Clear Speech

Technologies that Enable the Expression and Reception of Language

Synthesis Lectures on Assistive, Rehabilitative, and Health-Preserving Technologies

Editor
Ronald M. Baecker, ***University of Toronto***

Advances in medicine allow us to live longer, despite the assaults on our bodies from war, environmental damage, and natural disasters. The result is that many of us survive for years or decades with increasing difficulties in tasks such as seeing, hearing, moving, planning, remembering, and communicating.
This series provides current state-of-the-art overviews of key topics in the burgeoning field of assistive technologies. We take a broad view of this field, giving attention not only to prosthetics that compensate for impaired capabilities, but to methods for rehabilitating or restoring function, as well as protective interventions that enable individuals to be healthy for longer periods of time throughout the lifespan. Our emphasis is in the role of information and communications technologies in prosthetics, rehabilitation, and disease prevention.

Clear Speech: Technologies that Enable the Expression and Reception of Language
Frank Rudzicz
2016

Designed Technologies for Healthy Aging
Claudia B. Rebola
2015

Fieldwork for Healthcare: Guidance for Investigating Human Factors in Computing Systems
Dominic Furniss, Rebecca Randell, Aisling Ann O'Kane, Svetlena Taneva, Helena Mentis, and Ann Blandford
2014

Fieldwork for Healthcare: Case Studies Investigating Human Factors in Computing Systems
Dominic Furniss, Aisling Ann O'Kane, Rebecca Randell, Svetlena Taneva, Helena Mentis, and Ann Blandford
2014

Interactive Technologies for Autism
Julie A. Kientz, Matthew S. Goodwin, Gillian R. Hayes, and Gregory D. Abowd
2013

Patient-Centered Design of Cognitive Assistive Technology for Traumatic Brain Injury Telerehabilitation
Elliot Cole
2013

Zero Effort Technologies: Considerations, Challenges, and Use in Health, Wellness, and Rehabilitation
Alex Mihailidis, Jennifer Boger, Jesse Hoey, and Tizneem Jiancaro
2011

Design and the Digital Divide: Insights from 40 Years in Computer Support for Older and Disabled People
Alan F. Newell
2011

Clear Speech: Technologies that Enable the Expression and Reception of Language
Frank Rudzicz

ISBN: 978-3-031-00471-1 paperback
ISBN: 978-3-031-01599-1 ebook

DOI 10.1007/978-3-031-01599-1

A Publication in the Springer series
SYNTHESIS LECTURES ON ASSISTIVE, REHABILITATIVE, AND HEALTH-PRESERVING TECHNOLOGIES

Lecture #8
Series Editor: Ronald M. Baecker, *University of Toronto*
Series ISSN
Print 2162-7258 Electronic 2162-7266

Clear Speech

Technologies that Enable the Expression and Reception of Language

Frank Rudzicz
Toronto Rehabilitation Institute and
Department of Computer Science, University of Toronto

SYNTHESIS LECTURES ON ASSISTIVE, REHABILITATIVE, AND HEALTH-PRESERVING TECHNOLOGIES #8

ABSTRACT

Approximately 10% of North Americans have some communication disorder. These can be physical as in cerebral palsy and Parkinson's disease, cognitive as in Alzheimer's disease and dementia generally, or both physical and cognitive as in stroke. In fact, deteriorations in language are often the early hallmarks of broader diseases associated with older age, which is especially relevant since aging populations across many nations will result in a drastic increase in the prevalence of these types of disorders. A significant change to how healthcare is administered, brought on by these aging populations, will increase the workload of speech-language pathologists, therapists, and caregivers who are often already overloaded.

Fortunately, modern speech technology, such as automatic speech recognition, has matured to the point where it can now have a profound positive impact on the lives of millions of people living with various types of disorders. This book serves as a common ground for two communities: clinical linguists (e.g., speech-language pathologists) and technologists (e.g., computer scientists). This book examines the neurological and physical causes of several speech disorders and their clinical effects, and demonstrates how modern technology can be used in practice to manage those effects and improve one's quality of life. This book is intended for a broad audience, from undergraduates to more senior researchers, as well as to users of these technologies and their therapists.

KEYWORDS

computational linguistics, speech-language pathology, assistive technologies, rehabilitation science, machine learning

Dedicated to those overcoming barriers in communication

Contents

Preface

When I was a grad student, the purpose of my research was to improve the accuracy of speech recognition software for people with speech disorders. I started by working with cerebral palsy (CP), which remains the most common cause of hard-to-understand speech today. Most people with CP were not very well understood by speech recognition at the time—less than 1% of their words could be correctly recognized whereas a speaker without a speech disorder might be comfortably understood 85% or 90% of the time. It wasn't that their words didn't make sense—people with CP can normally understand and produce *language* just fine—it was that their voices are quite different from those of the general population, which can profoundly confuse speech software. It was my job to un-confuse the software.

Not being understood *almost all of the time* can be annoying in itself—and speech recognition certainly did a dismal job for people with CP. It was therefore perhaps somewhat frustrating that speech was often the *most effective* means of communication these individuals had. Although CP limits the control of the muscles of speaking (e.g., the tongue), CP *also* affects other muscles (e.g., those controlling the fingers). This means that while speech in CP can be approximately three times slower than typical speech, typing can be over a *hundred* times slower.

So if a computer can't understand what you say *and* it takes too long or is too difficult to type by hand, then merely participating in our modern society becomes a tremendous challenge. According to the U.S. Census bureau, less than 10% of people with severe disabilities are employed, partially due to difficulty in communication, which has considerable consequences for social and health well-being.

Something must be done.

So how could I make my own small dent toward cracking this huge problem? Since the *sounds* of speech in cerebral palsy were so difficult for computers to understand, I reasoned that it might help to "teach" the computer *why* those sounds were difficult—to teach it about differences in the *physical* origins of speech. How do you teach a computer? These days, we use MACHINE LEARNING where you basically program the computer to find patterns and relationships in data by itself, typically given lots of carefully curated examples that you provide. In my case, I needed to provide examples of speech sounds and their corresponding vocal tract movement, and for that I needed participants to come into the lab to have their voices and facial movements recorded during speech.

Many of the participants were in their early twenties and came in with their parents or other caregivers. One young man with CP was particularly talkative, and his father was equally eager to insert himself into the conversation, usually to repeat or to clarify what his son said. They were both very outgoing, and we had about as non-serious a chat as you can imagine in a research

setting, in the basement of a satellite building of the University of Toronto. At one point, the young man revealed that one of his main motivations for volunteering (and for getting his dad to take time off of work to drive him into the lab), was "girls." I told him that was not part of our research protocol. This young man's father then chimed in to say that it wasn't so much "girls" as it was a *particular* girl, and that she and his son were "courting,"[1] but communication between them remained difficult. The young man had tried a number of devices and programs to help him be understood, but he found each of them to be insufficient—he didn't feel like he could properly express himself. The alternative to talking through a computer was to talk through the filter of your father, which can also be non-ideal in courtship. He wanted to help us improve the technology.

Can advanced speech technology improve your love life? More data is required. However, what was clear to me from that exchange was that so much of who we are, collectively and as individuals, depends on our ability to communicate. Language is not just about communicating facts or making plans—to a large extent it defines how others perceive us and how we perceive ourselves in the world. Being able to define yourself in your own words—to speak for yourself—is liberating.

I hope that this book can bring together people who really should be talking together, especially technologists, therapists and clinicians, and people affected by speech disorders. Technologists need to know what challenges exist in the real world and how clinicians are currently meeting those challenges. Therapists need to know how artificial intelligence that can help to diagnose, monitor, and overcome issues of communication. Perhaps most importantly, people affected by speech and language disorders need to know that there is light at the end of the tunnel, and that technology is helping to provide that light.

> [Language has a] unique role in capturing the breadth of human thought and endeavour...We look back at the thoughts of our predecessors, and find we can see only as far as language lets us see. We look forward in time, and find we can plan only through language. We look outward in space, and send symbols of communication along with our spacecraft, to explain who we are, in case there is anyone there who wants to know. [Crystal, 1998]

Frank Rudzicz
February 2016

[1] Is that what kids do these days?

Figure Credits

Figure 3.1 Used with permission from Microsoft © 2016.

Figure 5.2 Based on: Paul Boersma. Functional Phonology: Formalizing the interactions between articulatory and perceptual drives. Ph.D. thesis, Universiteit van Amsterdam, September 1998.

Figure 5.3 Based on: Alexander B. Kain, John-Paul Hosom, Xiaochuan Niu, Jan P.H. van Santen, Melanie Fried-Oken, and Janice Staehely. Improving the intelligibility of dysarthric speech. *Speech Communication*, 49(9):743–759, September 2007.

Figure 6.2 Keith L. Moore and Arthur F. Dalley. *Clinically Oriented Anatomy*, Fifth Edition. Lippincott, Williams and Wilkins, 2005. Copyright © 2005 Lippincott, Williams & Wilkins. Used with permission.

Figure 6.4 Based on: Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition.* Prentice Hall, Second edition, 2009.

Figure 8.1 Based on: Fangxin Chen and Aleksandar Kostov. Optimization of dysarthric speech recognition. In *Proceedings of the 19th Annual International Conference of the IEEE*, volume 4, pages 1436–1439. Engineering in Medicine and Biology Society, November 1997.

Figure 8.2 Prasad D. Polur and Gerald E. Miller. Investigation of an HMM/ANN hybrid structure in pattern recognition application using cepstral analysis of dysarthric (distorted) speech signals. *Medical Engineering and Physics*, 28 (8):741–748, October 2006. Copyright © 2006 Elsevier. Used with permission.

Figure 9.1 Used by permission of the Tetra Society of North America.

Figure 9.3a Courtesy of National Public Website on Assistive Technology.

Figure 9.3b Courtesy of Poule, `https://en.wikipedia.org/wiki/Augmentative_and_alternative_communication#/media/File:Minimo.jpg`.

Figure 9.4 Courtesy of © 2016 Copyright PrAACtical AAC.

Figure 9.5 Courtesy of David J.C. MacKay.