Impact of intelligent personalization on communication effectiveness in dysarthria

Mansoor Pervaiz, Northeastern University, Boston, MA
Advised by Rupal Patel, Northeastern University, Boston, MA

THESIS:
Intelligent interfaces that personalize for behavior and context can enhance communication effectiveness for speakers with dysarthria

ABSTRACT
Communication is essential for conveying one’s needs, feelings, and ideas, and for building social bonds. Individuals with dysarthria feel isolated and seek help from clinicians to improve communication. Based on impairment severity, clinicians prescribe either speech therapy or assistive technology. Speech therapy is infrequent and while successful in clinic, adaptations are hard to maintain in daily conversation. Although assistive technologies provide an alternative method of communication for those with severe impairment, current approaches are slow, fatiguing and do not adapt to user’s varying situational needs. Recent advances in mobile and embedded technologies afford an opportunity to capture speech behavior and situational context. This dissertation proposes that intelligent interfaces can learn an individual’s behavior and context to improve effectiveness of communication for those with speech impairments.

1. INTRODUCTION
Over the past decade, there has been a wide spread adoption and increase in computation power of mobile devices (“Pew Internet & American Life Project. Internet, broadband, and cell phone statistics,” 2010). Given that these devices are carried with us throughout the day (Venta, Isomursu, Ahtinen, & Ramiah, 2008), a mobile device can provide substantial contextual data at any point in time (Fogg, 2007). Through embedded sensing (such as GPS location tracking, accelerometer-based exercise detection), and access to the user’s calendar, contacts and other personal information, mobile phone applications can infer where a user is and what (s)he is doing. These data make it possible to create just-in-time interventions that provide users with support at times when support is most needed (Intille, 2004; Patrick et al., 2009). This ability is especially useful for supporting interventions for communication impairments, in which the clinical setting does not adequately capture real-world communication scenarios. Interventions for communication impairments rely on self-reported data of user behavior and performance in varying environments once outside of the therapy room. However, the ubiquity of information regarding user behavior and context in daily communication provides a remarkable opportunity to target a client’s constantly varying needs with higher precision.
2. COMMUNICATION
Communication is multifaceted and serves multiple purposes including basic wants and needs, information transfer, social closeness, social etiquette, and self communication (D. Beukelman & Mirenda, 2005; Light, 1988). Communication activities are pervasive in the home, work, school, leisure and community settings and account for over 60% of the work day (Hinrichs, 1964). The ability to communicate effectively is a critical factor in higher quality of life. (Ruben, 2000). Communication impairments impact all aspects of life – education, medical care, employment, family and social involvement (D. Beukelman & Mirenda, 2005) – and can lead to participation that is less diverse, is restricted more to the home setting, and involves fewer friends and smaller social networks (Davidson, Howe, Worrall, Hickson, & Togher, 2008; Law, 2002).

3. COMMUNICATION IMPAIRMENT
There are many types of communication impairments. These impairments may impact cognition, language and or speech. This proposal focuses on the latter. Dysarthria is a motor speech disorder that can disrupt this complex system and alter speech patterns and movement of the muscles. Dysarthria is caused by a wide range of neurological conditions such as Parkinson's disease (PD), multiple sclerosis (MS), cerebral palsy (CP), traumatic brain injury (TBI), amyotrophic lateral sclerosis (ALS), myasthenia gravis (GS), Huntington’s disease (HD) and stroke (Duffy, 2012). Due to the broad range of etiologies, there are no known data on the incident rate of dysarthria in the general population. However, based on the incidence rates of dysarthria in PD, MS, and stroke alone, more than 3 million individuals in the U.S. are estimated to present with dysarthria (Darley, Brown, & Goldstein, 1972; de Lau & Breteler, 2006; Mackenzie, 2011; Miller, Deane, Jones, Noble, & Gibb, 2011; Mozaffarian et al., 2015; Tullman, 2013; Yao, Hart, & Terzella, 2013).

Perceptual characteristics of dysarthria include imprecise articulation, reduced or inappropriate prosody and voice quality distortions (Darley, Aronson, & Brown, 1969b, 1969a) that reduce speech intelligibility (Duffy, 2012). Many individuals consider the inability to communicate as one of the most difficult aspects of dysarthria (Pitcairn, Clemic, Gray, & Pentland, 1990). Therefore, to alleviate this strain speech interventions aim to improve communication abilities which may enhance social participation and improve quality of life (Eadie et al., 2006; O’Halloran, Hickson, & Worrall, 2008).

4. MANAGEMENT OF COMMUNICATION IMPAIRMENT
The primary goal of speech and language clinicians is to maximize the effectiveness, efficiency and naturalness of communication (Rosenbek, LaPointe, & Johns, 1985). For individuals with dysarthria, based on their severity, clinicians either opt to either improve efficiency and naturalness of an individual’s speech or to improve effectiveness of assistive devices (Duffy, 2012).

4.1. Speech Therapy
Currently, a wide variety of techniques and strategies are used to treat and manage dysarthria. Two primary strategies are extensively used to improve speech intelligibility...
in mild to moderate dysarthria: 1) reduced speech-rate (Yorkston, 2007), and 2) increased vocal loudness (Solomon, McKee, & Garcia-Barry, 2001; Ward, Theodoros, Murdoch, & Silburn, 2000). Reduced speech rate provides speakers with more time to attain articulatory precision (Duffy, 2012) and listeners with increased processing time (Nishio, Tanaka, Sakabibara, & Abe, 2011). Increased vocal loudness improves audibility (Neel, 2009) and increases the amplitude and coordination of articulatory muscles. The Lee Silverman Voice Treatment (LSVT) (Ramig, Bonitati, Lemke, Horii, & Others, 1994) applies this principle for individuals with Parkinson's disease (PD) (Ramig et al., 2001) and has shown promise for multiple sclerosis (Sapir et al., 2001), cerebellar dysfunction (Sapir et al., 2003), stroke (Mahler, Ramig, & Fox, 2009), cerebral palsy (Boliek et al., 2009; Fox & Boliek, 2012) and Down syndrome (Petska, Halpern, Ramig, & Robinson, 2006). Although successful in clinical settings, adherence in daily conversation is challenging because of the cognitive load of monitoring one’s speech, and fatiguing due to maximal effort required in loudness training. Smart devices (mobile phone and wearable devices) can improve adherence by adapting to user speech behavior and context. Mobile devices can monitor an individual’s speech characteristics and situational context to provide just-in-time adaptive feedback cues to monitor vocal characteristics in daily conversation and adapt to situational changes like location, ambient noise, and user fatigue.

4.2. Assistive Technologies
For individuals with severe dysarthria, speech production may be supplemented or replaced with augmentative and alternative communication (AAC) assistive devices. AAC encompasses a range of techniques from the use of sign language and letterboards to speech generating devices (SGD). A SGD includes a letter or picture based user interface to support message formulation and a text to speech module to output the message. Current SGDs are relatively passive means for translating user intentions into spoken output. Communicating via these systems is slow and physically demanding because it requires considerable effort to search for, and navigate to, desired items (Udwin & Yule, 1990). The proposed work posits that adding an adaptive or contextual intelligence layer will enhance communication efficiency.

5. CHALLENGES WITHIN CURRENT APPROACHES
Historically, therapists have relied on client self report as a window into user behavior and needs outside the clinical setting which is not always reliable. Furthermore, the efficacy of tradition intervention is limited by the lack of: 1) clinician cues in daily conversation, 2) personalization to situational context, and 3) personalization to user abilities. Access to multiple data streams that provide knowledge of user behavior and context has the potential to transform clinical practice.

5.1. Lack of clinician cues in daily conversations
Clinicians employ structured in-clinic treatment sessions and usually provide follow-up exercises to be completed outside the therapy session to reinforce and generalize speech behavior change. During structured treatment sessions, speech-language pathologists (SLPs) provide explicit feedback cues to improve intelligibility through rate reduction or vocal loudness elevation techniques. Explicit cues incorporate clear directions and
instructions that require an individual to consciously manipulate speech movements to reach a desired goal (Gazzaniga, Ivry, & Mangun, 1998). Outside of the structured setting (in daily conversations), clinicians expect individuals to adhere to training by relying on implicit self-cues like reactions to sensations or habits (Gazzaniga et al., 1998). However, some individuals require continued explicit cueing outside of the speech clinic to maintain speech goals (Morris, Iansek, McGinley, Matyas, & Huxham, 2005; Oliveira, Gurd, Nixon, Marshall, & Passingham, 1997), which can be achieved through technological innovation. Wearable devices that process user speech and provide real time feedback can improve adherence in daily conversation.

5.2. Lack of personalization to situational context

5.2.1. Speech training
Traditional loudness training requires maximal effort at all times, because it does not take situational context (e.g., the speaker's fatigue level or ambient noise of location) into account. This potentially unnecessary constant effort may lead to fatigue, which is a significant factor in neuromotor impairments and restricts participation (Brunstrom, 2001; Hartelius, Wising, & Nord, 1997). Therefore, users may benefit from an adaptive loudness requirement based on the ambient noise of their location. Such personalization may ease the need for maximal effort and reduce vocal fatigue.

5.2.2. AAC
Communication through SGD is 15 to 25 times slower than natural speech because it takes considerable time and effort to search for and navigate to desired items while constructing messages. This challenge is exacerbated by the upper motor limb disorders associated with neurological impairments (Bloxham, Dick, & Moore, 1987; Cohen & Rae-Grant, 2010; Flowers, 1976). Despite the slow communication rate, commercial AAC systems have not yet focused on accommodating to individual vocabulary needs that vary based on age, gender; situational context; social role, and living environments (D. Beukelman & Mirenda, 2005) and only support 55% of the vocabulary required by AAC users in adult social contexts (Bryen, 2008). Predicting user input based on situational context (location, time of day, and conversational partner data) can improve communication effectiveness of AAC devices, which is possible due to improvements in mobile sensing and computation.

5.3. Lack of personalization to user abilities
Despite having some residual speech, individuals with moderate dysarthria are taught to use AAC systems because unfamiliar conversational partners cannot understand them. However, even individuals even with severe dysarthria often reject AAC devices and continue to use their residual speech (Fager, Hux, Beukelman, & Karantounis, 2006; Smith and Ford, 1987). Affording those with moderate dysarthria a means to use their residual vocalizations to access their AAC device may improve adoption and ultimately enhance communication.

6. PROPOSED WORK
The information used for personal health care today largely comes from self-report surveys and infrequent doctor consultations. Mobile devices offer a unique opportunity
for communication researchers to improve impact of interventions to individuals more frequently and at a much more personal level. These devices not only serve as computing and communication devices but also come with a rich set of sensors that improves learning of individual behavior and context (Consolvo et al., 2008). Moreover, the near constant connectivity enables aggregation of further knowledge by offloading computation of large-scale sensor data processing to back-end servers. The ubiquity of mobile devices with advanced sensing and processing capabilities has the potential to transform outcomes for individuals with dysarthria across the severity range. For those with mild impairment, an in-situ adaptive speech behavior training mode is proposed. For speakers with moderate dysarthria, a multimodal system is proposed to empower users to access all viable methods of access. Finally, the proposed work makes use of situational data (location, time of day, and conversational partner) to enhance communication efficiency for AAC users with severe dysarthria.

6.1. Speech Behavior Change

6.1.1. Prior investigation

Use of technology in speech therapy is promising but primarily limited to telepractice (Constantinescu, Rosengren, Johnels, Zetterberg, & Holmberg, 2010). An exception is the incorporation of near real-time clinician feedback with LSVT Companion and real-time feedback with LAPP. LSVT Companion (Halpern et al., 2012) is a PDA application that extends the traditional loudness training: Lee Silverman Voice Therapy (LSVT) by allowing users to practice a subset of clinician sessions at home. However, this system doesn't provide any feedback during daily conversation, and expects maximum effort at all times. LAPP (McNaney et al., 2015) is a Google glass application that displays users a thumbs up as a visual feedback during daily conversation to achieve predetermined target volume. LAPP addresses the need of feedback in daily conversation. However, feedback is limited because of its Boolean (success/failure) design. This design doesn't allow a user to determine how close (s)he is to the target. Yet this information is important for motivation, particularly when a speaker is very close to the target loudness but is too fatigued to put in more effort.

6.1.2. Proposed work

The proposed work will build upon the previous work demonstrating that increased vocal loudness can improve speech clarity and extend previous technological interventions by: 1) providing higher resolution feedback in daily conversation, and 2) adjusting loudness target based on ambient noise levels in order to minimize fatigue and maintain appropriate loudness levels. This will involve development of a Google glass application that will monitor vocal loudness and provide visual feedback in real time, and evaluate the efficacy of the artifact through a user study.

6.1.3. Hypothesis

Individuals with mild dysarthria can modulate their vocal loudness in daily conversation to improve communication when provided with real time visual feedback.
6.2. Multimodal AAC:

6.2.1. Prior investigation
One challenge with acceptance of AAC devices is that individuals with moderately dysarthric speech prefer to use their own voice even though they are difficult to understand and need to repeat themselves often (D. R. Beukelman, Fager, Ball, & Dietz, 2007). To improve acceptance of AAC, researchers have investigated various approaches to incorporate dysarthric speech in AAC applications. They have applied automatic speech recognition to dysarthric speech (Ferrier, 1991; Treviranus, Shein, Haataja, Parnes, & Milner, 1991; Kotler & Thomas-Stonell, 1997) but were only moderately successful for mild dysarthria. Researchers have applied multimodal recognition to improve speech recognition by combining acoustic speech signals with surface electromyographic (sEMG) signals measured from the face and neck and found that the additional sEMG signals boosted recognition of disordered speech by as much as 12% (Deng et al., 2009). Rudzicz, Hirst, and van Lieshout (2012) demonstrated improved recognition of disordered speech using rich articulatory information from electromagnetic articulography (EMA).

6.2.2. Proposed work
This proposed research would extend previous work by introducing the use of lingual and labial kinematics along with dysarthric acoustics to predict the intended message of an individual. These data streams will be collected using LinKa translator (Huo, Wang, & Ghovanloo, 2008) which simultaneously captures acoustic speech, tongue movement and lip movement. This work will collect a dataset of dysarthric utterances with multiple streams of data: speech, lingual and labial kinematics and evaluate the prediction statics using machine learning approaches: Hidden Markov models and deep belief networks.

6.2.3. Hypothesis
Multiple input modalities (acoustic, lingual, and labial) improve recognition of dysarthric speech.

6.3. Context aware AAC

6.3.1. Prior investigation
Individuals vary their vocabulary based on their conversational partner (family, friends, or strangers), location (home, work or school) and time of day (morning, afternoon or evening). However, current AAC systems only support a very small subset of the total vocabulary needed by users. Some systems do use location for contextual information by allowing users to pre-program words based on different locations. Talk Rocket Go and Chatable use GPS location to load up specific and pre-prepared vocabulary pages that have been previously linked by the user to that location. Instead of pre-programming words, AAC systems can learn the words an individual uses based not only on their location but also time of day and conversational partner.

6.3.2. Proposed work
This work will extend previous research by collecting a corpus of every day text messages with contextual information (time of day, location, and conversational partner). Furthermore, this work will compare impact of contextual information on message
prediction by applying machine learning methodologies: dynamic bayesian network, transfer learning, and transformation rule.

6.3.3. Hypothesis
Contextual information can be leveraged to improve communication effectiveness (number of inputs required to construct target phrases).

7. THESIS COMMITTEE

7.1. Rupal Patel Thesis Advisor
Dr. Rupal Patel is a Professor at Northeastern University and holds joint appointments in the Department of Speech-Language Pathology and Audiology (SLPA) and the College of Computer and Information Science (CCIS). She specializes in understanding neuromotor control of speech disorders and designing and developing assistive communication technologies that leverage the user's capabilities. Given her domain knowledge and experience with the end user group, Dr. Patel will provide overall guidance across several domains including the design of the predictive algorithms and the user interface and the usability studies.

Email: r.patel@neu.edu
Homepage: http://www.cadlab.neu.edu/people.php?name=rupal_patel

7.2. Stephen Intille
Dr. Stephen Intille is an Associate Professor in the College of Computer and Information Science and Bouve College of Health Sciences at Northeastern University. His research focuses on the development of novel healthcare technologies that incorporate ideas from ubiquitous computing, user-interface design, pattern recognition, behavioral science, and preventive medicine. Areas of special interest include technologies for measuring and motivating health-related behaviors, technologies that support healthy aging and well-being in the home setting, and mobile technologies that permit longitudinal measurement of health behaviors for research, especially the type, duration, intensity, and location of physical activity.

Email: s.intille@neu.edu
Homepage: www.ccs.neu.edu/home/intille/

7.3. Frank Rudzicz
Dr. Rudzicz is currently a Scientist at the Toronto Rehabilitation Institute, a status-only assistant professor of Computer Science at the University of Toronto, co-founder and President of WinterLight Labs Inc., and President of the international joint ACL/ISCA special interest group on Speech and Language Processing for Assistive Technologies. He is the recent recipient of the Young Investigator award from the Alzheimer's Society of Canada, the Early Researcher award from the Government of Ontario, and the Excellence in Applied Research award from National Speech-Language & Audiology Canada. His work involves machine-learning, human-computer interaction, speech-
language pathology, rehabilitation engineering, signal processing, robotics, and linguistics.

Email: frank@spoclab.com

Homepage: http://www.cs.toronto.edu/~frank/Home.html

7.4. Yun Raymond Fu
Dr. Fu is an interdisciplinary faculty member affiliated with College of Engineering and the College of Computer and Information Science at Northeastern University. Dr. Fu's research interests are Interdisciplinary research in Machine Learning and Computational Intelligence, Social Media Analytics, Human-Computer Interaction, and Cyber-Physical Systems.

Email: yunfu@ece.neu.edu

Homepage: http://www1.ece.neu.edu/~yunfu/

8. REFERENCE
WILEY-LISS DIV JOHN WILEY & SONS INC, 111 RIVER ST, HOBOKEN, NJ 07030 USA.


