ACES: A Cross-Discipline Platform and Method for Communication and Language Research

Joshua Hailpern*, Marina Danilevsky†, Andrew Harris†, Sunah Suh†, Reed LaBotz†, Karrie Karahalios†
*HP Labs, Palo Alto, CA †University of Illinois, Urbana, IL
*joshua.hailpern@hp.com †(danilev1,harris78,sunahsuh,labotz1,kkarahal)@illinois.edu

ABSTRACT
While conducting research focused on individuals with impairments is vitally important, such experiments often have high costs (time and money), and researchers may be limited in the instructions they can give, or participant feedback they can gather (due to the impairment). We present how an impairment emulation system (ACES) can be used by researchers in the behavioral sciences. By repurposing this new technology within the context of a "traditional" psychology experiment, we were able to analyze impaired linguistic and communication in a manner that was not possible without a system such as ACES. Our experiment on 96 participants provided strong support for a theory in the aphasia psychology community, and uncovered new understandings of how people communicate when one interlocutor's speech is distorted with aphasia. These findings illustrate a new direction of HCI research that directly helps researchers in Psychology, Communication, and Speech and Hearing Science.

Author Keywords
Aphasia; Assistive Technology; Disabilities; Empathy; Emulation Software; Language; Speech; Messaging

ACM Classification Keywords
K.4.2 Social Issues: Assistive technologies for persons with disabilities

INTRODUCTION
Working with individuals that have impairments can be challenging for researchers from Computer Science to Psychology to Speech and Hearing Science. Researchers need expertise in the specific impairment, experimental costs are often higher to accommodate the physical/cognitive/language challenges, and the studies themselves can take longer and cost more to run [21]. In addition, if there is a language, cognitive or perceptual impairment, instructions given or gather able participant feedback may be limited.

Given these challenges, consider the large and lasting impact of a tool that can allow researchers (in technology or behavioral sciences) to more easily conduct experiments to aid/study individuals with impairments by running "simulated" studies with "typical" individuals. Such a tool/technique could reduce the impairment specific burden on researchers, while allowing them to test theories, run studies with more complex instructions, gather more detailed participant feedback, and quickly explore solutions before running a large N study with the target population. For those researchers with Institutional Review Boards, such a tool would expedite the approval process because participants would not be from a "vulnerable" population. Note that we are not advocating the abandonment of studies that directly work with individuals that have an impairment; rather we are advocating a solution to improve such studies, making them more time- and cost-effective by emulating a disorder through software.

In 2011, the **A**phasia **C**haracteristics **E**mulation **S**oftware, or **ACES**, was first published. ACES allows its users (e.g., caregivers, therapists and family) to experience, firsthand, the communication-distorting effects of aphasia [16]. The original goal was to allow users to build empathy and understanding by "walking in the shoes" of someone who has aphasia. Later that year, the authors validated the realism of their aphasia emulation by running (and passing) two Turing Tests [17]. We hypothesize that ACES' applications and implications are far broader than the original authors had intended and have implications for technologists and non-technologists alike.

This paper illustrates how emulation systems, like ACES, can be used by researchers in traditional Psychological, Communication or Speech and Hearing Science research. Unlike working directly with many individuals that have cognitive or communication impairments, ACES' users are neurologically typical, can provide full oral and written responses to questions, and can understand and cary out complex instructions. Through an experiment with 96 participants, we show that ACES-like systems can provide a cheap, fast and effective tool to conduct research on "simulated" individuals with impairments. We demonstrate how this approach can uncover new understandings about interpersonal communication, and provide new evidence to support theories of aphasia communication that *could not have been found without a system like ACES*. Thus, the foremost contribution of this paper is the demonstration that ACES-like CMC systems have many benefits to test and validate theories, and run "traditional" Psychology experiments targeting individuals with impairments.

RELATED WORK

As this work leverages an aphasia emulator, we describe aphasia and other tools that have emulated non-language disorders. We then highlight the relevance of ACES-like solutions for researchers by briefly discussing how language and communication is examined the existing technical and non-technical literature.

Aphasia

Aphasia is a term that describes an acquired language disorder that impairs an individual's ability to produce and understand language [5] in both written and spoken forms[6]. Aphasia is associated with individuals that have brain damage (e.g. stroke), though the manifestation (symptoms and severity) can vary based on the location and type of damage to the brain. Based on the variety of aphasia "flavors," classification systems were created to help researchers, clinicians, and individuals [15, 42]. HCI research on aphasia has largely focused the remedying communication challenges via image based communication in mobile phones [3], and day-to-day interaction [9, 1]. Also of note is the technology based research to aid individuals with aphasia in speech therapy [34] and scheduling their daily activities [31].

Disorder Emulation

In 1967, Weizenbaum et. al. introduced Eliza [47], a computer AI that attempted to emulate a Rogerian psychotherapist. Four years later, Colby et. al. published PARRY [11], and interactive computer model of paranoia. While both Eliza and PARRY were emulation systems, they did not allow a "typical" individual to experience what it was like to communicate *as* a psycho-therapist or have paranoia. These stand in sharp contrast to ACES, which allows the user to experience what it is like to have aphasia. While not a language-based emulation, Takagi's built an emulator [43] that allows a web developer to see (through obfuscation) the readability of web content to a blind screen-reader user.

The Study of Communication

Through a detailed examination of the subtle but important language patterns within conversations, researchers can uncover and quantify these changes in interpersonal communication. The examination of these patterns is not an invention of this research, nor of the computer age. Many researchers in fields such as Communication, Linguistics, and Psychology have explored how and why people communicate [14, 7, 10, 4, 48]. It is in this large corpus of existing work that we see future applications and uses of ACES, and we draw on these works in defining measures to study ACES' impact.

The examination of language and textual communication has permeated the field of human computer interaction, most notably in the field of **C**omputer-**M**ediated **C**communication (CMC). Issues of trust in computer, textual and other forms of CMC have been widespread [8, 49, 45, 32]. While some of these works has focused on examining language, others

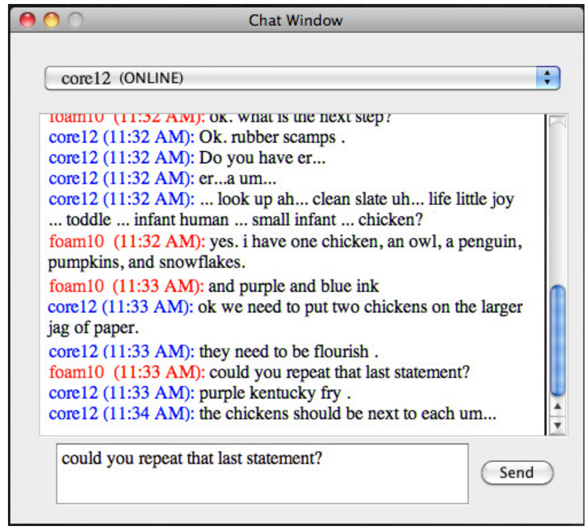


Figure 1. ACES Instant Message Window
Screenshot is of a conversation from the study in this paper. The current (red) user's partner (blue) has their text distorted.

explore the outcome of the interactions. Scissors [39, 40] examined the alignment of text and mimicry. Specifically in the domain of **I**nstant **M**essage (**IM**), Wang et. al. [46] studied, via hand-coding, the content of messages (e.g. as ideation, strategy, responses). Their focus was on the examination of conversational distance (using the divergence metrics for distribution of categories). Leshed's PhD. dissertation [28] examined language features to explore interpersonal behaviors on which people interact over IM. These works closely parallel the type of CMC questions posed in this paper (pertaining to linguistic impact and conversation outcomes). We draw on these works, and others in the CMC literature to ground many of the linguistic measures we employ.

A BRIEF OVERVIEW OF ACES

Given the prominence of ACES in this work (though the system itself is not a contribution of this paper), we briefly highlight its functioning here. For a complete description of its system and validation, please see [16, 17]. ACES is a configurable probabilistic model of the linguistic distortions associated with aphasia situated in an IM client. ACES' design is highly robust, and can emulate many types and severities of aphasia. When a user sits at a computer, they type their message into a IM window (Figure 1). This message is then distorted (based on the current settings in ACES' probabilistic model), and sent over the AOL IM network. The distorted message is then received by a conversation partner[1]. Nowhere in this process do either conversation participant role-play[2] or guess "how" the messages should be distorted [16]. Rather, the distortions provided by ACES are highly realistic, in that they passed two Turing Tests on speech disorder experts [17]. Further, ACES creates full logs of all conversations including the message *intended* to be sent, as well as the distorted message that *was* sent. In contrast to this prior work on ACES

[1] ACES allows the user to see or not see the distorted message. This parallels how some individuals with aphasia know they are making errors, while others do not.
[2] As currently done in many Speech Therapy classrooms.

which focused on the fidelity of the system and peoples emotional response, this paper examines the impact of distortions on the conversation itself (linguistics and quality).

MOTIVATION, SCOPE & RESEARCH QUESTIONS

We consider ACES-like systems to have many benefits to Researchers in Psychology and Communication beyond that of empathy building (as shown in [16]):

- **Controls:** ACES can consistently emulate a type of aphasia (unlike real individuals, who can have huge variation in their aphasia manifestations), allowing for a highly controlled experimental setup.
- **Customizable:** ACES' distortions can be easily manipulated to allow for conversations to simulate mild to severe aphasia across multiple sub-types, allowing researchers to test any specific community they wish to target.
- **Low Cost:** Unlike working with individuals that have an impairment [21], using ACES is lower in cost, time, and recruitment effort. In addition, there are few individuals with a given impairment within a population, whereas ACES' users can be nearly anyone from the population.
- **Feedback:** Many individuals with aphasia (or other impairments) have substantial challenges understanding instructions and/or responding to questions. All ACES participants can easily answer questions about their experience because they do not have aphasia themselves.
- **Intent:** With language disorders that alter grammar and syntax, it is very difficult (if not impossible) to truly understand what an individual intends to say. ACES provides researchers with transcripts of both what was said, and what individuals *intended* to say.

Research Questions

However, these broader potential applications of ACES have never been validated. Our goal is to demonstrate how systems like ACES can be used in traditional research settings in Psychology, Communication or Speech and Hearing Science and uncover new understandings of how people communicate when one interlocutor's speech is distorted with aphasia. To this end, we seek to answer the following research questions:

Question 1 Can ACES be used in a known experimental paradigm from Psychology, Communication or Speech and Hearing Science? *Showing that ACES can be integrated in traditional research setting.*

Question 2 Can ACES' distortions impact conversation style in a known way? *Showing that ACES interactions produce known changes in conversation style.*

Question 3 Can transcripts of *intended* conversation style be useful? *Showing that one of the main features of ACES transcripts over "real" transcripts has utility to researchers.*

Question 4 Can ACES be used to uncover new understandings of interpersonal interaction? *Showing that ACES is a rich platform for uncovering new phenomena.*

	Description
Step #1	Demographic Questions
Step #2	Conversation with Partner, Structure #1
Step #3	Questions About 1^{st} Conversation
Step #4	Conversation with Partner, Structure #2
	Participant Roles & Task are Switched
Step #5	Questions About 2^{nd} Conversation
Step #6	Final Set of Questions

Table 1. Study Session Order

It should be noted, that while we are seeking to design, and execute a traditional-style experiment that may be found in Psychology, Communication or Speech and Hearing Science, **it is not our goal to conduct every possible analysis that a research in said field would run.** Rather, we wish to demonstrate that this class of technology has new, broader applications beyond empathy building.

We use these questions to guide and shape our experimental design, measures and analysis.

STUDY DESIGN

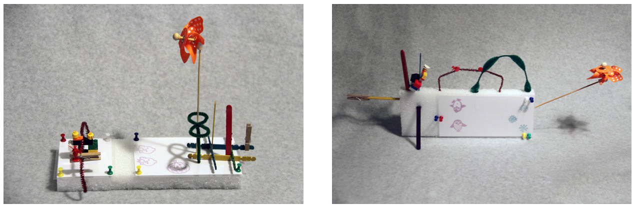
To directly address our Research Questions we followed a known experimental design[3] from Psychology and Communication called Novice-Novice Design [7, 10, 4, 48]. In these studies, two participants are asked to perform a specific activity together (in which neither have prior experience or expertise with[4]). One such example is the "map task," developed by Anderson & Bard in the UK [4]. In the "map task," each participant has a map with landmarks on it. One participant's map has a path drawn/highlighted on it, and that participant must give the path (via conversation) to their partner, so as to recreate the path. This experimental paradigm is used to explore language, communication, approach to problem solving, shared perspective, and many more other psychological and communication questions.

The specific style of Novice-Novice design we followed will be referred to as "**W**rite **I**t - **D**o **I**t" design (**WIDI**), following the work of [7, 10]. In this experimental context, two participants are separated and assigned a ***Task***. One of the participants (***Writer Task***) is told to instruct the other participant (***Doer Task***) in building a replica of a pre-built ***Structure*** which is only visible to the *Writer*. Both participants are allowed to communicate, though only the *Writer* is allowed to view and manipulate (though not take apart) the pre-built *Structure*, while only the *Doer* is allowed to view and manipulate the raw pieces.

Ninety-six individuals (grouped in subject-pairs) engaged in WIDI IM conversations using ACES. To test our above research questions, we utilized a between-subject 2x3 factorial design. The first factor in our factorial design compared the *Writer Task* to the *Doer Task*. The second factor in our factorial design compared a ***Control Group*** (where subjects did not experience aphasic distortions) to two ***Treatment Groups***: an ***Aphasia Writer Group*** (where the participant in the *Writer*

[3]Directly addressing Research Question 1.

[4]Compared to Expert-Novice where one participant has expertise in the activity.



Horizontal Structure *Vertical Structure*

Figure 2. Pictures of Structures Used in Experiments

Task experienced distortions of their text first-hand) and an ***Aphasia Doer Group*** (where the participant in the *Doer Task* experienced distortions of their text first-hand)[5]. We refer to the participants who have their text distorted as having the ***Aphasia Role***[6], and to the participants receiving distorted text as having the ***Typical Role***.

All participants received the same experimental protocol, same activity, and same *Structures* to build. Subjects did not know about any other *Groups*, and were blind with respect to which *Group* they were assigned.

Experimental Protocol

Each study **Session** consisted of 6 steps lasting a total 90 minutes (Table 1). A session involved two 30-minute IM conversations between a pair of participants (a subject-pair) who did not know each other. Each participant played both *Roles*: the *Aphasia Role* in one conversation and the *Typical Role* in the other. Each participant also performed both *Tasks*: the *Writer Task* in one conversation and the *Doer Task* in the other. Participants switched *Role* and *Task* between conversations (see Table 2). Subject-pairs remained physically separated. Each member of a pair was placed in separate identical rooms with identically set-up 21” iMac computers. All questionnaires were administered digitally.

All participants completed a brief demographic survey to assess their background and prior knowledge of aphasia (if any). This allowed us to ensure equal background and pre-knowledge across *Groups*. Upon completion of these pre-study questionnaires, all participants were given an identical explanation of the study protocol. All participants (regardless of *Group*, *Role* or *Task*) were told that they would have two thirty-minute IM conversations with each other [7].

During these conversations one member of the subject-pair would take on the *Writer Task*, while the other member of the subject-pair would take on the *Doer Task*. The *Writer* would be given a completed *Structure* (assembled and glued), while the *Doer* would be given a bag with all the pieces needed to recreate said *Structure* (plus some extras). Participants were told that after the first conversation, they would switch *Tasks* (ensuring that each participant would play both *Tasks*). Only participants in the *Aphasia Writer & Aphasia Doer Groups* were informed about aphasia and the aphasia distortions; both participants were told that the distortions would occur, and which participant would be having their text distorted.

Conversation #1		**Conversation #2**	
User A	User B	User A	User B
APH Writer	TYP Doer	TYP Doer	APH Writer
Structure #1	*Structure #1*	*Structure #2*	*Structure #2*
APH Writer	TYP Doer	TYP Doer	APH Writer
Structure #2	*Structure #2*	*Structure #1*	*Structure #1*
APH Doer	TYP Writer	TYP Writer	APH Doer
Structure #1	*Structure #1*	*Structure #2*	*Structure #2*
APH Doer	TYP Writer	TYP Writer	APH Doer
Structure #2	*Structure #2*	*Structure #1*	*Structure #1*
TYP Doer	TYP Writer	TYP Writer	TYP Doer
Structure #1	*Structure #1*	*Structure #2*	*Structure #2*
TYP Doer	TYP Writer	TYP Writer	TYP Doer
Structure #2	*Structure #2*	*Structure #1*	*Structure #1*

Table 2. Counter Balance Permutations of *Role*, *Task*, and *Structure*
APH = Aphasia Role — *TYP* = Typical Role

Two different *Structures* (see Figure 2) were used (one for each conversation). All participants used the same *Structures* (though the order of *Structures* were fully counterbalanced). *Structures* were built from assorted craft items and toy bricks (following [7]). While some pieces did overlap between the two *Structures*, each *Structure* contained components not found in the other *Structure*. Further, for any piece in a *Structure* “alternative extras” were available (different color, size, type, etc.) for the builder to choose from. This helped ensure a lively dialogue during the conversations.

At the end of both IM conversations, participants were administered a questionnaire to gauge their perspective on conversation quality. This questionnaire consisted of short answer questions, and a series of Likert Scale questions (see Dependent Measures below). Participants were remunerated $25 for their participation in the experiment. A $30 bonus was awarded to the winning team in each *Group*, as determined by the team whose assembled *Structure* was the most complete (see *Objective Quality* in Dependent Measures bellow).

Confound Counterbalancing

Given the number of other factors inherently present in this type of experimental setup, there were at least three potential confounding effects (*Structure* order, *Task* order, and *Role* order). To help control for these potential confounding effects, we used counterbalancing, a method commonly employed to help avoid confounding from order of task and presentation. When counterbalancing, all permutations of the confounders are included in an attempt to minimize any bias due to these confounders that are not central to the experimental question.

We fully counterbalance *Structure*, *Task*, and *Role* in Table 2, requiring each pair to switch both *Role* and *Task* between conversation #1and #2. This requires a minimum of 8 subject pairs to fully counterbalance the experimental design. Our cohort of ninety-six individuals therefore consisted of three *Groups*. Thirty-two subjects were in the *Aphasia Writer Group*. thirty-two in the *Aphasia Doer Group* and thirty-two in the *Control Group*. Thus we had 16 subject-pairs in each *Group* (doubling the minimum number of subject pairs to fully counterbalance the design). Because the final data set

[5]Given that the two *Tasks* require different goals, we were unsure if the language and communication would be differently effected by ACES distortions. We therefore employed three *Groups* rather than just two (Control vs. Treatment) by linking *Task* and *Role*, attempting to account for any inherent bias from the *Task*.

[6]In this experiment, participants in the *Aphasia Role* could not see the distorted message. This parallels how some individuals with aphasia do not know their errors.

[7]Participants were logged into IM accounts created for this study, not requiring subjects to disclose their own user names or passwords.

has an equal number of logs from conversation #1 and #2, we have also counterbalanced for learning effects.

Types of Aphasia

Based on the Boston classification system[15, 42], there are numerous types of aphasia each of which can manifest in varying severities. Subsequently, there are many permutations of distortions that could be applied in this study. To ground this experiment, we leveraged one of the models created and validated during the Turing Test study of ACES[17]. This model was based on an Anomic subject named Wolf from the unpublished data files used in [29, 30] provided to the researchers by Lise Menn, University of Colorado. Individuals with Anomic aphasia have difficulty with selecting and producing correct content words, though their grammar is generally correct. For example, words may be replaced by other words that are semantically related ('birthday' with 'anniversary' or 'cake'), that have no semantic relationship (cat with airplane), that have similar phonetic sounds ('population' with 'pollution'), or non-words ('castle' with 'kaksel').

Dependent Measures

Conversation Quality can be measured in two important and different ways [23, 12]. First is ***Objective Conversation Quality***, or a measure of the effectiveness of the conversation, that can be measured consistently regardless of the persons measuring. *Objective Quality* does not take opinion or subjective judgment into account. The second measure is ***Perceived Conversation Quality***, or how each conversation partner felt about their experience or conversation. This technique dispenses with the objectively measurable features in favor of more intangible issues of assessing performance.

To produce an *Objective Quality* score, each *Structure* built during a conversation will be scored on completeness and accuracy. For both *Structures* a list of all the connections was created. A *connection* is where one piece touches another, or how one piece is deformed/shaped/placed. A connection can be graded (as applicable) on both placement and orientation (1 point each). For every piece put on a *Structure* which does not belong, one point was deducted. Both *Structures* are of the same relative complexity[8]. To ensure cross *Structure* comparisons we can use percentage completion rather than raw score.

To assess each participant's *Perceived Conversation Quality*, a post conversation questionnaire was utilized. The questionnaire consisted short answer questions, the **I**owa **C**ommunication **R**ecord (**ICR**) [13] and the **I**nterpersonal **C**ommunication **S**atisfaction **I**nventory (**ICSI**) [20]. The ICR[9] and ICSI[10] are a series of Likert Scale questions, whose mean is an indication of *Perceived Quality*.

In addition to the two measures of *Conversation Quality*, ACES logged all IM messages sent. For the *Aphasia Writer & Aphasia Doer Groups*, both intended *and* distorted versions of each message was logged. Thus, the language used during the activity can also be analyzed (see next subsection).

[8]The horizontal and vertical *Structures* have 57 and 58 connections respectively.

[9]Ten 9 point likert scale questions, the lower the ICR score, the better the conversation.

[10]Nineteen 7 point likert scale questions, the higher the ICSI score, the better the conversation. Because the measure is intended to assess spontaneous conversations, some questions did not relate to the experimental design. We therefore asked a subset of 10 question (#1,2,3,5,8,9,11,12,14,16,18).

	Control Group	Aphasia Writer Group	Aphasia Doer Group
Size (N)	32	32	32
Group Age	30.75 (10.76)	29.75 (9.41)	27.00 (11.69)
Sex (% male)	46.88	43.75	50.00
Educational Attainment †	2.00 (1.30)	1.94 (1.22)	1.44(1.37)
% Know what Aphasia is	43.75	46.88	31.25
% Taken Class on Aphasia	21.88	25.00	12.50
% Worked with Aphasia	3.12	6.25	0.00
% Family/Friend with Aphasia	3.12	3.12	0.00

Table 3. Write-it Do-it Participants
Continuous variables presented as mean (sd), p-values calculated by GEE
† Education attainment scores were 0=High School, 1=AA, 2=BS/BA, 3=MS/MA, 4=PhD

Language, Linguistics and Parts of Speech

To address the research questions directly targeted in this paper (specifically Q2 and Q3), we must examine the language uses by participants (as recorded by ACES logs). Raw number of lines (how many messages sent) and words (how many words were sent over all their messages) are coarse linguistic metrics that can be extracted from ACES logs. Parts of Speech (POS) is an important set of linguistic metrics for researchers in psychology, communication, NLP [22] as well as CMC [33, 39]. Measures such as POS, and number of lines/messages, are accepted measures for analyzing conversations, that have been used in the past to examine aphasic speech [37]

At a high level, POS can be divided into Function Words (e.g. articles, pronouns, interjections etc) and Content Words (e.g. nouns, verbs, adjectives). Function Words provide the grammatical structure to a sentence, while Content Words provide the the meaning to a sentence. To this end, we extracted Function Words and Content Words, and further subdivided Content Words into Adjectives, Verbs, Nouns, and Adverbs. There is, however, a problem with using raw occurrence of each POS, since, as users talk more (increasing the number of words generated), the occurrence of each POS will increase as well. To normalize across users, we therefore examine the ratio of each POS metric to the total words produced, rather than raw counts (allowing for comparison between the 3 *Groups*).

Analysis

To address our above Research Questions, we propose the following analytical process. By conducting a WIDI experiment, and demonstrating that participants can perform the assigned activity we demonstrate that ACES can be used within a known experimental paradigm (Question 1).

There exists a well established theory in the Psychology literature, Aphasia Adaptation Theory, that states that while aphasic distortions are imposed upon and individual by their impairment, they can learn behaviors that change their speech patterns in constructive ways [27, 19, 24]. These changes may be "preventative adaptation" or "corrective adaptation" [36, 35]. Regardless, [26] suggest that many of the "symptoms" or characteristic output of individuals with aphasia may

Language	Control Group	Aphasia Writer Group	Aphasia Doer Group
***Writers'* Lines**	84.53 (40.63)	99.06 (54.81)	81.63 (40.90)
***Writers'* Words**	797.60 (173.78)	602.19 (243.53)	771.88 (249.80)
***Doers'* Lines**	57.72 (27.29)	68.47 (41.97)	56.25 (22.78)
***Doers'* Words**	282.88 (103.72)	376.81 (192.56)	228.16 (124.67)

Table 4. Raw Count (#) of Users' Language
Values represent Mean (sd)

Writer's Language	Control Group	Aphasia Writer Group	Relative Change
% Function Words	0.43 (0.02)	0.40 (0.06)	▽
% Content Words	0.57 (0.02)	0.60 (0.06)	△
% Adjectives	0.12 (0.03)	0.13 (0.03)	–
% Verbs	0.13 (0.02)	0.11 (0.02)	▼
% Nouns	0.28 (0.03)	0.32 (0.05)	▲
% Adverbs	0.05 (0.01)	0.04 (0.01)	▼

A. Control & Aphasia Writer Groups' Comparative Statistics
Arrows Indicate direction of change from Control to Aphasia Writer Group

Writer's Language	Aphasia Doer Group	Aphasia Writer Group	Relative Change
% Function Words	0.43 (0.03)	0.40 (0.06)	▽
% Content Words	0.57 (0.03)	0.60 (0.06)	△
% Adjectives	0.11 (0.02)	0.13 (0.03)	△
% Verbs	0.13 (0.02)	0.11 (0.02)	▼
% Nouns	0.27 (0.03)	0.32 (0.05)	▲
% Adverbs	0.05 (0.01)	0.04 (0.01)	▼

B. Aphasia Doer & Aphasia Writer Groups' Comparative Statistics
Arrows Indicate change direction from Aphasia Doer to Aphasia Writer Group

Table 5. Comparative Statistics for Adaptation Theory
Values represent Mean (sd) Part of Speech usage
▽△ is significance $p \leq 0.05$, ▲▼ is significance $p \leq 0.01$ with GEE

not be the disorder itself, but the adaptation of the individual to the impairment. However, it would be very difficult to test this theory with individuals that have aphasia, because we cannot know what those individuals *intend* to say, or examine their manipulations of their language. Given that ACES logs the intended messages sent by participants in the *Aphasia Writer & Aphasia Doer Groups*, we can compare their syntactic changes to those individuals in the *Control Group*. The existing literature attributes many specific syntactic changes made by individuals with aphasia to Adaptation Theory. Based on the theory, because patients know that it's hard to talk they concentrate on producing shorter sentences [25] with high-information words, such as nouns [38], and dropping low information words, like function words, verbs, adverbs, etc [44, 27]. If Aphasia Adaptation Theory holds, we would expect to see such changes in our logs (in the intended messages, before distortion is applied), and would we could confirm Research Questions 2 and 3

Lastly, to address Question 4 we can examine the relationship between *Perceived Quality*, *Objective Quality* and linguistic features of the conversations (see following section on Language, Linguistics and Parts of Speech).

Statistical Tests

The scalar nature of our quantitative measures would suggest using a Two-Sample Wilcoxon Rank-Sum (Mann-Whitney) test, a more conservative metric than a T-Test as it makes no assumptions about the data distribution (normal or otherwise). However, a Rank-Sum does not account for the correlated nature of the data: each pair of participants had two conversations, and their interactions will clearly be correlated. Generalized Estimating Equations (GEE) [18] with a linear regression[11] were used to account for these correlations.

To test for correlations, we likewise used GEEs to account for the correlation in data. In this correlation analysis, we produce a coefficient that indicates the slope of the regression (and the direction of correlation). However, this coefficient is not the same as an R^2 or correlation coefficient, and cannot be treated as such. Therefore, the GEE tests' coefficients are not as easy to interpret as a Pearson's Coefficient. We therefore *also* ran a pairwise correlation coefficient using the Pearson's Correlation test. Sadly, Pearson's Correlation test does not account for the correlated nature of the data, so we cannot report the p-value, and the calculated correlation coefficient value should only be treated as an indication of the correlation magnitude (and not a hard-and-fast 'true' value). Pearson's Correlation coefficient values range from -1.0 (negatively correlated) to 1.0 (positively correlated).

[11] Linear regressions were used to test associations with scalar responses as outcomes.

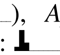
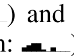
By having a fully counterbalanced study design, we minimize the impact of learning/order effects. For a robust analysis, we examine the impact of order/learning through GEE, comparing both conversation quality, and linguistic measures.

SUBJECTS

All 96 participants did not know each other, and were assigned randomly to a *Group*, *Task* and *Role*. Participants were recruited from a large state school, and the surrounding town. No particular prior knowledge about aphasia was required for inclusion. At the beginning of each session, we asked participants a series of demographic questions. Table 3 contains the mean and standard deviation for participants in each cohort. In addition, we tested to see if our cohorts' makeup was statistically different (Table 3). There were no statistical differences in age, gender, educational attainment, or prior knowledge of aphasia between any of the *Groups*.

RESULTS

All participants in our study were able to understand and perform the WIDI activity. No subject-pair was able to complete the activity (given that the thirty minute time provided was too short). Across all three *Groups*, the *Writers* communicated more (produced statistical more Lines and Words) as compared to the *Doers* ($p \leq 0.01$), see Table 4.

Overall, there was a variation in structure completeness (*Objective Quality*) within and between the *Control Group* (μ=31.52%, sd=13.98, histogram: ), *Aphasia Writer Group* (μ=17.23%, sd=6.12, histogram: ) and *Aphasia Doer Group* (μ=30.71%, sd=18.19, histogram: ). In statistical analysis, the difference between the *Aphasia Writer Group* and both the *Control Group* and *Aphasia Doer Group* was statistically significant ($p<0.01$). However, there was no statistical difference between the *Control Group* and the *Aphasia Doer Group* ($p=0.84$).

Quality Measure	Control Group	Aphasia Writer Group	Aphasia Doer Group
Objective (%)	31.52 (13.98)	17.23 (6.12)	30.71 (18.19)
Doer's ICSI	5.48 (1.25)	4.64 (1.31)	4.91 (1.25)
Writer's ICSI	5.15 (1.41)	4.53 (1.22)	4.99 (1.07)
Doer's ICR	3.74 (1.17)	4.79 (1.15)	4.13 (1.14)
Writer's ICR	3.79 (1.11)	4.87 (1.20)	4.26 (0.96)

***A. Summary Statistics for* Conversation Quality**

Values represents Mean (sd) of each measure of Conversation Quality

Quality Measure	CG to AWG	CG to ADG	ADG to AWG
Objective	▼	–	▼
Doer's ICSI	▽	–	–
Writer's ICSI	▽	–	–
Doer's ICR	▲	–	△
Writer's ICR	▲	–	△

***B. Comparative Statistics for* Conversation Quality**

Arrows Indicate change in direction

▽△ is significance $p \leq 0.05$, ▲▼ is significance $p \leq 0.01$ with GEE

Table 6. Comparative and Summary Statistics for *Conversation Quality*
ICR (lower is better), ICSI (higher is better)
Control Group (CG), *Aphasia Writer Group* (AWG), *Aphasia Doer Group* (ADG

	Objective	Writer ICSI	Writer ICR	Doer ICSI	Doer ICR
Objective	–	–	–	–	–
Writer ICSI	0.09	–	–	–	–
Writer ICR	-0.22	-0.60 ‡	–	–	–
Doer ICSI	0.14	0.22	-0.10	–	–
Doer ICR	-0.08	-0.25	0.11	-0.84 ‡	–

***A.* Control Group**

	Objective	Writer ICSI	Writer ICR	Doer ICSI	Doer ICR
Objective	–	–	–	–	–
Writer ICSI	0.17	–	–	–	–
Writer ICR	-0.36 †	-0.74‡	–	–	–
Doer ICSI	-0.02	0.10	-0.09	–	–
Doer ICR	-0.05	-0.13	0.15	-0.75 ‡	–

***B.* Aphasia Writer Group**

	Objective	Writer ICSI	Writer ICR	Doer ICSI	Doer ICR
Objective	–	–	–	–	–
Writer ICSI	-0.03	–	–	–	–
Writer ICR	0.30	-0.39 †	–	–	–
Doer ICSI	0.44†	-0.27	0.34 †	–	–
Doer ICR	-0.11	0.41†	-0.12	-0.72‡	–

***C.* Aphasia Doer Group**

Table 7. Correlations between *Conversation Quality*
Pearson's Correlation Coefficients for Indication of Magnitude
† is $p \leq 0.05$, ‡ is $p \leq 0.01$ with GEE
ICR (lower is better), ICSI (higher is better)

Aphasia Adaptation Theory

To examine the presence (or absence) of aphasia adaptation theory, we compared the intended language and POS usage for participants playing the *Aphasia Role* with participants in the *Control Group* (performing the same *Task*). When comparing the *Writer* in the *Control Group* (no distortion) to the *Writer* in the *Aphasia Writer Group* (with distortions) we see a large shift in language (Table 5A). While the number of messages (Table 4) statistically remained the same ($p > 0.05$), the number of words fell ($p \leq 0.01$). When examining what types of words were reduced or sacrificed, we observed that participants sacrificed function words, verbs, and adverbs, while increasing noun production.

When comparing the *Doer* in the *Control Group* to the *Doer* in the *Aphasia Doer Group* (with distortions) we see no statistically significant shift in language. Number of lines, words and POS distributions remained unchanged ($p > 0.05$). Subsequently, when we compared the *Writer* in the *Aphasia Writer Group* (with distortions) to the *Writer* in the *Aphasia Doer Group* (without distortions) we observed the same relative change as seen between the *Control Group* and the *Aphasia Writer Group* (Table 5B).

Conversation Quality

When we compare the conversation quality between the three *Groups* (Table 6B), we see that across all 5 measures of *Conversation Quality*, there is a statistical difference between the conversation quality in the *Control Group* and the *Aphasia Writer Group*. We also see several significant differences between the *Aphasia Writer Group* and *Aphasia Doer Group* (though not in the ICSI measures). There was no statistical difference between the conversation quality of the *Control Group* and the *Aphasia Doer Group*. Though not shown, there was no statistical difference in *Perceived Quality* between the *Writer* and *Doer* within each of the three cohorts[12].

[12]Lack of statistical significance does not equate to statistical similarity. Further analysis would be needed to prove the distributions are the same.

In theory, the 5 measures of *Conversation Quality* should all reflect the same thing: the quality of the conversation. To examine the inter-relationship between these five measures, we calculated correlations between each measure of *Conversation Quality* using a GEE. For those measure of *Conversation Quality* that were statistically correlated, we present Pearson's Correlations coefficients in a correlation matrix (Table 7).

Across all three *Groups*, the only two consistently significant correlations were between the measures of *Perceived Quality* (ICSI and ICR). Further, the correlation between *Writers*' ICSI and *Writers*' ICR, and between the *Doers*' ICSI and *Doers*' ICR are both negative with a high effect size. Having a negative correlation is expected, because while the higher the ICSI outcome is the better the conversation, the ICR uses lower scores to signify better conversations.There was *no* significant correlation found between *Objective Quality* and any of the *Perceived Conversation Quality* measures.

DISCUSSION

We now discuss the results within the context of our four Research Questions.

Question 1

Can ACES be used in a known experimental paradigm from Psychology, Communication or Speech and Hearing Science?

This experiment demonstrates that ACES can be integrated within a typical experimental design from the Psychology and Communication literature. Participants were able to understand the activity and their assigned *Tasks*. Participants also had varied objective performance, which was to be expected given the existing literature on this class of study design [7, 10, 4, 48]. Further, with the introduction of a distorted or impaired communication channel (*Aphasia Writer Group*), performance decreased[13] (also, as would be expected). We can therefore see that ACES can be integrated in traditional research setting.

Question 2

Can ACES' distortions impact conversation style in a known way?

Our analysis did find striking similarities between the intended messages sent by our participants with distortions, and behavior described in Aphasia Adaptation Theory. When *Writers* had their text distorted, they increased the percentage of words that were nouns while sacrificing verbs, adverbs, and function words. This suggest that *Writers* adopted a more curt method of communication, focusing on the object rather than the action. In other words, participants were adapting their language in a way that is evocative of how real non-fluent aphasics behave[14]. These findings show that ACES interactions produce known changes in conversation style. Perhaps more importantly, the parallels between our subjects adaptations and individuals with aphasia lends strong supporting evidence that the ACES emulator is suitable for further psychological testing.

Question 3

Can transcripts of intended *conversation style be useful?*

One of the challenges with Aphasia Adaptation Theory is that it concerns what an individual does (consciously or subconsciously) to adapt to their impairment *before* communicating and having their language be distorted. Using traditional transcripts containing only what an individual "said" would not be able to differentiate what participants changed in their language, versus which changes were generated due to distortions. Therefore our findings and their support for Aphasia Adaptation Theory could only have been conducted and uncovered with transcripts of *intended* messages. ACES, coupled with these findings, opens the door to other questions about user adaptation and changes to linguistic distortions (from aphasia or other language disorders).

Question 4

Can ACES be used to uncover new understandings of interpersonal interaction?

Our analysis of *Conversation Quality* uncovered that there is no relationship between *Objective Quality* and *Perceived Quality* (Table 7), however our two measures of *Perceived Quality* were highly correlated. Further, for both *Objective Quality* and *Perceived Quality*, Aphasic Distortions made *Conversation Quality* worse[15] (see Table 6).

[13] Interestingly, we did not see a difference between the *Control Group* and the *Aphasia Doer Group*. This observation opens up a new question; is the Shannon Information [41] sent from the *Doer* to the *Writer* not as important as the information conveyed from the *Writer* to the *Doer* (so therefore any distortion does not matter or has minimal impact)? This may be supported by the statistically lower number of Lines and Words being sent from the *Doer* to the *Writer*. Another explanation is that the "amount" of information that needs to get across from the *Doer* to the *Writer* smaller/easier to understand (e.g. short pragmatics or continuers like "ok", or "got it"). It is outside of the scope of this paper to explicitly measure Shannon Information in each message, so we therefore cannot assess if the Writer has more (in quantity) information, or more complex information to convey. However we see this as a promising future research question.

[14] While this one study isn't definitive "proof" of Adaptation Theory, it does strengthen the arguments for the theory of Aphasic adaptation.

[15] While we do not see statistically significant differences in ICSI scores between the Aphasic Writer and Aphasic Doer Cohorts, this may be due to statistical power.

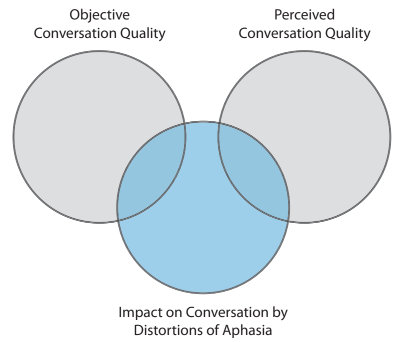


Figure 3. Venn Diagram of Conversation Quality and Distortion Impact

Thus, when considering these two findings in concert, we conclude that distortions of aphasia (by ACES) impact *Conversation Quality* in two distinct, and independent ways, as illustrated by Figure 3, by affecting both the product (or goal) of the conversation (*Objective Quality*), as well as the interpersonal interaction(*Perceived Quality*). These new understandings about interpersonal communication are clearly grounded in the quality of the measures used. While the Objective Conversation Quality is, by definition, objectively observable, perceived quality is not. However, by having two measures that have a high effect size correlation (*correlation coefficient* > 0.6) and are highly significant ($p < 0.05$), we can infer that they are valid and verifiable measures. This supports the conclusion that aphasia (or, to be precise, ACES' distortions which simulate aphasia) impacts both how we feel about the interaction as well as the conversation output quality/goal. It is important to note that gathering this type of user feedback (perceived conversation quality metrics via questionnaires) would be extremely difficult with individuals who have aphasia. These findings, and the unique ability to gather them from "typical" individuals using ACES highlights how ACES is a rich platform for uncovering new phenomena.

LIMITATIONS

Whenever there is interaction that is not face-to-face, many backchannels (e.g. gestures or intonation) are lost. This is a limitation (or control) of many study designs and many forms of CMC (e.g. IM or Email). While this limitation is important to note, Novice-Novice Designs are well accepted study designs to investigate interpersonal communication. Further, individuals with aphasia actively use many CMC applications [2], and their linguistic deficits in writing are generally consistent with those of the person's spoken language [6].

IMPLICATIONS & FUTURE WORK

Beyond explicitly answering our four research questions, we see multiple implications of the findings from our study for both technologists and non-technologists.

CMC Emulation Systems

While prior studies on ACES have shown the realism of its distortions[17] and its impact on improving user empathy[16], this is the first study to show that CMC emulation systems can impact *how* users communicate and the quality of their conversations. Moreover, these changes appear to be similar to those observed in individuals with aphasia. To our knowledge, this is the first such system demonstration of its kind (for any disorder).

These findings (coupled with the prior work on ACES) suggest the need to create system that emulate other language[16] or perception disorders. Beyond explicit disorders, technologists could investigate systems that build empathy and study communication in other contexts (e.g. English as a second language, or exposure to cultural differences). ACES research shows that CMC systems can do more than just facilitate communication. They can build empathy[16] and explicitly study communication when one interlocutor's speech is impaired.

Aphasia & Communication

We believe a rich area of future exploration is using ACES-like systems to test theories of communication, and exploring some of the current gaps in understanding how people with disorders communicate. Given the strong initial results and rich data set collected during this experiment, there are many more exciting questions that may be asked of this data. What linguistic changes (be they semantic, syntactic or structural) impact *Conversation Quality*? What types of questions asked by participants provoke the most useful responses? As alluded to above, what is the difference in Shannon Information between the two *Tasks*? owever these and other questions fall well outside the scope of this paper and this publication venue, given its technical grounding. Ideally, these questions would be addressed by Psychologists and published in Psychology research journals. Given how many provocative questions have emerged from this study, we are highly optimistic that our work here has demonstrated that this new technology has a wide variety of future applications.

Based on the new findings related to the relationship between types of *Conversation Quality*, we hypothesize that there are two aspects of conversation that can be targeted for "improvement." First, we can attempt to mediate the information conveyed to generate a more productive output (*Objective Quality*). We can also attempt to change how people feel and react to challenging interactions (*Perceived Quality*). While both goals are important, it is the task of improving perceived conversation that **must** be targeted, since improving only the objective output itself may not improve the perceived quality (since the interaction will still be challenging). Given previous work on ACES improving empathy[16], understanding and patience for conversation partners, we hypothesize that by providing users (and potential conversation partners, clinicians, and/or doctors) with the experience of communicating with aphasia (through ACES), we can effectively improve perceived conversation quality (this would need to be verified in a future experiment.)

These future directions of ACES beg the questions " how far can this system be pushed?" and "where does the line between emulation and broad implications for treatment lie?" Given the parallels between adaptation theory shown in this paper, and the turing test validation shown in [17], we are optimistic. However, exploring this relationship for aphasia, or other disorder emulation, is an important direction for future exploration.

[16]ACES-like systems could be tailored to their users by analyzing their linguistic patterns and word usage (e.g. by analyzing emails), thus personalizing the distortions.

Research with Individuals with Impairments

Finally, it should be noted that to conduct this study we recruited 96 participants, none of whom had aphasia. Subsequently, this study required no special accommodations or extra time to accommodate a language impairment. All instructions *and* questions were given once, and participants' responses to questions were easily understood. This is a striking contrast to the additional complexity of running a study with participants that have an impairment [21]. As a result, more participants were able to be included (increasing N), over a shorter time-frame (3 months), with less financial cost. Further, we were able to control the exact manifestation of aphasia, which would not be possible in a traditional setting. These are benefits for technologists and non-technologists alike. For example, imagine that a communication researcher theorizes that, if an individual with aphasia alters their language in some particular way or uses an new UI, they will have an improved experience with their conversation partner. ACES could be used to test this theory, examining such an intervention's impact on the language and conversation quality. Then, based on these initial findings, the researcher could then run a better designed large-N study with individuals who have aphasia (or re-run the design to adjust for the findings).

CONCLUSION

This paper illustrates how a CMC impairment emulation system (ACES) can be integrated within a traditional Psychology and Communication research experiment. Not only did our participants understand and complete the activity, but we appear to provide new evidence towards validating an existing theory in the Psychology literature on aphasia. Moreover, this supporting evidence would be nearly impossible to construe without a system like ACES. Further, our ACES study was able to uncover new understandings about interpersonal communication and its impact on *Conversation Quality*. By situating ACES within this "traditional" experimental context, we show that this new technology can provide many unique and important benefits to researchers (control, customization, low cost, user feedback, and pre-impairment intent) that are not present when working directly with individuals that have an impairment. We believe that our novel and important findings for researchers in Psychology, Speech and Hearing, and Communication research. Further, the findings highlight the need for and potential applications of ACES-like CMC systems as a new direction for HCI research with direct and tangible real-world impact.

ACKNOWLEDGMENTS

We would like to thank our participants. We would also like to thank Gary Dell and Sarah Brown-Schmidt whose advice throughout this experimental design, execution and analysis proved invaluable.

REFERENCES

1. Al Mahmud, A., Gerits, R., and Martens, J. Xtag: designing an experience capturing and sharing tool for persons with aphasia. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, ACM (2010), 325–334.
2. Al Mahmud, A., and Martens, J. Understanding email communication of persons with aphasia. In *Extended abstracts on Human factors in computing systems*, ACM (2011), 1195–1200.
3. Allen, M., McGrenere, J., and Purves, B. The design and field evaluation of phototalk: a digital image communication application for people. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, ACM (2007), 187–194.
4. Bard, E., Anderson, A., Chen, Y., Nicholson, H., Havard, C., and Dalzel-Job, S. Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language 57*, 4 (2007), 616–641.
5. Benson, D. *Aphasia, Alexia and Agraphia: Clinical Neurology and Neurosurgery Monographs*. Churchill Livingstone, New York, 1979.
6. Benson, D. *The neurology of thinking*. Oxford University Press, USA, 1994.
7. Beun, R., and Cremers, A. Object reference in a shared domain of conversation. *Pragmatics &# 38; Cognition, 6 1*, 2 (1998), 121–152.
8. Bos, N., Olson, J., Gergle, D., Olson, G., and Wright, Z. Effects of four computer-mediated communications channels on trust development. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, ACM (2002), 135–140.
9. Boyd-Graber, J., Nikolova, S., Moffatt, K., Kin, K., Lee, J., Mackey, L., Tremaine, M., and Klawe, M. Participatory design with proxies: developing a desktop-pda system to support people with aphasia. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM (2006), 151–160.
10. Brown-Schmidt, S., and Tanenhaus, M. Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive science 32*, 4 (2008), 643–684.
11. Colby, K., Weber, S., and Hilf, F. Artificial paranoia. *Artificial Intelligence 2*, 1 (1971), 1–25.
12. Cushman, W., and Rosenberg, D. Human factors in product design. *Advances in human factors/ergonomics 14* (1991).
13. Duck, S., Rutt, D., HOY, M., and STREJC, H. Some evident truths about conversations in everyday relationships all communications are not created equal. *Human communication research 18*, 2 (1991), 228–267.
14. Giles, H. *A study of speech patterns in social interaction: Accent evaluation and accent change*. PhD thesis, University of Bristol, Bristol, UK, 1971.
15. Goodglass, H., Goodglass, and Kaplan. *Boston Diagnostic Aphasia Examination: Stimulus Cards–Short Form*. Lippincott Williams & Wilkins, 2001.
16. Hailpern, J., Danilevsky, M., Harris, A., Karahalios, K., Dell, G., and Hengst, J. Aces: Promoting empathy towards aphasia through language distortion emulation software. In *Proceedings of the ACM's SIG CHI Conference 2011 Conference.*, CHI 2011, ACM (Vancouver, BC Canada, 2011).
17. Hailpern, J., Danilevsky, M., and Karahalios, K. Aces: Aphasia emulation, realism, and the turing test. In *Proceedings of the ACM SIGACCESS- ASSETS 2011 Conference.*, ASSETS 2011, ACM (Dundee, Scotland, 2011).
18. Hardin, J., and Hilbe, J. *Generalized estimating equations*. Chapman and Hall/CRC, New York, 2003.
19. Hartsuiker, R., and Kolk, H. Syntactic facilitation in agrammatic sentence production. *Brain and Language 62*, 2 (1998), 221–254.
20. Hecht, M. The conceptualization and measurement of interpersonal communication satisfaction. *Human Communication Research 4*, 3 (1978), 253–264.
21. Henry, S. *Just ask: integrating accessibility throughout design*. Lulu. com, 2007.
22. Jurafsky, D., Martin, J., Kehler, A., Vander Linden, K., and Ward, N. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*, vol. 163. MIT Press, 2000.
23. Kirvesoja, H. *Experimental ergonomic evaluation with user trials: EEE product development procedures*. Oulun yliopisto, 2001.
24. Kolk, H. A time-based approach to agrammatic production. *Brain and Language 50*, 3 (1995), 282–303.
25. Kolk, H., and Heeschen, C. Adaptation symptoms and impairment symptoms in broca's aphasia. *Aphasiology 4*, 3 (1990), 221–231.
26. Kolk, H., and Heeschen, G. The malleability of agrammatic symptoms: A reply to hesketh and bishop. *Aphasiology 10*, 1 (1996), 81–96.
27. Kolk, H., and Van Grunsven, M. Agrammatism as a variable phenomenon. *Cognitive Neuropsychology* (1985).
28. Leshed, G. *Automated language-based feedback for teamwork behaviors*. PhD thesis, Cornell University, 2009.
29. Menn, L., Kamio, A., Hayashi, M., Fujita, I., Sasanuma, S., and Boles, L. The role of empathy in sentence production: A functional analysis of aphasic and normal elicited narratives in Japanese and English. *Function and Structure* (1998), 317–356.

30. Menn, L., Reilly, K., Hayashi, M., Kamio, A., Fujita, I., and Sasanuma, S. The interaction of preserved pragmatics and impaired syntax in Japanese and English aphasic speech. *Brain and language 61*, 2 (1998), 183–225.

31. Moffatt, K., McGrenere, J., Purves, B., and Klawe, M. The participatory design of a sound and image enhanced daily planner for people with aphasia. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2004), 407–414.

32. Nguyen, D., and Rosé, C. Language use as a reflection of socialization in online communities. *ACL HLT 2011* (2011), 76.

33. Niederhoffer, K. G., and Pennebaker, J. W. Linguistic style matching in social interaction. *Journal of Language and Social Psychology 21*, 4 (2002), 337–360.

34. Piper, A., Weibel, N., and Hollan, J. Introducing multimodal paper-digital interfaces for speech-language therapy. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, ACM (2010), 203–210.

35. Ruiter, M., Kolk, H., and Rietveld, T. Speaking in ellipses: The effect of a compensatory style of speech on functional communication in chronic agrammatism. *Neuropsychological rehabilitation 20*, 3 (2010), 423–458.

36. Ruiter, M. B. *Speaking in ellipses: The effect of a compensatory style of speech on functional communication in chronic agrammatism*. PhD thesis, Radboud University Nijmegen, Nijmegen, Netherlands, 2008.

37. Saffran, E., Berndt, R., and Schwartz, M. The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language 37*, 3 (1989), 440–479.

38. Salis, C., and Edwards, S. Treatment of written verb and written sentence production in an individual with aphasia: A clinical study. *Aphasiology 24*, 9 (2010), 1051–1063.

39. Scissors, L., Gill, A., Geraghty, K., and Gergle, D. In cmc we trust: the role of similarity. In *Proceedings of the 27th international conference on Human factors in computing systems*, ACM (2009), 527–536.

40. Scissors, L., Gill, A., and Gergle, D. Linguistic mimicry and trust in text-based cmc. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM (2008), 277–280.

41. Shannon, C., and Weaver, W. *The mathematical theory of communication*, vol. 19. University of Illinois Press Urbana, 1962.

42. Shewan, C., and Kertesz, A. Reliability and validity characteristics of the western aphasia battery (wab). *Journal of Speech and Hearing Disorders 45*, 3 (1980), 308.

43. Takagi, H., Asakawa, C., Fukuda, K., and Maeda, J. Accessibility designer: visualizing usability for the blind. In *ACM SIGACCESS Accessibility and Computing*, no. 77-78, ACM (2004), 177–184.

44. Thompson, C., Ballard, K., Tait, M., Weintraub, S., and Mesulam, M. Patterns of language decline in non-fluent primary progressive aphasia. *Aphasiology 11*, 4-5 (1997), 297–321.

45. Toma, C. Perceptions of trustworthiness online: The role of visual and textual information. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM (2010), 13–22.

46. Wang, H., and Fussell, S. Groups in groups: Conversational similarity in online multicultural multiparty brainstorming. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM (2010), 351–360.

47. Weizenbaum, J. Contextual understanding by computers. *Communications of the ACM 10*, 8 (1967), 480.

48. Wilkes-Gibbs, D., and Clark, H. Coordinating beliefs in conversation. *Journal of Memory and Language 31*, 2 (1992), 183–194.

49. Wilson, J., Straus, S., and McEvily, B. All in due time: The development of trust in computer-mediated and face-to-face teams. *Organizational Behavior and Human Decision Processes 99*, 1 (2006), 16–33.