Charlie Chaplin and gesture training in severe aphasia: A controlled double-blind single-case experimental design

Clémence Vibrac, Amélie Avias, Pierre-Olivier François, Marie-Eve Isner-Horobeti, Agata Krasny-Pacini

Charlie Chaplin and gesture training in severe aphasia: a controlled double-blind single-case experimental design.

<u>Author's copy of accepted manuscript:</u> Charlie Chaplin and gesture training in severe aphasia: a controlled double-blind single-case experimental design

Ann Phys Rehabil Med 2020 Feb 4;S1877-0657(20)30031-2.
doi: 10.1016/j.rehab.2019.12.010. Online ahead of print.

Clemence Vibrac, ST[1,2,3]; Amelie Avias, ST[3]; Pierre-Olivier François, ST[4]; Marie-Eve Isner-Horobeti, MD, PhD[4,5]; Agata Krasny-Pacini, MD, PhD[4,6]

[1] Pôle Ambroise Paré service ORL. Hôpitaux Civils de Colmar 39 avenue de la liberté 68024 Colmar cedex, France
[2] Pôle Psychiatrie service Psychiatrie Infanto Juvénile. Hôpitaux Civils de Colmar 39 avenue de la liberté 68024 Colmar cedex, France
[3] Centre de Formation Universitaire en Orthophonie de Strasbourg, 4 rue Kirschleger 67000 Strasbourg, France
[4] Pôle de Médecine Physique et de Réadaptation. Institut Universitaire de Réadaptation Clemenceau Strasbourg. 45 boulevard Clémenceau, F-67082 Strasbourg cedex, France
[5] Strasbourg University, Fédération de Médecine Translationnelle de Strasbourg, EA 3072 "Mitochondrie, stress oxydant et protection musculaire", France
[6] Strasbourg University, Unité Inserm 1114 Neuropsychologie Cognitive et Physiopathologie De La Schizophrenie, Département De Psychiatrie, Hôpital Civil De Strasbourg, 1 Place De L'hôpital, 67091 Strasbourg Cedex F-67000 Strasbourg, France

<u>Corresponding author:</u>
Dr. Agata Krasny-Pacini
Institut Universitaire de Réadaptation Clemenceau-Strasbourg
10 avenue Achille Baumann
BP 80096, 67402 Illkirch Cedex France
Telephone: 0033 671-28-41-52 (mobile)
E-mail: <u>agatakrasnypacini@gmail.com</u>

Abstract

Objective: to explore whether two types of gesture interventions can improve communication in severe aphasia**. Design:** pilot study performed at home in routine care using a controlled double-blind single-case experimental design (SCED), using a multiple baseline across three subjects and across two behaviors (gesture and naming)**. Setting:** home-delivered intervention by an outreach team for patients with acquired brain injury. **Participants:** three male patients with stroke-induced severe chronic aphasia, non-functional perseverative speech and severe associated impairments. **Interventions:** a passive gesture intervention in which patients watch movies selected for their intensive use of gesture and an active gesture intervention in which patients actively practice gestures through Visual Action Therapy. **Main Outcome Measure(s):** naming score, gesture score and non-verbal subscale of the Lillois Test of Communication **Results:** Gesture interventions improved the ability to gesture a list of words (Tau-U= 0.38-0.67 for combined gesture interventions effect) and increased non-verbal communication activity in all three patients. Benefits were maintained at three-month follow-up. **Conclusions**: mute films, that use intensive non-verbal communication may provide a useful add-on to speech therapy. Improving naming in very severe and chronic aphasia may not be feasible and more effort could be devoted to improving gesture-based and non-verbal communication.

Key Words: aphasia, stroke, gesture, intervention, speech therapy, single-case experimental design

Introduction

Aphasia following a stroke is a frequent and disabling condition that decreases quality of life. A number of approaches have been proposed to treat aphasia, with currently no consensus[1], but the use of gesture has been proposed as a way to enhance aphasia recovery [2]. In this approach, gestures are used to facilitate spoken output but also to compensate communication activity. Compensatory communication relies primarily on symbolic gestures that express some type of meaning[2] (e.g., hand shaped as a house, use of objects or actions such as a glass/drinking, familiar actions such as thumbs up … -see Sekine et al. for details[3]), and has shown strong communicative in people with aphasia[4–6]. Therefore training gesture in aphasia may result in better effectiveness of communication in everyday life (1) by enhancing the use of meaningful gestures by the patients to increase the comprehension of their speech; (2) by facilitating their spoken production: (3) more generally by promoting the use of non verbal communication skills. It is however unknown if this may be achieved in the presence of very severe aphasia, associated with apraxia (which hinders gesture production), and executive dysfunction (which hinders efficient planning of discourse and inhibition of meaningless gestures and spoken productions).

Although the combination of observation and practice of gestures (e.g. Visual Action Therapy- VAT)[6] seems to constitute an effective approach [8], the mere observation of an action (e .g. on a videoclip) may enhance word production[9].

The aim of the study was to explore whether two types of gesture interventions [(1) a passive gesture intervention in which patients only watch movies selected for their intensive use of gesture and non-verbal communication (e.g.: Charlie Chaplin) and (2) an active gesture intervention in which patients actively practice gestures through VAT] can improve communication in patients with stroke-induced severe chronic aphasia, through improvement of (1) meaningful gesturing ability; (2) naming ability; (3) nonverbal communication skills.

METHODS:

The study was a pilot study performed at home in routine care, using a controlled multiple baseline design across 3 subjects and 2 behaviors (gesture and naming). The first phase was a baseline without specific intervention (but non-specific time with therapist), the two following phases were gesture intervention phases applied in a balanced order (see table 1).

Table 1. Baseline and intervention description.	
Baseline	No specific aphasia training but time with therapist of same intensity (45-min sessions) as specific intervention administered in the following phases. During baseline, therapists spent time with participants, using materials not targeting aphasia but rather other cognitive functions: non-verbal logic (Logix®), mental flexibility and spatial reasoning (Tangram), visuo-spatial representation (Connect 4)
Passive gesture intervention	Watching mute films with the therapist (45-min sessions). The films were selected for the intensive use of gestures and facial expressions of their characters: mute movies (Charlie Chaplin, Buster Keaton), pantomimes (Mime Bizot), stories using sign language. In every session, different types of films were shown, to keep participants interested. Therapists did not intervene during movie screening, unless the participant asked something. At the end of the sessions, therapists asked participants questions to keep them engaged in the activity (e.g., Did you understand the movie? Are there gestures you could use? How did you like the movie?). None of the words of the repeated measures was represented in the movies.
Active gesture intervention	Visual Action Therapy, in which participants are trained to make gestures using cards and real-life objects (a telephone, a hammer, a rubber) in a progressive and structured protocol published by Helm-Estabrooks.[7] All items from the original Helm-Estabrooks protocol were used, as was half of the 12-word set. It was expected that the intervention would generalize to the untrained items of the 12-word set.

Phases length was determined by a restricted Marascuilo-Busk procedure[10], with the restriction of having at least 5 measures per phase[11] i.e. phase changes of each patient were selected at random among k= 5 possible time points, so that each phase had minimum 5 measures and maximum 10 measures, with the restriction of never having the same phase length for two patients (to respect the staggered sequential introduction of intervention in SCEDs). Note that in this procedure, the number N of subjects must be less than k. The higher the difference between k and N, the more permutations (= k !/(k-N) !) the design has and the more power to detect a significant change above 0.05 probability.

Target behaviors were measured every other intervention session. In each phase, patients had four to five 35-45-minute-long individual intervention sessions weekly, with one of the 2 speech therapists conducting the study, in their home. Phase length was determined after 5 points of baseline were obtained. The patients were not blinded to the content of each phase but they were blinded to the hypothesis of the study and to the fact that only the second and third phases had a specific content in relation to their aphasia.

Target behaviors: At each measurement, patients had to name (target behavior 1) and gesture (target behavior 2) a set of 12 words (e.g.: key, snake, glass, to dig, to color, to mix..) previously identified as impossible to name (no spoken production), selected at random from a list of 100 words administered prior to the study. The content of each set was patient-specific and remained the same throughout the phases. Gestures were scored as follows: 0: no gesture produced; 1: gesture initiated but without meaning; 2: expressive gesture but imprecise: 3: precise and expressive gesture that can be easily understood. Naming was scored as follows: 0: no spoken production; 1: inappropriate spoken production; 2: semantic or phonemic paraphasia 3: correct naming. Each measure being based on 12 words, the gesture and the naming score could vary between 0 and 36. An increase in scores meant the patient was improving. The scoring was performed based on video recording of the patient, visualized in randomized order. Assessors were therefore blinded to intervention phase when scoring the patient. Two assessors independently scored 20 % of measures for each patient and each phase, in order to calculate the inter-rater reliability of the scoring system.

Other measures included: (1) the Boston Diagnostic Aphasia Examination naming score; (2) the Lillois Test of Communication (LTC) [12,13] – non-verbal subscale (a PACE-like assessment in which patient have to make the assessor discover the picture they are looking at using any form of communication). LTC is a standardized assessment of functional communication, with an adequate inter-rater reliability, routinely used in French-speaking aphasic patients, which assesses most aspects of communication activity (motivation to communicate, body language etc…). These measurements were administered four times: (1) before baseline; (2) after the first gesture intervention; (3) after the second intervention; (4) three months after the end of all interventions (follow-up). This last measure aimed at testing maintenance of effect and further included a supplemental measure of target behaviors (12 word gesture and naming).

Inclusion criteria: patients aged 18 years and older, presenting with an ischemic left-sided stroke and a chronic (onset > three years) aphasia-related severe communication disability, defined as a complete inability to communicate with the physician at the medical visits, without the help of proxies. Inclusion were concurrent (i.e. all patients started baseline at the same time). There were no exclusion criteria.

The interventions and evaluation were performed at the patients' homes. The study conformed to Helsinki Declaration. Patients gave a written informed consent. Procedural fidelity was measured based on video-recording of the sessions by a speech therapist who did not participate in the session.

Data analysis was performed using both visual aids (using https://manolov.shinyapps.io/Overlap/) and statistical methods adapted to single-case experimental designs : randomization tests[10] based on permutations for the overall effect of the interventions for all the patients and Tarlow's Tau-U [14] for the individual effect of the interventions for each patient. Regarding the individual effect of the interventions for each patient, the following comparisons were performed: (1) baseline versus cumulative effect of both gesture interventions: (2) baseline versus first gesture intervention; (3) first gesture intervention versus second gesture intervention. When the number of measures at baseline is seven or less, Tarlow's Tau-U method may fail to detect even high degrees of trend. For this reason, in case of significant results, we checked whether Tau U effect sizes were maintained if a systemic baseline correction was used for phases of length of seven of less (i.e. using a conservative and over-correcting approach where even non-significant trends are corrected for). Randomization test was performed using ExPRT Package [15], using a within-cases comparisons of means, following a restricted Marscuilo-Busk procedure[10].

The three patients included were males, at least 4 years post middle cerebral artery ischemic stroke, presenting with a right-sided hemiplegia, severe apraxia and severe communication disability due to aphasia (Aphasia Severity Rating Scale of 0 in all patients) but also due to executive dysfunction. All had continuous intensive on-going speech therapy since their stroke (two to four sessions weekly). Other characteristics of the included patients, phase's length of each patient and interventions order are presented in table 2.

Table 2: Characteristics of patients, phase length of each patient and intervention order

	Patient 1	Patient 2	Patient 3
Age, years	65	71	43
Time since stroke, years	6	8.5	4
Aphasia Severity Rating Scale	0	0	0
BDAE naming score pre-intervention	0	33	72
Communication	Limited to three perseverative words, including « yes /no » code understandable only to familiar persons	Correct use of yes/no. Short sentences, with mainly perseverative non-functional content (e.g. « I know everything », repeated 20 times during each medical visit)	Some naming ability but no functional communication (perseverative use of single words), good non verbal abilities (2 SD above severe aphasic patients – see figure 1)
Deficits associated to the right-sided hemiplegia, severe apraxia and severe aphasia		Seizures Behavioral and cognitive dysexecutive syndrome (with severe apathy)	Behavioral and cognitive dysexecutive syndrome (with disinhibition and impulsivity)
<u>Baseline (A)*</u>			
Content	Aspecific time with therapist	Aspecific time with therapist	Aspecific time with therapist
Phase A length	10 sessions	14 sessions	12 sessions
Number of measurement points**	5	7	6
<u>First specific intervention (B)*</u>			
Content	Passive Gesture Intervention	Active Gesture Intervention	Passive Gesture Intervention
Phase B length	12 sessions	16 sessions	18 sessions
Number of measurement points**	6	8	9
<u>Second specific intervention (C)*</u>			
Content	Active Gesture Intervention	Passive Gesture Intervention	Active Gesture Intervention
Phase C length	16 sessions	10 sessions	16 sessions
Number of measurement points**	8	5	8

*for clarity of statistical comparisons, and because gesture interventions were applied in a randomized order (and therefore differed among patients), the letters A, B, C refer to the order in which phases are compared: A refers always to baseline, B to the first specific intervention irrespective of its content and C to the second specific intervention irrespective of its content.
** Target behaviors were measured every other intervention session.

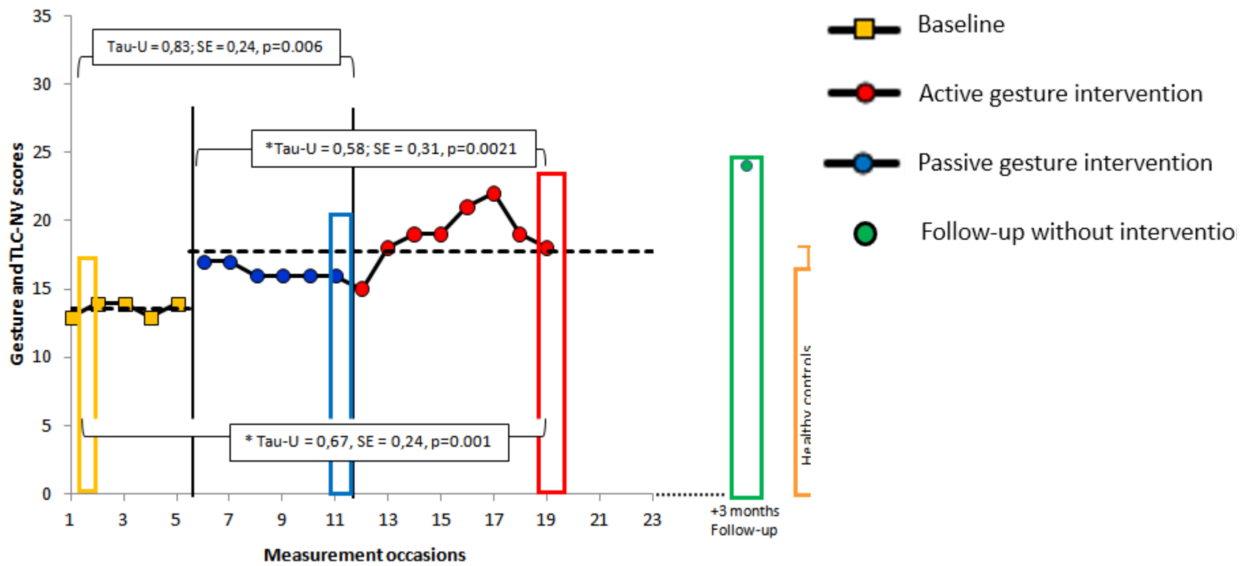
RESULTS

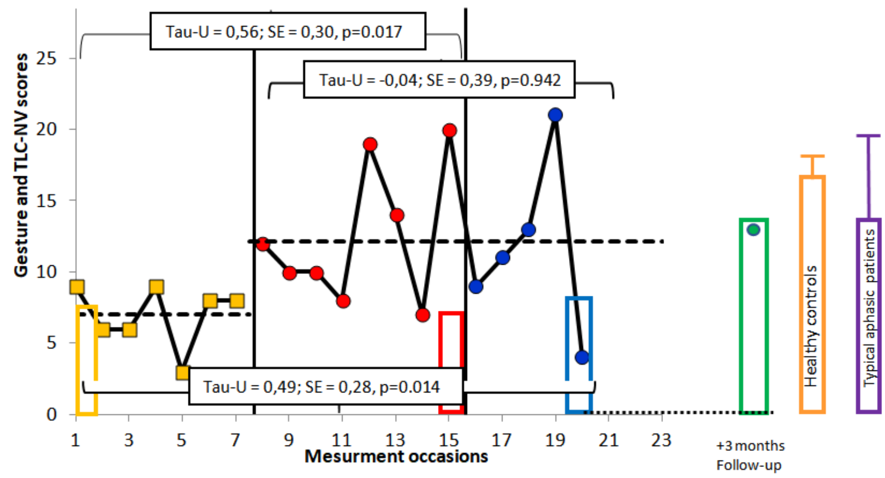
Procedural fidelity checklists indicated 83% correct implementation of intervention. Inter-rater reliability of target behaviors was found to be good (correlation of 0.92 for gesture scores and 0.94 for naming scores).

Gesture interventions improved the ability to gesture a list of words and increased non-verbal communication activity in all three patients (see figure 1 and supplemental data 1). Active gesture intervention was more effective than the passive intervention for gestures in both patients who begun with the passive gesture and for naming in patient 3 (see figure 2).

Figure 1: Evolution of Gesture and Non-verbal Communication scores over time, across patients and across phases.

Colored squares and circles represent gesture raw scores. Rectangles represent the Lillois Test of Communication non-verbal subscale scores (TLC-NV): yellow, blue, red and green rectangles represent the non-verbal score of the three patients included in the study respectively in baseline, after passive gesture intervention, after active gesture intervention and at follow-up; purple rectangles represent the typical non-verbal scores of a sample of severe aphasic patients based on Darrigrand et al.[13] *; orange rectangles represent the typical non-verbal scores of healthy controls, extracted from LTC normative data. Error bars are 1SD. Maximum gesture score is 36 and maximum LTC-NV score is 30. Vertical lines indicate phase changes (i.e. change in intervention content). Horizontal dotted lines represent mean scores of baseline and mean scores of combined gesture interventions.*


Baseline
Active gesture intervention
Passive gesture intervention
Follow-up without interventio
Gesture and TLC-NV scores
Measurement occasions
Tau-U = 0,83; SE = 0,24, p=0.006
*Tau-U = 0,58; SE = 0,31, p=0.0021
* Tau-U = 0,67, SE = 0,24, p=0.001
+3 months
Follow-up


Gesture and TLC-NV scores
Mesurment occasions
Tau-U = 0,56; SE = 0,30, p=0.017
Tau-U = -0,04; SE = 0,39, p=0.942
Tau-U = 0,49; SE = 0,28, p=0.014
+3 months
Follow-up
Healthy controls
Typical aphasic patients

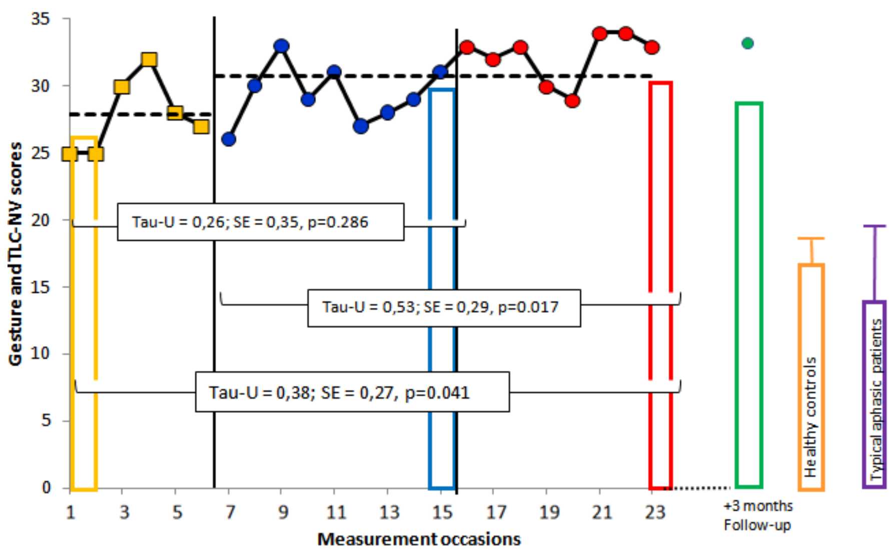
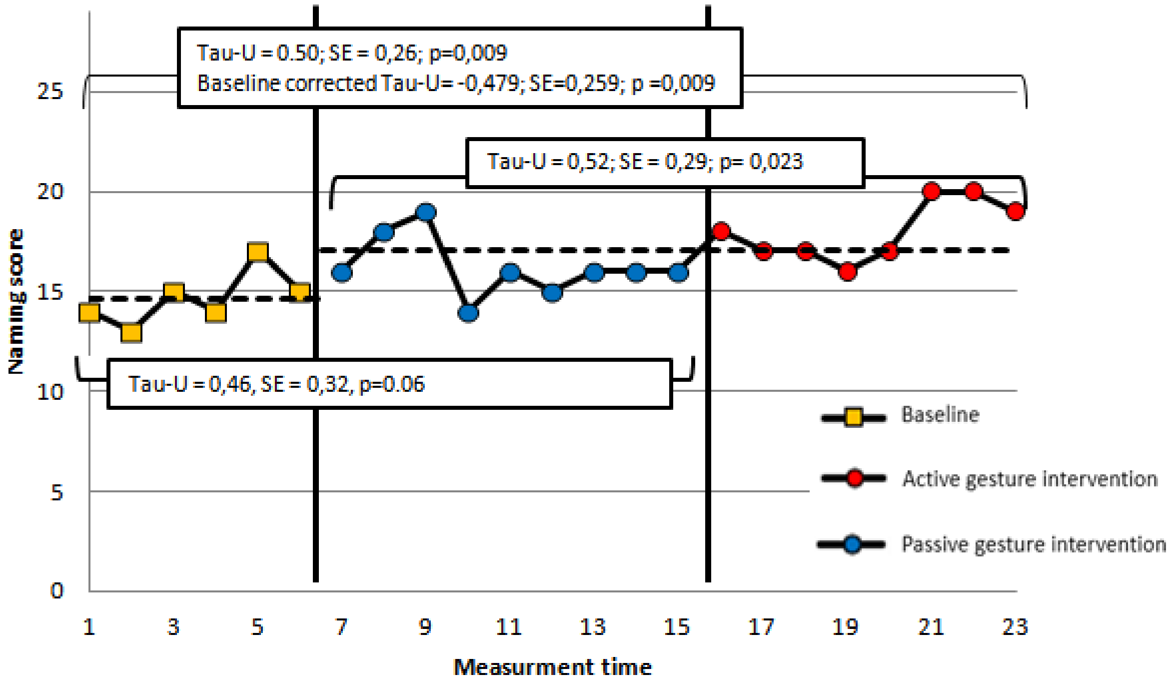

Gesture and TLC-NV scores
Measurement occasions
Tau-U = 0,26; SE = 0,35, p=0.286
Tau-U = 0,53; SE = 0,29, p=0.017
Tau-U = 0,38; SE = 0,27, p=0.041
+3 months
Follow-up
Healthy controls
Typical aphasic patients

Figure 2: Evolution of naming score across phases for patient 3.



For patient 3, naming visual analyses (see figure 2) suggested a non-monotonic trend towards improvement of baseline, and baseline correction seemed necessary, even if the trend was not detected as significant using Tarlow's method. Using the conservative baseline correction proposed by Tarlow, overall gesture intervention (passive + active interventions taken together) suggested a statistically significant inverse effect i.e. worsening in performance (baseline corrected Tau-U= -0,479; SE=0,259; p =0,009). The patient did not worsen on his naming scores, as documented by increased level (i.e. mean scores of a phase, represented by the horizontal dotted line) between baseline and gesture intervention, but naming showed a more pronounced improvement during baseline than during gesture interventions. Naming did not show changes in the two other patients (see supplemental data 2). BDAE naming scores remained unchanged (0 for patient 1, 28-35 for patient 2, 66-72 for patient 3).

Overall, statistical tests using randomization tests showed that the combined effect of both gesture interventions tended to improve gesture expression abilities (p = 0.067), but not naming ability (p = 0.53). Gesture interventions improved the ability to gesture a list of words and increased non-verbal communication activity (although the latter cannot be confirmed statistically due to the small number of measurement occasions, related to LTC administration time: 90 minutes). Benefits were maintained at three-month follow-up.

Discussion

This is, to our knowledge, one of the very few single-case experimental design (SCED) papers that used (1) a controlled, double blind design, (2) with an adequate baseline and control for time with therapist, (3) monitored treatment fidelity and (4) adequate target behavior explored for its inter-rater reliability based on video recordings of patients, scored in random order. This paper meets all current standards for SCEDs[11]. Choosing phases length for a patient at random is recommended in SCEDs but is very rarely done, while it considerably improves the methodological rigor of single case research. The study limitation is that the Marscuilio-Busk procedure may have had insufficient power (k !/(k-N) ! = 60 possible permutation) to statistically show an intervention global effect by the stastistical randomization test. Higher power would have been achieved if the design allowed to randomly select longer baselines (i.e. longer than 10 measures), allowing more k possible starting points and therefore more permutations. It is however always challenging in SCEDs for both therapist and patients to spend more time in baselines and less in intervention. A future direction for SCED research could be to collect longer baselines while training the patient on another patient-relevant function, unrelated to the intervention being tested. Here the baseline was a non-specific time with therapist, while it would have been better to use it for relevant training (e.g.: attention, executive functioning training). Nonetheless, the SCED design was strong and shows the feasibility of conducting research on small samples, especially with patients who are often excluded from clinical trials because of the numerous associated deficits, while they represent those most in need of innovative effective interventions.

A number of statistical procedures are now available to interpret SCED data, but all have their limitations. The presence of improvement trends in baseline scores is a major issue, as illustrated by patient 3' naming scores which showed a more pronounced improvement during baseline than during passive gesture intervention. A reasonable interpretation of this result is that naming showed a practice effect that probably reached a ceiling by measurement time 9 (session 18), followed by a lack of further improvement with the gesture intervention. This suggests that the patient improved more with the non-specific baseline intervention content, than with the gesture interventions. This raises the issue of improvements unrelated to intervention such as time spent with therapist. Patient 3 was young, single and quite isolated (lived in a village and was unable to drive). It is likely that spending time with motivated young speech therapists played a more significant role in improving the patient's scores than the content of the intervention itself. In this study, time with therapist was the same during all the phases which is a methodological strength and allowed to detect this non-specific, statistically significant effect.

Nonverbal modalities may be used as verbal facilitators/learning enhancement to improve verbal communication or as a communication strategy to improve total communication[16]. The latter is supported by our study, as gesture scores and TLC-NV scores improved, while the former was supported by the results of patient 3 only for whom active gesture intervention only showed significant gains in naming. Previous research has shown that gesture training alone has nonsignificant effects on verbal production and should be combined with verbal training, which is probably a reason why the intervention did not have the expected effect on naming[2].

Patients' ability to gesture words improved more with the active gesture intervention based on VAT, compared to watching mute films. All 3 patients had apraxia, which has been shown to predict the comprehensibility of gesturing irrespective of aphasia severity [17]. Apraxia may have prevented patients from benefiting from the passive intervention while the repetitive step-by-step use of gestures in VAT enabled learning of gestures in the presence of apraxia. This supports growing evidence for the use of gesture in treating aphasia[16,18], use of gesture training protocols[2] (here VAT) and the necessity of intensive treatments (here sessions weekly) [1]. However to achieve better results, other modalities (e.g. drawing, music…) of training should be included. On-going studies should confirm that Multimodal approaches, (as opposed to Constraint-Induced Aphasia Therapy limited to spoken modality) are the most effective for patients with severe aphasia[18].

Patient 1, who had started with the passive intervention, improved his gesture scores and non-verbal communication just by watching movies, though this improved further in the Visual Action Therapy. Watching mute films, which use intensive non-verbal communication may constitute a useful add-on to speech therapy in aphasic patients, as movies allow a more ecological use of communication (as opposed to the VAT that required more artificial imitation of gestures). Generalization to spontaneous discourse is a major challenge of aphasia rehabilitation[2]. The use of films may enhance this generalization because it is closer to the natural situations of communication, which is key necessity in gesture treatments [19]. Future studies could explore the effectiveness of this approach as an add-on to speech therapy

Improving naming in severe and chronic aphasia may not always be feasible and more effort could be devoted to improving gesture-based compensatory communication and non-verbal aspects of communication.

We have no Conflict of Interest

1. Brady, M. C., Kelly, H., Godwin, J., Enderby, P. & Campbell, P. Speech and language therapy for aphasia following stroke. *Cochrane Database Syst. Rev.* CD000425 (2016) doi:10.1002/14651858.CD000425.pub4.
2. Rose, M. L., Raymer, A. M., Lanyon, L. E. & Attard, M. C. A systematic review of gesture treatments for post-stroke aphasia. *Aphasiology* **27**, 1090–1127 (2013).
3. Sekine, K., Rose, M. L., Foster, A. M., Attard, M. C. & Lanyon, L. E. Gesture production patterns in aphasic discourse: In-depth description and preliminary predictions. *Aphasiology* **27**, 1031–1049 (2013).
4. Rose, M. L., Mok, Z. & Sekine, K. Communicative effectiveness of pantomime gesture in people with aphasia. *Int. J. Lang. Commun. Disord.* **52**, 227–237 (2017).
5. de Beer, C. *et al.* How Much Information Do People With Aphasia Convey via Gesture? *Am. J. Speech Lang. Pathol.* **26**, 483–497 (2017).
6. Hogrefe, K., Ziegler, W., Weidinger, N. & Goldenberg, G. Comprehensibility and neural substrate of communicative gestures in severe aphasia. *Brain Lang.* **171**, 62–71 (2017).
7. Helm-Estabrooks, N., Fitzpatrick, P. M. & Barresi, B. Visual action therapy for global aphasia. *J. Speech Hear. Disord.* **47**, 385–389 (1982).
8. Ertelt, D. & Binkofski, F. Action observation as a tool for neurorehabilitation to moderate motor deficits and aphasia following stroke. *Neural Regen. Res.* **7**, 2063–2074 (2012).
9. Bonifazi, S. *et al.* Action observation as a useful approach for enhancing recovery of verb production: new evidence from aphasia. *Eur. J. Phys. Rehabil. Med.* **49**, 473–481 (2013).
10. Levin, J. R., Ferron, J. M. & Gafurov, B. S. Comparison of randomization-test procedures for single-case multiple-baseline designs. *Dev. Neurorehabilitation* **0**, 1–21 (2016).
11. Tate, R. L. *et al.* Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychol. Rehabil.* **23**, 619–638 (2013).
12. Rousseaux, M., Daveluy, W. & Kozlowski, O. Communication in conversation in stroke patients. *J. Neurol.* **257**, 1099–1107 (2010).
13. Darrigrand, B. *et al.* Communication impairment and activity limitation in stroke patients with severe aphasia. *Disabil. Rehabil.* **33**, 1169–1178 (2011).
14. Tarlow, K. R. An Improved Rank Correlation Effect Size Statistic for Single-Case Designs: Baseline Corrected Tau , An Improved Rank Correlation Effect Size Statistic for Single-Case Designs: Baseline Corrected Tau. *Behav. Modif.* **41**, 427–467 (2017).

15. Gafurov, B. S. & Levin, J. R. *ExPRT (Excel Package of Randomisation Tests): Statistical Analyses of Single-Case intervention Data*. (2017).
16. Pierce, J. E., O'Halloran, R., Togher, L. & Rose, M. L. What Is Meant by 'Multimodal Therapy' for Aphasia? *Am. J. Speech Lang. Pathol.* **28**, 706–716 (2019).
17. Hogrefe, K., Ziegler, W., Weidinger, N. & Goldenberg, G. Non-verbal communication in severe aphasia: influence of aphasia, apraxia, or semantic processing? *Cortex J. Devoted Study Nerv. Syst. Behav.* **48**, 952–962 (2012).
18. Rose, M. L. *et al.* Constraint-induced or multi-modal personalized aphasia rehabilitation (COMPARE): A randomized controlled trial for stroke-related chronic aphasia. *Int. J. Stroke Off. J. Int. Stroke Soc.* 1747493019870401 (2019) doi:10.1177/1747493019870401.
19. de Ruiter, J. P. & de Beer, C. A critical evaluation of models of gesture and speech production for understanding gesture in aphasia. *Aphasiology* **27**, 1015–1030 (2013).

Supplemental data 1: details of gesture scores analyses

Visual analysis and overlap-based effect size: Analyzing visually the graphs, level (i.e., the mean scores of each phase) improved in all participants during gesture interventions. For participant 2, measurement variability increased during gesture intervention possibly because of intervention intensity inducing fatigue. Participant 3 showed an important variability at baseline and an important increase in scores in the first 4 measures (possibly because of a practice effect) followed by a decrease in scores in the last 2 baseline points. Overlap was calculated by using Tarlow's Tau-U taking into account the presence of data trends if necessary (Table 2). Tarlow's method first checks for trends at baseline and then corrects the data, depending on whether there are significant trends or not (baseline corrected Tau-U).

Tarlow's method indicated that no baseline correction was needed (non-significant trend for all comparisons). Therefore, results of the first 4 columns of Table 2 are reported in Figure 1. When the number of measures at baseline is ≤ 7, Tarlow's Tau-U method may fail to detect even high degrees of trend. For this reason, with significant results, we checked whether Tau-U effect sizes were maintained if a systemic baseline correction was used for phase length ≤ 7 (i.e., using a conservative and over-correcting approach where even non-significant trends are corrected for). This is reported in Table 2.

Table 2. Comparison of Tarlow's Tau-U values with and without baseline correction by using http://ktarlow.com/stats/tau/

	Tau	SE_{Tau}	**p**	**Baseline Corrected Tau ‡**	SE_{Tau}	**p**
Participant 1						
A versus (BC)	0.67*	0.241	0.001	0.67*	0.241	0.001
A versus B	0.826*	0.241	0.006	0.826*	0.241	0.006
B versus C	0.581*	0.308	0.021	0.581*	0.308	0.021
Participant 2						
A versus (BC)	0.490*	0.276	0.014	0.49*	0.276	0.014
A versus B	0.564*	0.302	0.017	0.564*	0.302	0.017
B versus C	-0.036	0.392	0.942	Not computed because phase length ≥ 8 reliably rejects the need for trend correction		
Participant 3						
A versus (BC)	0.381*	0.273	0.041	-0.434*	0.266	0.019
A versus B	0.260	0.353	0.286	-0.381	0.338	0.11
B versus C	0.529*	0.291	0.017	Not computed because phase length ≥ 8 reliably rejects the need for trend correction		

‡ computed if comparing phase length ≤ 7 measures.

*Statistically significant at $p<0.05$

Letters A, B, C refer to the order in which phases are compared: A always refers to baseline, B the first specific intervention regardless of its content and C the second specific intervention regardless of its content (see Table 1 and Figure 1 for the participant-specific phase B and C randomized content).

Results were unchanged for participants 1 and 2. For participant 3, comparison of baseline versus passive gesture intervention remained non-significant, but the effect of the overall gesture intervention conferred an aberrant statistically significant worsening in performance (bcTau-U = -0.434*, SE = 0.266, p = 0.019). This is not an exceptional finding and was raised by Tarlow: if the projected trend line crosses the ceiling of the measurement scale, Tau-U values tend to become statistically significant toward worsening (because participants cannot follow a trend that goes beyond the maximum score, in this case 36). The following figure illustrates the Theil-Sen regression line used in baseline corrected Tau-U (graph obtained through www.manolov.shinyapps.io/overlap) and detrended data that are used for bcTau-U calculation.

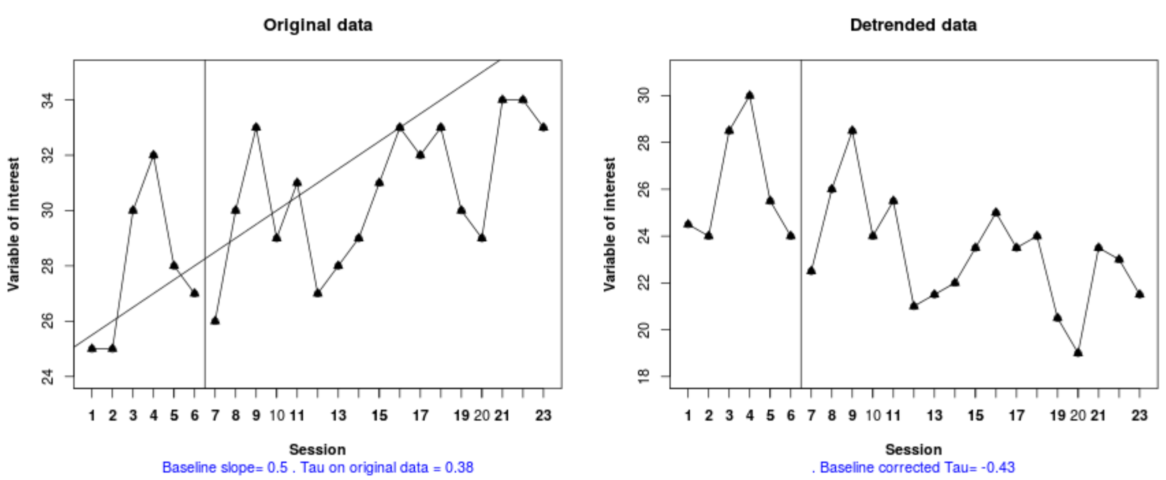


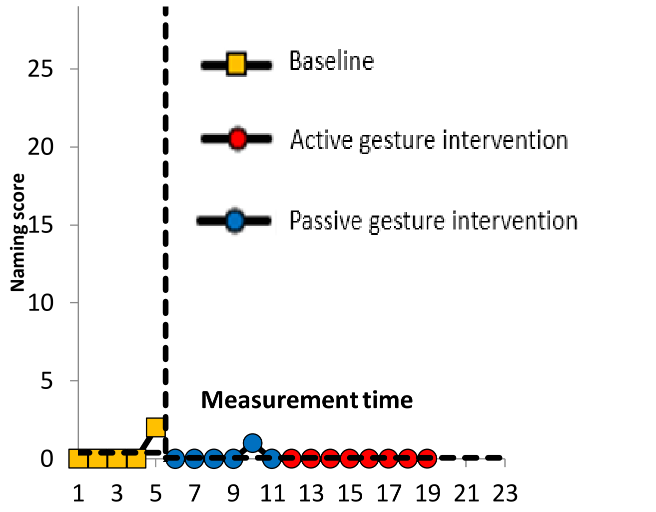
Figure 2. Theil-Sen regression line used in baseline-corrected Tau-U and detrended data (graph obtained by using www.manolov.shinyapps.io/overlap)

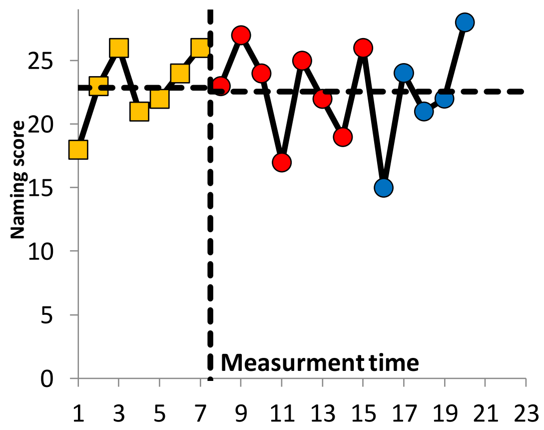
Because bcTau-U baseline correction uses a correction of linear trend, it can lead to biased baseline correction if the trend is not linear: the baseline of participant 3 was visually not linear and not even monotonic; therefore, it was decided to interpret the intervention effect without baseline correction in the main results of the paper.

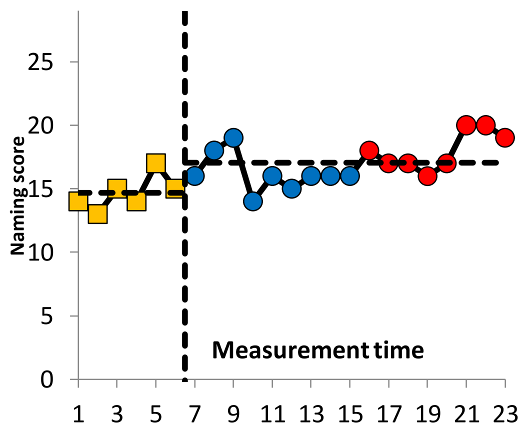
Another issue with Tarlow’s method is that Tau-U effect sizes decrease as the number of measures in intervention phase increases. This explains why for participant 1, for example, Tau-U was greater when computed for A versus B (0.83) as compared with A versus the cumulative effect of B and C (0.67).

Supplemental data 2. Naming results.

	Tau-U	SE_{Tau}	p
Participant 1			
A versus (BC)	-0.202	0.318	0.435
A versus B	-0.084	0.425	0.892
B versus C	-0.320	0.358	0.312
Participant 2			
A versus (BC)	-0.008	0.316	1.000
A versus B	0.040	0.365	0.907
B versus C	-0.109	0.390	0.714
Participant 3 ‡			
A versus (BC)	0.497 *	0.256	0.009
A versus B	0.462	0.324	0.061
B versus C	0.517 *	0.294	0.023







Highlights

- Improving naming in severe and chronic aphasia may not be feasible.
- Non-verbal modalities may be used as a communication strategy to improve total communication in severe aphasia but also in some other conditions as verbal facilitators to improve verbal communication.
- Mute films that use intensive non-verbal communication may be a useful add-on to speech therapy.
- A controlled double-blind single-case experimental design is feasible and useful for study of individuals with severe disabililty and associated deficits who are usually excluded from large clinical trials.
- Statistical analysis must take into account the improvement trend in baselines.