nonfluent aphasia \(21\text{--}23\). 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 1 of our participants were able to reach 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 nonfluent aphasia. Schmid and Ziegler \(24\) found that compared to normal participants, patients with aphasia were impaired when visually discriminating both speech and nonspeech oral movements. Furthermore, the study by Schmid and Ziegler \(24\) revealed that the presence of apraxia of speech, 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 nonfluent speech includes the same frontal areas modulated by visual speech perception in normal participants. The present findings suggest that treatment of speech production in nonfluent 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 1 goal in mind: to demonstrate that visual perception of others’ speech-motor movements can be taxed as a medium to improve speech production in nonfluent 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 (eg, higher-frequency words for more severe aphasia; lower-frequency words for milder aphasia); (2) the number of words included in the treatment phase (eg, fewer words for more severe aphasia to increase initial task success); (3) time interval between the picture and onset of the spoken word (eg, lengthening the time for patients with slower reaction time); and (4) session length (longer sessions for patients with better endurance). It is also worth mention-
ing 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 (eg, 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 after 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 laboratory is currently working on this issue using functional MRI. 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 step-wise 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.25 An alternative approach would be to administer treatment utilizing targets of greater complexity (eg, using mostly multisyllabic or low-frequency words), because a number of studies have found that treatment of more complex structures can facilitate greater generalization to less complex stimuli.

All of the participants in the current study had Broca’s aphasia. Yet the overall theme of this article focuses on nonfluent aphasia. Based on the Western Aphasia Battery, nonfluent aphasia would also include transcortical motor aphasia and global aphasia. Thus, it is pertinent to consider whether patients with other kinds of nonfluent 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 Western Aphasia Battery, their language profiles varied considerably with regard to, for example, speech content and repetition scores. Treatment success also varied widely, in which participants 2-6 and participant 9 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, in which 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 (eg, Broca’s or transcortical motor aphasia). Conversely, treatment success did not appear to be related to the severity of apraxia of speech, 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 nonfluent 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 nonfluent aphasia, a population in which speech output seldom improves significantly in the chronic phase of the disorder.

Acknowledgments

The authors thank Adair Kopani and Jerry Wilcoxson for their assistance in data collection.

Sources of Funding

This work was supported by grants to Julius Fridriksson from the NIDCD (DC005915 and DC008355), and a grant to Chris Rorden from the NINDS (NS054266).

Disclosures

None.

References


Treating Visual Speech Perception to Improve Speech Production in Nonfluent Aphasia
Julius Fridriksson, Julie M. Baker, Janet Whiteside, David Eoute, Jr, Dana Moser, Roumen Vesselinov and Chris Rorden

*Stroke*. 2009;40:853-858; originally published online January 22, 2009;
doi: 10.1161/STROKEAHA.108.532499

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/40/3/853

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in
*Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org//subscriptions/