# Journal of Assistive Technologies

Co-designing a speech interface for people with dysarthria

Foad Hamidi Melanie Baljko Connie Ecomomopoulos Nigel J. Livingston Leonhard G. Spalteholz

# Peer-reviewed paper

# Co-designing a speech interface for people with dysarthria

Foad Hamidi, Melanie Baljko, Connie Ecomomopoulos, Nigel J. Livingston and Leonhard G. Spalteholz

Foad Hamidi and Dr Melanie Baljko, both based at York University, Toronto, Canada. Connie Ecomomopoulos is Independent Researcher, at Toronto, Canada. Dr Nigel J. Livingston and Leonhard G. Spalteholz, both based at CanAssist, University of Victoria, Victoria, Canada.

**Abstract**

**Purpose** – *The purpose of this paper is to describe the development and evaluation of CanSpeak which is an open-source speech interface for users with dysarthria of speech. The interface can be customized by each user to map a small number of words they can speak clearly to commands in the computer system, thereby adding a new modality to their interaction.*

**Design/methodology/approach** – *The interface was developed in two phases: in the first phase, the authors used participatory design to engage the users and their community in the customization of the system, and in the second phase, we used a more focussed co-design methodology during which a user of the system became a co-designer by directly making new design decisions about the system.*

**Findings** – *The study showed that it is important to include assistive technology users and their community in the design and customization of technology. Participation led to increased engagement, adoption and also provided new ideas that were rooted in the experience of the user.*

**Originality/value** – *The co-design phase of the project provided an opportunity for the researchers to work closely with a user of their system and include her in design decisions. The study showed that by employing co-design new insights into the design domain can be revealed and incorporated into the design that might not be revealed otherwise.*

**Keywords** *Co-design, Multimodality, Assistive technology, Dysarthria of speech, Participatory design, Speech recognition*

**Paper type** *Research paper*

## 1. Introduction

Participatory design (PD) (Greenbaum and Kyng, 1991; Schuler and Namioka, 1993) and specifically co-design (Sanders and Stappers, 2008), is a design methodology that emphasizes the incorporation of user domain knowledge and recognizes the importance of collaborating and co-creating with the user. An essential technique of PD is to actively engage the user of the technology in its design. Originally, PD was used to develop technologies for the workplace; however, it became apparent that it also offers great value to many other areas of design and development including assistive technology (AT) (Kensing and Blomberg, 1998).

CanSpeak is a software application that improves the accessibility of software systems that require the use of computer mouse and keyboard. It functions by providing speech as an alternative mode for input actions, and is designed for use by individuals whose speech is affected by dysarthria. CanSpeak is highly customizable and its functioning can be tailored to the user's particular speech preferences and characteristics (which for users with dysarthria can vary to a large degree from user to user). CanSpeak uses a set of "keywords," each of which is associated with a command in the software application that the user seeks to use

Received 26 October 2014
Revised 5 January 2015
Accepted 7 January 2015

The authors would like to thank all the volunteer participants, as well as Sandra Salamon and Dr Torsten Felzer.

(Hamidi *et al.*, 2010). The user's input speech is mapped to keywords using automatic speech recognition (ASR). The user and/or their caregivers can customize the keywords and commands based on the user's preferences and needs. In this paper, we describe how we used PD to come up with the initial design of CanSpeak and how we used it to further refine it.

The objective of this paper is as follows: to discuss the use of PD in the design of ATs using a concrete example, to describe the knowledge outputs generated from co-designing a variant of CanSpeak in a series of case studies, and to advocate for the broader uptake of PD in the design of AT. To support this process, we discuss our reflections on the application of PD and co-design to AT.

The focus of this paper is on the inclusive design methodologies used. From the beginning, we worked closely with several users with disabilities and members of their community (i.e. teachers, parents and caregivers) who provided us with suggestions and usage data that we incorporated into the design of the interface. Later on, we modified our method to include a user with disability, and a co-author of this paper, as a member of our design team.

## 2. Background

The rationale for employing PD as a design methodology for AT is compelling. PD (Greenbaum and Kyng, 1991; Schuler and Namioka, 1993) and co-design (Sanders and Stappers, 2008) are design methodologies that emphasize the incorporation of user domain knowledge and recognize the importance of collaborating and co-creating with the user. An essential technique of PD is to actively engage the user of the technology in its design. Originally, PD was used to develop technologies for the workplace; however, it has become apparent that it also offers great value to many other areas of design and development (Kensing and Blomberg, 1998). As we will argue below, PD is appropriate for designing AT for a number of reasons.

In speaking about design methodologies and design domains, it is important to be clear about roles and context. Computing-related knowledge outputs, as generated by academic activity, are often envisioned as finding their application through technology transfer processes (i.e. the various processes by which scientific and technological developments are made accessible to a wider range of users, who can then further develop and exploit the technology into new products, processes, applications, materials or services, oftentimes through commercialization). Certainly, the AT sector has a large number of commercial entities which design, develop, manufacture and sell AT digital products. In market climates that operate under accessibility provisions that are mandated by governmental legislation, the seller-buyer relationship can be more complex than the typical producer-consumer relationship. For instance, the AT user may not be the buyer of the AT product (e.g. a school board or governmental agency may be the purchaser and the administrator of a program that allocates AT products to the program clients). Commercial AT solutions tend to be deployed in formal contexts (e.g. through formal assessment in technology access clinics, by clinicians or other professionals, such as speech-language pathologists). We certainly feel that PD has an important role to play in the design of commercialized AT.

However, and just as importantly, the AT sector is also deployed through non-commercial channels. Motivated individuals, such as parents or teachers, retired engineers (e.g. working for non-profit entities, such as the Tetra Society) and academic researchers, animate these flows. The AT solutions may include the development, installation and/or configuration of open-source AT software solutions, or the design and/or hacking of digital devices. These AT solutions tend to be highly specific to a single or a few users.

In academia, researcher in computing-related fields with focus on AT, oftentimes develop AT prototypes, which get taken up as AT solutions by users. In this context, researchers are faced with the task of designing digital technologies for users whose abilities, desires and needs differ from themselves greatly. This may be the case even for those researchers who have disabilities themselves and are themselves users of AT. Each person's use scenario may be unique and it is important to develop and examine methods that allow for the incorporation of different perspectives and experiences into design. Thus, it is essential for technology designers to work closely with representatives from the population they are designing for, so that they can have a better understanding of the potentials and perils of their creation when used by real people. Additionally, the importance of accommodating designers and co-designers who themselves have disabilities and thus, "insider knowledge," in the design of AT is becoming increasingly apparent (Rogers and Marsden, 2013).

In addition, it is important to engage people who have disabilities but who might not have ready access to fabrication and design methods, since this will support the democratization of design and will challenge the traditional relationship between technology providers and its users, bridging the gap between these two populations (Hurst and Tobias, 2011).

A further potential benefit has been articulated in terms of technology uptake (vs abandonment). The high rate of AT abandonment is well studied (Phillips and Zhao, 1993). The application of PD generates the possibility of technologies that are created by people with disabilities for people with disabilities. This might contribute to a stronger sense of ownership of the technology.

A variety of AT projects have previously used PD. Wu *et al.* (2004) used PD with a group of six individuals with memory loss (amnesia) to develop a computational tool to help with the problem of disorientation. The researchers found that working with this target population as design partners allowed them to gain a better understanding of their living conditions and to gain access to the personal expertise of potential users through "mutual learning." The researchers used paper tools, such as meeting agendas and drawn use-case scenarios, to aid with the design. The research resulted in the development of a memory aid software application to be used on a personal digital assistant (PDA).

In another study, PD was used with people with aphasia in the development of an enhanced with sound and images (ESI) planner for PDAs. Users provided feedback on different stages of prototyping, from brainstorming to low-fidelity paper prototyping to high-fidelity software prototyping (Moffatt *et al.*, 2004). Initial brainstorming and prototyping were conducted with a participant with amnesia who passed away before the end of the study. Subsequently, three other people with amnesia collaborated with the researchers in later stages of prototype development. The ESI planner was evaluated against a not enhanced with sound and images (NESI) planner, with eight users participating as evaluators (users with aphasia). Although the researchers observed that the performance of the participants was very diverse, the results provided useful knowledge about design tradeoffs (e.g. the NESI planner could be operated more quickly, but the ESI planner could be used more accurately).

Galliers *et al.* (2012) also engaged potential users with aphasia in a series of workshops in which five participants used gestures (rather than spoken or written language) to express ideas about software and paper game prototypes and to provide feedback about them. The researchers found that this method empowered the participants and also challenged the notion of researcher as "fixer." This study emphasized that communication is a challenge when using PD with users (or potential users) of AT; communication is a key process, and to support this, it is important to devise appropriate communication modes and tools (i.e. accessible expressive tools and language).

In a study where a communication application for individuals with cognitive disabilities was to be developed, Dawe used an early prototype as a "technology probe" (Dawe, 2007). The concrete instantiation of a prototype can help overcome difficulties that potential users may have in expressing their ideas verbally and in engaging in what otherwise could be abstract or hypothetical discussions (Cohen and Graupe, 1980). The probe was developed on the basis of qualitative data (interviews with 20 families of children with cognitive disabilities). It was to stimulate communication in the form of suggestions and ideas from two individuals with cognitive disabilities and their families. Dawe argued that the use of a technology probe is an effective means of engagement with individuals with significant cognitive disabilities because it supports the necessary communication processes. The use of a probe allowed the participants to express ideas through actions in relation to the technology. Additionally, having a concrete object (even in early versions) that could be changed and modified in response to user suggestions and requests offered the potential for the development of a sense of ownership and empowerment in the user and the idea that his or her input was reflected in a concrete manner in the technology being designed.

Anthony *et al.* (2012) conducted a one-day PD workshop with 12 postsecondary students with various learning disabilities. In the workshop, students interacted with two prototypes in teams and provided feedback and design ideas to the researchers. Post-workshop evaluation surveys showed that the students found the workshop engaging and relevant. They felt empowered by being included in the design process and motivated to engage in discussions and hands-on activity. The researchers identified important factors for the success of the approach to be the consideration of communicative differences and the focus on relating the projects to the participants' personal lives.

In a study with 18 older adult participants, Davidson and Jensen (2013) involved potential users in the design of a smartphone app to monitor health. Participants were divided into five groups, each working on one of five possible applications. Analysis of the resulting designs showed that participants identified several novel health metrics (i.e. metrics that were not already included in existing apps). Also, the breadth of design suggestions indicated that the participants had a more holistic view of health that allowed them unique insights into design; this holistic view was not completely in convergence with that of the technology designers.

The above projects make two points apparent: first, including users in the design process can uncover valuable domain expertise that can be incorporated into design (Visser *et al.*, 2005) and; second, a challenge for applying PD to the development of AT is to come up with appropriate methods that allow for the expression and integration of this expertise (Sanders and Stappers, 2008). These communication tools can be alternatives to formal questionnaires and interviews and may involve dialogue around lo-fi and/or high-fi prototypes.

The high degree of both inter- and intra-variability in the speech of users with dysarthria of speech makes developing ASR applications that aim to recognize this kind of speech accurately challenging (Rosen and Yampolsky, 2000). Despite these challenges ASR has been used for more than three decades to provide environmental and computer control for people with dysarthria (Clark and Roemer, 1977; Cohen and Graupe, 1980). While in this paper, we focus on design methodology rather than the ASR system, it should be noted that similar techniques using small-vocabulary ASR systems have been developed for dysarthric speech in the past (Kewley-Port *et al.*, 1991; Rosengren *et al.*, 1995). For example, in the STARDUST project, a speaker-dependent, small-vocabulary interface was developed to provide environmental control for user with dysarthria of speech (Hawley *et al.*, 2003). The system was tested with five individuals with speech dysarthria and improvements in accuracy rates were observed for three users. However, all participants were undergoing speech therapy intervention at the same time as using the system, so it is difficult to say if the improvements in performance were due to the ASR training or speech intervention. Additionally, it was observed that the physical and health conditions of users made extensive training difficult for them (Hawley *et al.*, 2003).

## 3. CanSpeak

CanSpeak is a software application that improves the accessibility of software applications that require the use of the computer mouse and/or keyboard (Hamidi *et al.*, 2010). It functions by providing an alternative mode for input actions via the mode of speech, and is designed for use by individuals with dysarthria of speech. It is used as an intermediate interface between the user and any application that uses discreet input (e.g. from the keyboard, switches or mouse clicks). CanSpeak makes use of a set of "keywords," each of which is associated with a command in the software application that the user seeks to use). Any of these keywords can be triggered using ASR.

The commands and keywords can be customized based on user preferences and needs (which can vary to a large degree from user to user). The user and/or his or her community (i.e. caregivers, teachers and parents) can change the keywords to ones that are easy to say clearly for them. The ASR module in the system does not require any training on the user's input speech, but the system does require configuration to ensure a good fit between the characteristics of the user's speech and the set of keywords. The keyword set consists of a small number of words specifically tailored to each user. Each of the keywords in the set is associated with an input action (e.g. a key event that corresponds to typing a letter of the alphabet or a digit, or the invocation of a system command). Each time a keyword is recognized, the system sends the corresponding input action to the application that is using the interface. The application then interprets the input action according to the interaction context. The number of keywords depends on the application and needs of the user. In the first iteration, the user desired text entry functionality; so 47 keywords were used. However, in other use scenarios, we have used as little as three keywords.

CanSpeak's user interface consists of a window that displays the keyword set and their associated mappings (e.g. to the letters of the alphabet or system commands). The CanSpeak window is positioned beside the user interface of the application with which CanSpeak is communicating the mapped commands. For example, if CanSpeak is used to access the internet, its window is displayed beside the browser interface. To save screen space, the CanSpeak interface can be modified such that only some keywords are displayed. The interface is shown on the left hand side of Figure 1.

CanSpeak is written in Java and uses the open-source Sphinx-4 speech recognition engine (Walker *et al.*, 2004). It uses modular design, which makes it easy to customize. Additionally, the design makes it easy to integrate it into different applications. For example, we have combined it with KeySurf, a keyboard-based WWW navigation application (Spalteholz *et al.*, 2008). The resulting application, WebSpeak, is a multimodal internet navigation interface that combines voice and keyboard input. The importance of this type of modular design is that it allows user-tailored modes of inputs (e.g. speech for one user but mechanical switches for another); this approach is particularly emphasized in AT where the user population has varied needs and abilities (Grammenos *et al.*, 2006; Kawai et al., 1996). In terms of functionality, the interface simply replaces keyboard keystroke or mouse click input with input from the speech module. Figure 1 shows the interface of CanSpeak.

CanSpeak's initial design and subsequent refinements were developed using PD. Variants of CanSpeak that were each tailored to different user scenarios were developed over time. CanSpeak was originally designed as an interface to aid with web navigation, but was subsequently generalized into a customizable intermediary module that can be used with different software applications.

The idea for the interface was inspired by the interaction needs of a colleague who faced accessibility barriers to conventional input devices such as the keyboard and mouse (due to motor impairment arising from cerebral palsy) and barriers to the use of speech input

**Figure 1** CanSpeak interface (left) beside a web browser window

(due to dysarthria). This user circumvented the barriers through the use of keyboard input using the mouth. However, he became motivated to try to combine speech and typing, since typing was highly fatiguing. Given these characteristics, we decided to develop a small-vocabulary speech input module to be used in combination with other input modes (e.g. keyboard, touchscreen, etc.) to form a multimodal interface and help decrease the fatigue caused by using one mode of interaction (e.g. typing) by making speech available as an alternative input mode. Thus, from the start, during the PD stages of initial exploration and discovery process (Spinuzzi, 2005), the project involved a close working relationship with a participant who provided valuable feedback on the design of the interface.

ASR accuracy improves when the number of possible input spoken keywords is limited. Therefore, the interface would ideally be used in combination with other input modes. However, it can also be used on its own. Initially, the target use scenario was accessing the WWW, and so the first use of CanSpeak was as part of the WebSpeak interface. Later on, we realized that this interface can be used in other contexts and modified it for different users.

CanSpeak's design is based on two main ideas that we discuss next: first, input restriction, the simplification and limiting of the speech recognition task, and; second, customization for each user.

### *3.1 Input restriction*

Previous research has recommended that automatic recognition tasks be kept simple (Franco *et al.*, 2000). Although the speech engine used, Sphinx-4, is capable of large vocabulary continuous speech recognition, we have configured it for recognizing a small-vocabulary set: this modification greatly simplifies the task of the engine and improves performance. We have also limited the speech input to isolated words and phrases of less than three words, on the basis of evidence that people with severe dysarthria perform better with this type of speech than with longer utterances (Rosen and Yampolsky, 2000). Another method used to improve recognition rates in a later version of the system was to cluster keywords into small groups or categories of minimal pairs (i.e. words that differ from each other in only one sound or phoneme). This method allows for the speech recognition engine to recognize a group of words rather than a single word. Both input restriction and categorization simplify the recognition task; the first method limits the scope of recognition and the second allows for less needed precision on what is recognized.

### *3.2 Customization*

There are two ways to customize CanSpeak. The first is through keyword selection, which is the process of selecting words that are easy to say for each user. In the experiments described in the next section, we worked with the user and his or her caregivers, teachers and parents to come up with keyword alternatives but the interface is designed such that the keyword selection can be done without the help of researchers. Thus, a configuration interface is provided for the modification of keywords that are associated with commands. Specific keyword-command mappings are stored in each user's individual profile. Our approach was to use a modified version of the NATO phonetic alphabet as the default starting point and to make substitutions and modifications as desired. The reason for choosing the NATO phonetic alphabet as default is that these keywords were originally chosen to be distinguishable from each other in noisy conditions and thus are easier for the engine to recognize. However, some of the keywords were not common familiar words (e.g. "Zulu") and were replaced (e.g. "Zebra") in the default keyword set.

We believe the best method for finding a good set is to take advantage of the user's own knowledge as well as feedback from people familiar with their speech (i.e. his or her speech therapists, parents, teachers or caregivers) about which words are easier for him or her to pronounce and for other listeners to understand. Keyword selection presents several challenges. First, the keyword-command mapping should be intuitive and easy to recall. Second, the keywords themselves should be easily and clearly articulable by the user. Third, the differences among the keywords (in terms of acoustic confusability) should be above the threshold of ASR misrecognition. Thus, the selection of a suitable keyword set is not trivial[1].

The second way to customize CanSpeak is through mapping design, which involves decisions as to what commands or input each of the keywords is associated with. Working closely with users of the system, many different possibilities were identified and incorporated into the design, starting from simple alphanumeric mappings to mouse clicks and finally to key combinations (e.g. Ctrl + C). Currently, researchers conduct mapping design, as it involves communication between CanSpeak and other applications, which requires programming skill. In the future, we plan to provide ways for users and their community to change mapping design themselves.

## 4. Methodology: co-designing CanSpeak

### *4.1 Pilot study*

We conducted a pilot study to assess the performance of the ASR module using the default keyword list (i.e. consisting of the NATO phonetic alphabet). These initial tests were done with users with clear speech. Two of the users had mild accents. In total, 15 participants with non-dysarthric speech participated using the modified keyword set. During this phase, several problematic words were identified and a modified set of NATO phonetic alphabet keywords was developed. The original recognition rate of 92.7 percent improved to 95.8 percent, once five out of 47 words were modified.

The purpose of the pilot study was to make sure the initial system worked adequately and to remove any software implementation and configuration errors before starting user studies with participants with disabilities. The pilot study was conducted in our research lab and we recruited our colleagues and students as participants in this phase. Conducting a pilot study before starting user studies is recommended to avoid having to discard data because of unnecessary errors in technology setup (van Teijlingen and Hundley, 2002).

### *4.2 Phase 1*

During this phase we conducted a user study with four participants with cerebral palsy and dysarthria of speech. Three of the participants are female and one male. Their ages ranged from 19 to 35. At the time of the study, all participants were students enrolled in an adult literacy-learning program.

Participants 1, 3 and 4 use computers on a regular basis. Participants 1 and 3 usually use a regular keyboard to type, but find typing to be fatiguing. Participant 4 uses his upper torso (head and lips movement) to type on a regular keyboard fitted with a waterproof shield. All three of these participants expressed interest in using their speech as an alternative input mode when tired from using their regular input routine. They were also interested in combining speech input with word prediction for writing. Participant 2 was not a regular computer user; she did not have sufficient motor control to use a standard keyboard or mouse and did not use AT to access the computer.

The experiment consisted of the following steps: first, an initial assessment session using the default keyword set developed in the pilot study was conducted. This was followed by an interface customization session, during which a new keyword set was developed using the participants' suggestions. In some cases, suggestions from the participant's teachers, caregivers and parents were incorporated. Finally, another assessment session using the new keyword set was conducted. The assessment task consisted of an aural prompt (playback of a recorded voice), which the user had to repeat. Prior to each assessment session, the procedure was described to the participant and the participant warmed up by saying a few words.

During the first assessment session, after a brief warm up, the participant was prompted to say each of 47 words in the default keyword set. The keywords were presented to the participant in a random order. The ASR system processed their input and recorded the prompts and recognized words in a log file.

Next, in an interface customization session, one of the researchers spent a one-hour sessions with each of the four participants. During this time, the keyword set was refined and problem words were identified and replaced. The same researcher conducted all of the customization sessions. During these sessions, he interviewed the participant and worked with him or her to identify candidate words to be included in the customized list. This phase differed from automatic acoustic model training in that, while the system was used to verify the recognition ease of candidate keywords, it did not attempt to adapt itself to the user automatically by analyzing a large training dataset but was customized utilizing the user's knowledge and (the knowledge of his or her teachers, caregivers and parents) directly. During the interface customization session, the researcher consulted with the parents, caregivers and teacher of participants 1 and 2, and incorporated their suggestions into the customized list. For participants 3 and 4, the researcher only used suggestions from the participants themselves. For both groups, the researcher incorporated keywords that he observed were clearly pronounced by each participant as well. The outcome of this session for each participant was a new customized keyword set that was stored in a user profile.

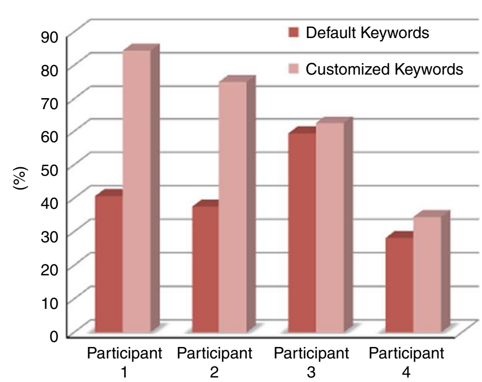
A second assessment session was then conducted during which the ASR module was tested with the modified keyword set. The second assessment session was identical in procedure to the first assessment session and a second log file recording the results was created.

Figure 2 shows the results of the two assessment sessions. Recognition accuracy improved remarkably for participants 1 and 2 and improved marginally for participants 3 and 4. The best results were achieved for participants 1 and 2. The accuracy rate for both participants doubled with the customized list (from 40.6 to 84.3 percent for participant 1 and from 37.5 to 75 percent for participant 2). The accuracy rates showed small improvement for participants 3 and 4. They increased from 56.2 to 62.5 percent showing a 6.3 percent increase for participant 3 and from 28.1 to 34.3 percent showing a 6.2 percent increase for participant 4. In comparison with accuracy rates from the pilot study, the rates for users with dysarthric speech are low. However, in AT lower recognition rates are acceptable for applications in which error correction is not costly or speech can be supplemented by other input modes (Rosen and Yampolsky, 2000).

The results show a dramatic improvement in accuracy rates for participants for whom feedback from parents, caregivers and teachers was incorporated. However, in contrast, the improvements were marginal for participants for whom this information was not available. This feedback, when present, helped create a better keyword set and thus dramatically improved the performance. The results show that input from the user's community, when incorporated, helped with the keyword customization process. Thus, identifying and engaging other stakeholders (especially communication partners) in the customization process is important and should be taken into account. We believe that keyword sets can be further improved by soliciting and incorporating input from the user's community more systematically (e.g. by using questionnaires).

We found the selection of a suitable keyword set a nontrivial problem. The customization of the word list for each participant was performed by trial and error. Moreover, the researcher was not familiar with the participants prior to the testing. We anticipate that this process will be more effective if a speech-language pathologist evaluates the user's speech in a short session and

**Figure 2** Phase 1: keyword recognition accuracy



suggests a list of words that are easy for him or her to pronounce. This process can also be made more effective by incorporating suggested lists compiled by speech-language pathologists for each word.

From a design perspective, in addition to helping find suitable keywords, the participants made suggestions about the interface look and feel (e.g. having different colors on the keyword panel to make it easier to read them), thus engaging in the prototyping stage. However, the focus was on customizing the interface for each person, rather than making major changes that might apply to other users of the system or that would cause major changes to the interface. At this point, while we appreciated and valued the participants' input, we were mentally not yet ready for viewing them as co-designers. This step came in the next iteration of the design process: in phase 2 of the study.

The main outcome of phase 1 was a robust working prototype that was tested with participants with dysarthria of speech. During phase 2 of the design, we used a cooperative prototyping approach (Grønbæk *et al.*, 1997) during which we used this working prototype as a communication tool and a starting point to try out new ideas suggested by the participant. In previous work, using such a "technology probe" has been shown to be effective as a communication tool (Dawe, 2007) and as we will present in the next section, our observations confirm this.

### *4.3 Phase 2*

The second phase started off as an in-depth user study with a single participant. The participant is in her 30s. She has Friedreich's ataxia, a genetically inherited neurodegenerative condition. She has some use of the keyboard and mouse. Although she is able to type, the sustained use of the keyboard and/or mouse is fatiguing for her. She has moderate to severe dysarthria of speech. She has some prior experience with speech recognition software (Dragon Naturally Speaking) but even after training the system accuracy rates were very low for her and it took much repetition and effort to get a word recognized. She uses a wheelchair for mobility. This profile is typical of her underlying condition (Friedreich's ataxia).

This user was interested to obtain a multimodal interface to access the WWW and to use her e-mail client (among other tasks). Prior to this study, she has no design experience and, although an avid computer user, she had no programming or technology development experience.

Initially, the purpose of this phase was to engage the participant in a more focussed iteration of PD and to customize the CanSpeak system so that it would be tailored for her specific needs. However, from the very outset, the participant was interested in exploring ways that through novel changes the interface could be made useful for other people. Thus, by making excellent design suggestions that were incorporated in the design and tested by her, she became a co-designer of the system and a co-author of this paper. For clarity we will refer to her, as the participant in the rest of the paper.

In total, eight half-hour sessions were organized over three months between the participant and the first author, roughly half of the sessions were spent on setting up and testing the system and the other half brainstorming and prototyping with the participant. The sessions were kept relatively short to avoid straining the participant's voice. The interface, along with a microphone setup, was left at the participant's residence so that she could use it in-between visits.

The main feature of the co-design process was a questioning of the assumptions on which the original design was built and, through this questioning, to come up with alternatives. Many good ideas started with "what if" questions. These inspired ideas in the first author, as well as, in the participant. During this phase, a common language that focussed on the function rather than the underlying mechanism of the prototype was developed and used through dialogue, an approach recommended in previous research (Spinuzzi, 2005). Two aspects of the system were modified radically to incorporate the participant's ideas: mapping design and keyword selection.

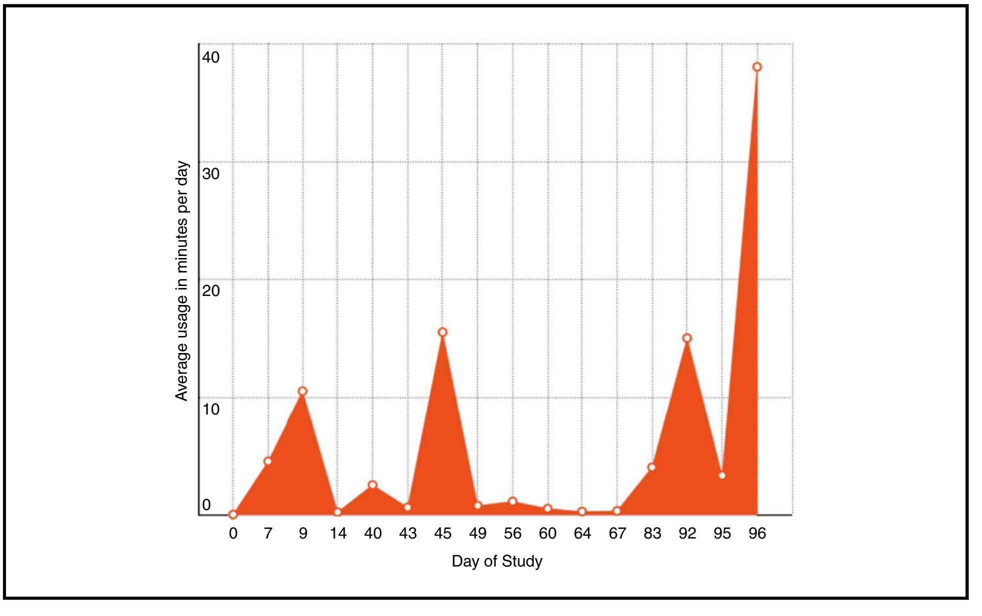
*4.3.1 Mapping design*. Based on initial interviews with the participant, we decided to use speech commands to activate common functions she was using regularly (e.g. activate text composition and e-mail applications, typing short messages). After an initial interaction period, she mentioned that for her, typing is not too difficult and only selecting multiple keys (e.g. Ctrl + C) is difficult.

She also mentioned that it would be better if the interface was not restricted to one or two applications and could be used over several applications. Taking these recommendations in mind, we mapped the keywords to combination keys. The third and final change came after a few weeks of the participant interacting with the new version of the program. She mentioned that it would be more helpful if instead of combination keys we could map the keywords to mouse clicks: right-click, left-click and double-click. After this change, the participant started to use the system more and has since expressed satisfaction with its function. Over time, she found the most useful application of the interface was using it to open e-mail attachments by right clicking on them and opening or saving them to another location on the computer. Prior to collaborating with her, our design team had not thought of this new functionality.

*4.3.2 Keyword selection*. Because of the severity of her dysarthria, it quickly became apparent that the default setting of the interface, as developed during phase 1, had to be changed for her. An initial test showed only a 12.9 percent accuracy rate. Thus, the number of keywords was reduced to four. During the first three sessions a set of four keywords were selected that provided a more robust recognition rate for her. The accuracy was increased dramatically to 77 percent, a significant improvement ($t_{71} = 7.78$, $p < 0.01$). After examining recorded voice samples and interviewing the participant and her caregiver, we decided that perhaps categorizing a number of similar words together and indicating them as one choice might improve the results. In other words, the voice engine was customized so that it categorized each incoming keyword into a group of words (as opposed to mapping them to a single word). Each group contained a set of four words that were minimal pairs with each other, meaning they differed only in one phoneme. From the perspective of the participant, the interaction was the same and she had to say one word per command. While in this case, this approach did not improve recognition, we still believe that for some participants, it might prove effective as the categorization of the keywords into small pockets further simplifies the ASRtask as the recognition needs not to be as precise as before and could correspond to a category rather than a single word. Unfortunately, this technique did not improve recognition accuracy and the recognition rate was 73 percent. This was not a significant difference ($t_{71} = 0.27$, ns).

Figure 3 shows the intensity of use over time. Following, we provide an interpretation of the data based on our interaction with the participant and interviews during the design process. During the

**Figure 3** Phase 2: intensity of use over time



first two weeks of use, there was initial interest and enthusiasm associated with the novelty of the system. This was followed by some time where the low recognition rates of the system were a barrier to successful interaction. Around day 43, the categorization scheme that allows for better recognition was introduced. This resulted in more use before a period of not use. This period coincided with holidays during which time the participant was travelling and did not use the system. Around day 67, new design ideas with respect to control mapping were applied which resulted in higher use intensity.

## 5. PD and AT: a discussion

The results clearly show the benefits of co-creating AT with representatives of the user population. By seeking out and incorporating the users' perspective, PD has the potential of facilitating more relevant designs. The idea of working closely with people with disabilities when developing AT is well accepted. However, we are proposing a more radical perspective: one that not only uses the participant's help in evaluating a design but, actually, empowers him or her to directly make and shape design decisions. While this method might take more time and need more logistic planning, the results might include novel and unusual ideas that other methods might not be able to reveal.

As much as a presentation of our experimental results from the development of CanSpeak, this paper is an account of our relationship with the users of the system. Initially, the system was developed with a specific individual. In phase 1 of the study, the participants were helping to customize the system to match their individual speech and giving feedback about the user interface's look and feel. In phase 2, however, the participant soon went beyond customizing and started to suggest major changes to the design, modifying the initial design completely and, in effect, making it something new. While in phase 1, the researchers were the people in charge of the technology (i.e. modifying and tweaking it behind the scenes), in phase 2, they became co-designers and discussed design decisions openly with the participant.

The cooperative prototyping approach used in phase 2 was important in facilitating communication and collaboration. If we had approached the participant without an initial design, our collaboration might not have been as successful. Thus, results from phase 1, which helped in designing a robust working prototype, helped set the stage for phase 2. In fact, in some situations, it might be a good idea to conduct initial prototyping before starting the co-design process. We believe the key to creativity is setting good limits and boundaries and, in this sense, having a working prototype can be an excellent starting point for the user to imagine possible changes, as it provides a set of limitations on how the system can do and would look like. It should be said that it is important for the prototype not to be too polished and detailed as this might give the impression that the design is almost finished and cannot be changed.

Working from the prototype, we were able to come up with many variations of the system: ones that replaced keystrokes, key combinations and/or mouse clicks with voice commands. Having these variations allows for the inclusions of preset configuration and customizable options for future users.

An important part of the project was taking into account the community in which the participants are situated. We recognize that identifying and taking into account the community of the user is important. Similar to contextual design, we view the user as not a single individual using a technology in isolation but as an entity situated within a whole ecosystem of people and technologies (Beyer and Holtzblatt, 1997). Thus, it is important to mine knowledge in the community of the user by conducting interviews, informal observation and gather feedback on a technology by demonstrating its possible use and soliciting feedback.

In the current project, community was involved in two different ways. In phase 1, input from the participants' community was used to come up with a suitable keyword set for each person. In phase 2, we contacted our co-designer through a support network for people with

Friedreich's ataxia in Canada (http://friedreichsataxia.tripod.com). Throughout the co-design process, the participant was motivated to share the results of the research with her friends on the network. Her vision was always to share the results of her work with other people who have similar disabilities.

## 6. Conclusions and future work

We have presented an example of using PD and co-design in the development, refinement and customization of a small-vocabulary speech interface, CanSpeak, for users with dysarthria of speech. CanSpeak was designed and assessed in two phases: in phase 1, we worked closely with four participants and their families to customize the interface for their use and in phase 2, we used the working prototype developed in phase 1 as a starting point to co-design a new version of the interface with a participant with dysarthria of speech. CanSpeak can be customized in two ways that affect its performance and use significantly. Keyword selection is done by users and/or members of their community and involves choosing keywords that are easy to say. Mapping design involves associating specific system commands and input with the keywords.

Future work can improve both ways to customize CanSpeak. Keyword selection is currently guided by trial and error and can be improved by providing a series of guidelines (possibly developed in collaboration with a speech-language pathologist) as to what kinds of keywords would be good candidates. These guidelines can be provided for different groups of users with similar speech disorders. Currently, mapping design has to be done by researchers. We are working on a version of the program where mappings can be provided as configuration options.

In our experience, we found co-designing with users with disabilities rewarding, as they possess invaluable tacit domain knowledge. The participant/co-designer in phase 2 is an active member of a community of people with the same disability. She is interested in sharing CanSpeak with members of her community. In future, we plan to co-facilitate co-design sessions, similar to phase 2, using CanSpeak with more participants with disabilities.

## Note

1. We use an example to illustrate keyword selection: while "Alpha" was initially used and was easy to pronounce for several users, two users had difficulty with it. The first user decided to use a longer multi-syllable word, "Alligator" instead. This worked well for her but for the second user neither "Alpha" nor "Alligator" worked well. Instead, she came up with the word "Away," which worked well. As illustrated by this example, the length and/or phonetic complexity of keywords are not necessarily indicative of their usefulness in this context; because of familiarity or frequency of use, sometimes longer or phonetically complex words are easy to pronounce clearly for a particular user.

## References


Anthony, L., Prasad, S., Hurst, A. and Kuber, R. (2012), "A participatory design workshop on accessible apps and games with students with learning differences", in Kurniawan, S. (Ed.), *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility in Boulder, CO, ACM Press*, *New York, NY*, pp. 253-4.

Beyer, H. and Holtzblatt, K. (1997), *Contextual Design: Defining Customer-Centered Systems*, Morgan Kaufmann, Burlington, MA.

Clark, J.A. and Roemer, R.B. (1977), "Voice controlled wheelchair", *Archives of Physical Medicine and Rehabilitation*, Vol. 58 No. 4, pp. 169-75.

Cohen, A. and Graupe, D. (1980), "Speech recognition and control system for the severely disabled", *Journal of Biomedical Engineering*, Vol. 2 No. 2, pp. 97-107.

Davidson, J.L. and Jensen, C. (2013), "What health topics older adults want to track: a participatory design study", in Ladner, R. (Ed.), *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility in Bellevue, WA, ACM Press*, *New York, NY*, pp. 1-8.

Dawe, M. (2007), “Let me show you what I want: engaging individuals with cognitive disabilities and their families in design”, in Gilmor, D. (Ed.), *CHI'07 Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, ACM Press*, *New York, NY*, pp. 2177-82.

Franco, H., Abrash, V., Precoda, K., Bratt, H., Rao, R., Butzberger, J., Rossier, R. and Cesari, F. (2000), “The SRI EduSpeak system: recognition and pronunciation scoring for language learning”, *Proceedings of Integrating Speech Technology in (Language) Learning in Dundee*, pp. 123-8.

Galliers, J., Wilson, S., Roper, A., Cocks, N., Marshall, J., Muscroft, S. and Pring, T. (2012), “Words are not enough: empowering people with aphasia in the design process”, in Halskov, K., Winschiers-Theophilus, H., Lee., Y., Simonsen, J. and Bødker, K. (Eds), *Proceedings of the 12th Participatory Design Conference in Roskilde, Denmark, ACM Press*, *New York, NY*, pp. 51-60.

Grammenos, D., Savidis, A., Georgalis, Y. and Stephanidis, C. (2006), “Access invaders: developing a universally accessible action game”, in Miesenberger, K., Klaus, J., Zagler, W. and Karshmer, A. (Eds), *Proceedings of International Conference on Computers Helping People with Disabilities in Linz, Austria, Springer*, *Berlin and Heidelberg*, pp. 388-95.

Greenbaum, J. and Kyng, M. (Eds) (1991), *Design at Work: Cooperative Design of Computer Systems*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Grønbæk, K., Kyng, M. and Mogensen, P. (1997), “Toward a cooperative experimental system development approach”, *Computers and Design in Context*, MIT Press, Cambridge, MA, pp. 201-38.

Hamidi, F., Baljko, M., Livingston, N. and Spalteholz, L. (2010), “CanSpeak: a customizable speech interface for people with dysarthric speech”, *Proceedings of the International Conference on Computers Helping People with Special Needs (ICCHP10), Springer*, *Berlin*, pp. 605-12.

Hawley, M., Enderby, P., Green, P., Brownsell, S., Hatzis, A., Parker, M., Carmichael, J., Cunningham, M., ONeill, P. and Palmer, R. (2003), “STARDUST: speech training and recognition for dysarthric users of assistive technology”, in Craddock, G.M.(Ed.), *Assistive Technology – Shaping the Future*, IOS Press, Amsterdam, pp. 959-64.

Hurst, A. and Tobias, J. (2011), “Empowering individuals with do-it-yourself assistive technology”, in Yesilada, Y. (Ed.), *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility in Dundee, Scotland, ACM Press*, *New York, NY*, pp. 11-8.

Kensing, F. and Blomberg, J. (1998), “Participatory design: issues and concerns”, in Poltrock, S. and Grudin, J. (Eds), *Proceedings of Computer Supported Cooperative Work in Seattle, WA, ACM Press*, *New York, NY*, pp. 167-85.

Kewley-Port, D., Watson, C.S., Elbert, M., Maki, D. and Reed, D. (1991), “The Indiana speech training aid (ISTRA) II: training curriculum and selected case studies”, *In Clinical Linguistics and Phonetics*, Vol. 5 No. 1, pp. 13-38.

Moffatt, K., McGrenere, J., Purves, B. and Klawe, M. (2004), “The participatory design of a sound and image enhanced daily planner for people with aphasia”, in Dykstra-Erickson, E. and Tscheligi, M. (Eds), *Proceedings of the SIGCHI conference on Human factors in computing systems in Vienna, ACM Press*, *New York, NY*, pp. 407-14.

Phillips, B. and Zhao, H. (1993), “Predictors of assistive technology abandonment”, *Assistive Technology*, Vol. 5 No. 1, pp. 36-45.

Rogers, Y. and Marsden, G. (2013), “Does he take sugar?: moving beyond the rhetoric of compassion”, *Interactions*, Vol. 20 No. 4, pp. 48-57.

Rosen, K. and Yampolsky, S. (2000), “Automatic speech recognition and a review of its functioning with dysarthric speech”, *Journal of Augmentative and Alternative Communication*, Vol. 16 No. 1, pp. 48-60.

Rosengren, E., Raghavendra, P. and Hunnicutt, S. (1995), “How does automatic speech recognition handle severely dysarthric speech?”, in Placencia Porrero, I. and de la Bellacas, P. (Eds), *Proceedings of the 2nd TIDE Congress in Paris, France*, IOS Press, Amsterdam, pp. 336-9.

Sanders, E.B.N. and Stappers, P.J. (2008), “Co-creation and the new landscapes of design”, *CoDesign: International Journal of CoCreation in Design and the Arts*, Vol. 4 No. 1, pp. 5-18.

Schuler, D. and Namioka, A. (Eds) (1993), *Participatory Design: Principles and Practices*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Spalteholz, L., Li, K.F., Livingston, N. and Hamidi, F. (2008), “Keysurf: a character controlled browser for people with physical disabilities”, in Ma, W.Y., Tomkins, A. and Zhang, X. (Eds), *Proceedings of International World Wide Web Conference in Beijing, ACM Press*, *New York, NY*, pp. 31-40.

Spinuzzi, C. (2005), "The methodology of participatory design", *Technical Communication*, Vol. 52 No. 2, pp. 163-74.

van Teijlingen, E. and Hundley, V. (2002), "The importance of pilot studies", *Nursing Standard*, Vol. 16 No. 40, pp. 33-6.

Visser, F.S., Stappers, P.J., Van der Lugt, R. and Sanders, E.B. (2005), "Contextmapping: experiences from practice", *CoDesign*, Vol. 1 No. 2, pp. 119-49.

Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., Wolf, P. and Woelfel, J. (2004), "Sphinx-4: A flexible open source framework for speech recognition", Technical report, Sun Microsystems, Menlo Park, CA.

Wu, M., Richards, B. and Baecker, R. (2004), "Participatory design with individuals who have amnesia", in Clement, A. and van den Besselaar, P. (Eds), *Proceedings of the 8th Participatory Design Conference in Toronto, ON, ACM Press*, *New York, NY*, pp. 214-23.

## Further reading

Beukelman, D.R. and Mirenda, P. (1998), *Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults*, 2nd ed., Paul H. Brooks Publishing Co., Baltimore, MD.

Bødker, K., Kensing, F. and Simonsen, J. (2004), *Participatory IT Design: Designing for Business and Workplace Realities*, MIT Press, Boston, MA.

Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B.B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N. and Eiderbäck, B. (2003), "Technology probes: inspiring design for and with families", in Cockton, G. and Korhonen, P. (Eds), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems in Fort Lauderdale, FL, ACM Press*, *New York, NY*, pp. 17-24.

Kawai, S., Aida, H. and Saito, T. (1996), "Designing interface toolkit with dynamic selectable modality", in Glinert, E.P. and Jaffe, D.L. (Eds), *Proceedings of the 2nd International ACM SIGACCESS Conference on Computers and Accessibility in Vancouver, BC, ACM Press*, *New York, NY*, pp. 72-9.

Loi, D. (2004), "Document/reflect/create – cultural probes in teaching and learning environments", in Clement, A. and van den Besselaar, P. (Eds), *Proceedings of the 8th Participatory Design Conference in Toronto, ON, ACM Press*, *New York, NY*, pp. 123-6.

Oviatt, S., Coulston, R., Tomko, S., Xiao, B., Lunsford, R., Wesson, M. and Carmichael, L. (2003), "Toward a theory of organized multimodal integration patterns during human-computer interaction", in Darrell, T., Maybury, M. and Wahlster, W. (Eds), *Proceeding of the 5th International Conference on Multimodal Interfaces in Vancouver, BC, ACM Press*, *New York, NY*, pp. 44-51.

## About the authors

Foad Hamidi is a PhD candidate at the Lassonde School of Electrical Engineering and Computer Science at the York University and a Member of the GaMay Lab. His research interests include intercultural collaboration, assistive technology and digital activism. The underlying theme of his work is using technology for empowerment and community engagement. Foad Hamidi is the corresponding author and can be contacted at: fhamidi@cse.yorku.ca

Dr Melanie Baljko is an Associate Professor of Computer Science and Engineering at the York University and the Head of the GaMay Lab. Her research concerns the design and evaluation of computational systems, such as those that provide assistance for individuals with disabilities and that are used clinically for speech disorders. She has a PhD and MSc in Computer Science from the University of Toronto and a BMath from the University of Waterloo.

Connie Ecomomopoulos is a user and Co-designer of Assistive Technology. She has been an active member of the Freidreich's Ataxia Made Easier (F.A.M.E.) community and has collaborated with GaMaY Lab since 2012.

Dr Nigel J. Livingston received the PhD Degree in Biometeorology from the University of British Columbia, Vancouver, BC, Canada, in 1986. He is a Professor in the School of Public Health and Social Policy at the University of Victoria and the founder of the assistive technology organization CanAssist, University of Victoria, Victoria, BC, Canada. He has over 30 years of experience in the design of environmental instrumentation and assistive technology.

Leonhard G. Spalteholz received a BEng. Degree in Computer Engineering and a MASc Degree in computer engineering, both at the University of Victoria, Victoria, BC, Canada, in 2006 and 2012, respectively. He has been working in the field of assistive technology since 2004 and has been the Manager of software projects at CanAssist, University of Victoria, Victoria, BC, Canada, since 2008. His research experience includes a low-cost zero-setup eye tracking system for people with Rett syndrome, a gait analysis system using wireless sensors, and novel web navigation software for persons unable to use a mouse. He specializes in adapting research results into practical tools for the disabled.