ONLINE SUPPLEMENT

Performance Variability as a Predictor of Response to Aphasia Treatment

E. Susan Duncan, MA, MS[1], Tanya Schmah, PhD[2,3], Steven L. Small, PhD, MD[3,4]

1. Departments of Cognitive Sciences and Neurology, University of California, Irvine, CA, USA

2. Rotman Research Institute, Baycrest, Toronto, ON, Canada

3. Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada

4. Departments of Neurology, Neurobiology and Behavior, and Cognitive Sciences, University of California, Irvine, CA, USA

Corresponding Author:

E. Susan Duncan

Room 3130 Biological Sciences III, Department of Neurology

UC, Irvine

Irvine, CA 92697

Email: duncane@uci.edu

Phone: 949-824-3783

Fax: 949-824-3360

Running Title: Performance Variability and Aphasia

SUPPLEMENTAL METHODS

Therapy Description

The important therapeutic features in IMITATE include visual observation, oral repetition, speaker variability, ecologically valid stimuli, high intensity, graded incremental learning and variability in gradation[1,2] . Visual observation refers to the use of audiovisual stimuli in which the speaker's moving face, lips and mouth are visible, in contrast to the control REPEAT therapy, in which the patients viewed a static image of each speaker while hearing their voice utter the target word or phrase. Oral repetition was performed during both IMITATE and REPEAT therapy sessions. Speaker variability was implemented by the presentation of each target word or phrase by each of six talkers. This was used for both therapies and was theorized to aid in generalization to a wider variety of talkers. Ecological stimuli refers to the use of real words and phrases that might be used by an English speaker in the course of daily activities (rather than non-speech oral movements, isolated syllables or nonsense words), as well as to the display of a visible talker such as one might engage with in normal communicative interactions. The use of real words and phrases was common to both groups, but visual observation was used only for patients receiving IMITATE. Use of ecologically valid stimuli was physiologically motivated by the shared substrates found for observation and execution of action, including oral communicative actions, on a cellular level in studies of macaque cortex[3,4] and in a motor cortical network model supported by fMRI findings in neurologically intact human subjects[5]. Such stimuli may enhance the efficacy of neural connectivity by matching observed actions, such as speech, to internal motor representations existing within the viewer's repertoire of action performance[6].

High intensity was implemented by requiring patients from both groups to participate in 90 minutes (3 30-minute sessions) daily, 6 days a week, for 6 weeks. This level of intensity, far

greater than the level that can typically be provided by a trained therapist, was chosen based on the positive correlation between intensity and therapeutic outcome[7]. This massed practice used in both the experimental and control therapies is consistent with other biologically motivated behavioral therapies for aphasia rehabilitation, notably Constraint Induced Aphasia Therapy (CIAT[8]). The IMITATE subjects were also exposed to the entire stimulus block of a target before producing a block repetitions, while the REPEAT group had a period for repetition following each of the 6 presentations. The massed stimulation of the IMITATE therapy has been theorized to prime the neurophysiological system and facilitate the generation or re-instatement of neural pathways[9].

Graded incremental learning was addressed by advancing patients in both groups through levels featuring successively more difficult stimuli, such as longer words, more complex phonology, more varied word classes, and longer word sequences (phrases and sentences). Each patient started the therapy at a level that was judged to be appropriate to pre-treatment repetition capability, and advanced through a level each week, unless the clinician performing the weekly repetition test (described in Methods) judged that repetition of a level was necessary. Finally, variability in gradation was implemented in both therapies by occasionally providing simpler stimuli at higher levels as well as more difficult stimuli at lower levels according to a probabilistic algorithm. In addition to being consistent with the variability of daily communicative demands, this approach sought to combine two conflicting bodies of evidence on learning strategies that suggest that either a simple[10] or a complex[11] origin (i.e., initial level of difficulty) yields maximum benefit.

Selection of Lexical and Phrasal Stimuli

Therapy items for IMITATE were selected by an algorithm that was designed to consider parameters including number of letters, phonemes, and syllables, part of speech, written frequency, familiarity, frontal and total visibility, and phonemic complexity. These values were derived from the MRC Psycholinguistic Database[12], the Kucera and Francis corpus[13], the Hoosier mental lexicon[14], and various measures of viseme content[15,16]. Measures of phonemic complexity were calculated by coding stimuli for presence of consonant blends in the initial position. Visibility of consonants and vowels was assessed on a 4-point scale that assigned high values to high-visibility productions, like consonants /p/, /b/, /m/, /f/, and /v/. Combining all criteria yielded a final pool of 2568 words, which was augmented by the addition of 68 words that were added due to high functional utility (e.g., “blue”, “March”, “chair”, “Monday”). Words contained between 1 and 4 syllables (mean = 1.42) and between 1 and 12 phonemes (mean = 4.09).

The stimulus set also included 405 phrases that were chosen due to common use and high functional utility for people with aphasia (e.g., “sit down”, “watch out”, “nice to see you”, “please pass the salt”). These were selected from a large variety of English language textbooks, travel guides, and intuition. Phrases were assigned a value on the basis of the number of words and syllables, as well as verb and preposition frequency. Phrases contained two to nine syllables (mean = 4.03) and two to five words (mean = 3.36).

Twelve treatment levels were developed, which increased gradually in complexity. Individual patients were assigned to a treatment level based on their level of functioning. The stimuli analyzed for this study were selected on the basis of the level to which the individual was

expected to advance over the course of therapy. This higher level was selected to diminish ceiling effects.

Demographic and Lesion Information for All Subjects

Select demographic and lesion information is included in Table I.

Intra-individual Variability Measure

For each time point (Week 0 or Week 6) and each subject, we computed the variance of the per-word scores in the pooled Word blocks, and the same for the pooled Phrase blocks. Our intra-individual variability measure for repetition score is the square root of the mean of the variances in the two blocks (Word and Phrase), which is comparable to a pooled standard deviation but gives equal weight to the Words and Phrases. This is analogous to our definition of the repetition Mean (see above), which gives equal weight to Words and Phrases, despite the Phrases block containing a larger, and more variable, number of words than the Words block.

Exclusion of Subjects from Intra-individual Variability Analysis

Beginning with all 18 subjects for which intra-individual variability data was available, i.e. all except Subject 2, we calculated full and partial Pearson correlation coefficients for the relationship between pre-treatment intra-individual variability and Improvement. The full correlation coefficient was $r = 0.79$. The partial correlation coefficient, controlling for age, months post onset and number of sessions completed as potential confounding variables, was $r = 0.78$. The partial correlation coefficient, with these confounding variables plus an additional one of pre-treatment repetition mean performance, was $r = 0.76$.

We then excluded Subject 4 as an outlier, since its pre-treatment mean performance lies more than 3 standard deviations below the group mean, see Figure 1B. Excluding this subject only, we recalculated the correlation coefficients between intra-individual variability and Improvement in post-therapy repetition mean, with results shown in Table II. Note that the full correlation coefficient of r = 0.75 (excluding Subject 4) is very similar to the value r=0.79 obtained using all subjects. The same is true of the partial correlation coefficient, with Age, Months Post Onset, and Number of Sessions Completed as confounding variables: r = 0.76 (excluding Subject 4) versus r = 0.78 using all subjects. However, with the additional confounding variable of Pretreatment Mean, the partial correlation coefficient is now r = 0.47 (excluding Subject 4) versus r = 0.76 using all subjects, and is not significant at the p=0.05 level. This means that, when Subject 4 is excluded, Pretreatment Mean and the other confounding variables listed predict Improvement well enough that including Pretreatment intra-individual variability as an additional predictor does not significantly improve the prediction.

Noting the effect of controlling for pre-treatment mean in the context of the overall high performance scores (group mean 79.4%, SD 18.8%), we considered that ceiling effects might have artificially constrained both intra-individual variability and Improvement scores. The asymptotic appearance of Figure 1B, showing pretreatment repetition Mean vs. intra-individual variability, further reinforces this possibility. This scatter plot shows several near-ceiling means, and also a strong negative association (excluding the outlier, Subject 4) between Mean and intra-individual variability. Therefore, our intra-individual variability measure may be artificially low for subjects with high Mean performance, due purely to a ceiling effect, especially since our intra-individual variability measure is based on variance, the estimation of which gives high weight to extreme values. In order to compensate partially for this, we recalculated the above

correlation coefficients using a restricted set of subjects. We calculated the overall mean of the per-subject standard errors of the repetition Mean, which was 2.2% pre-treatment and 1.9% post treatment, and excluded those subjects with a repetition Mean, either pre- or post- treatment, within one mean standard error of the ceiling (100%). (Note that the standard error of the repetition Mean score (SEM) equals one half of the square root of the sum of squared SEMs for Words and Phrases separately.) The subjects excluded are listed in the main text, and marked with crosses in Figure 1. We also used analogous criteria to exclude subjects with near-ceiling Words mean scores from all calculations using the Words scores only; and similarly for the Phrases-only calculations. These criteria led to exclusion of different subsets of subjects for the Words-only and Phrases-only calculations, as listed in the main text. Table 3 contains results analogous to Table I, but with the listed subjects excluded.

SUPPLEMENTAL TABLES

Subject	Sex	Age	Months Post-Onset	Treatment Group	# of Sessions Completed	Lesion Size (% LH)	Lesion Location	Aphasia Classification
1	F	72	17	I	88	9.64	FIPT	Broca's
2	M	60	5	R	84	9.72	FI	Broca's
4	M	63	7	I	54	7.52	FIPT	Broca's
5	M	56	16	R	90	3.31	FPT	Broca's
6	M	65	8	I	90	6.36	TP	Conduction
9	F	46	28	I	108	17.86	FPT	Broca's
10	F	31	11	I	106	10.62	FIPT	Anomic
11	M	58	13	R	101	19.78	FIPT	Trans. Motor
12	F	55	22	R	108	0.95	BG	Anomic
13	M	36	78	R	108	12.35	FIPT	Broca's
14	M	37	51	R	105	10.06	TIPO	Broca's
15	M	70	120	I	99	26.34	FIPT	Anomic
16	M	58	29	I	108	3.25	FI	Anomic
17	M	57	130	I	107	13.52	FIPT	Anomic
18	M	55	81	R	79	11.54	FIPTO	Wernicke's
19	M	42	124	R	53	5.21	FI	Anomic
20	M	60	7	I	103	11.42	TPO	Trans. Sensory
21	M	43	15	I	108	12.44	FIT	Broca's
22	M	49	29	I	108	11.24	FIPT	Broca's

Table I: Individual data for each of the 19 subjects. The fourth column gives months post stroke onset at time of enrollment. The fifth column gives treatment group (IMITATE or REPEAT). Lesion size is given as the percentage of left hemisphere (LH) voxels included in the lesion mask. Lesion location is abbreviated as follows: F- Frontal, P- Parietal, T-Temporal, I-Insular, O-Occipital, BG- Basal Ganglia. Aphasia classifications are as determined by the Western Aphasia Battery-Revised (WAB-R) and Transcortical is abbreviated (Trans).

Repetition measures:	r	p
Intra-individual variability vs. Improvement	0.75	0.0005
(controlling for Age, MPO, aphasia type, NSC)	0.73	0.0050
(controlling for Age, MPO, NSC, aphasia type, and Mean)	0.48	0.1153
Mean vs. Improvement	-0.64	0.0055

Table II: Full and partial correlation coefficients for Pretreatment intra-individual variability and Improvement in Mean accuracy, for the repetition test. Partial correlations control for Age, Months Post Onset (MPO), aphasia type (fluent vs. nonfluent), and Number of Sessions Completed (NSC) and pre-treatment Mean. The full correlation between pre-treatment Mean and Improvement is also included. Subject 4 has been excluded as an outlier from all calculations.

SUPPLEMENTAL FIGURE

Figure S1. Visual depiction of experimental design. Two pre-therapy assessments (Weeks -6,0) were separated by a 6-week interval during which no therapy was provided. Between Weeks 0 and 6, subjects participated in six weeks of therapy. No therapy was provided in the interval between the two post-therapy sessions (Weeks 6,12).

SUPPLEMENTAL REFERENCES

1. Small SL. A Biological basis for Aphasia Treatment: Mirror Neurons and Observation-Execution Matching. *Poznań Stud Contemp* 2009, 45: 313-326.

2. Lee J, Fowler R, Rodney D, Cherney L, Small SL. IMITATE: An intensive computer-based treatment for aphasia based on action observation and imitation. *Aphasiology* 2010; 24: 449-465.

3. Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. *Cognitive Brain Res* 1996; 3: 131-141.

4. Ferrari PF, Gallese V, Rizzolatti G, Fogassi L. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur J Neurosci* 2003; 17: 1703-1714.

5. Mashal N, Solodkin A, Dick AS, Chen EE, Small SL. A network model of observation and imitation of speech. *Front Psychol* 2012; 3: 84.

6. Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G. Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 2002; 297: 846-848.

7. Bhogal SK, Teasell R, Speechley M. Intensity of aphasia therapy, impact on recovery. *Stroke* 2003; 34: 987-993.

8. Pulvermüller F, Neininger B, Elbert T, et al. Constraint-induced therapy of chronic aphasia after stroke. *Stroke* 2001; 32: 1621-1626.

9. Ferguson A. Clinical Forum Learning in aphasia therapy: It's not so much what you do, but how you do it!. *Aphasiology* 1999; 13: 125-150.

10. Elman JL. Learning and development in neural networks: The importance of starting small. *Cognition* 1993; 48: 71-99.

11. 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 R* 2003; 46: 773-787.12. Coltheart, M. (1981). The MRC psycholinguistic database. *Q J Exp Physiol* 1981; 33: 497-505.

13. Kucera, H, Francis, WN. *Computational analysis of present day American English*, Providence, RI: Brown University Press; 1967.

14. Nusbaum, HC, Pisoni, DB, Davis, CK. Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on speech perception progress report* 1984; 10: 357-376.

15. Bement, L, Wallber, J, DeFilippo, C, Bochner, J, Garrison, W. A new protocol for assessing viseme perception in sentence context: The lipreading discrimination test. *Ear Hearing* 1988; 9: 33-40.16. Owens, E, Blazek, B. Visemes observed by hearing-impaired and normal-hearing adult viewers. *J Speech Lang Hear Res* 1985; 28: 381-393.