

Remote Speech Technology for Speech Professionals - the CloudCAST Initiative

*Phil Green¹, Ricard Marxer¹, Stuart Cunningham¹, Heidi Christensen¹,
Frank Rudzicz^{2,3}, Maria Yancheva³, André Coy⁴,
Massimiliano Malavasi⁵, and Lorenzo Desideri⁵*

¹University of Sheffield, United Kingdom,

²Toronto Rehabilitation Institute, Canada, ³University of Toronto, Canada,

⁴University of West Indies, Jamaica, ⁵AIAS Onlus Bologna, Italy

Abstract

Clinical applications of speech technology face two challenges. The first is data sparsity. There is little data available to underpin techniques which are based on machine learning and, because it is difficult to collect disordered speech corpora, the only way to address this problem is by pooling what is produced from systems which are already in use. The second is personalisation. This field demands individual solutions, technology which adapts to its user rather than demanding that the user adapt to it. Here we introduce a project, CloudCAST, which addresses these two problems by making remote, adaptive technology available to professionals who work with speech: therapists, educators and clinicians.

Index Terms: assistive technology, clinical applications of speech technology

1. Introduction to CloudCAST

In this working paper, we introduce CloudCAST, a Leverhulme Trust International Network funded from January 2015 for 3 years. The network partners are The University of Sheffield (United Kingdom), AIAS Onlus Bologna (Italy), The University of the West Indies (Jamaica), and the University of Toronto (Canada).

In recent years, there has been significant progress in Clinical Applications of Speech Technology (CAST) in diagnosis of speech disorders [1], tools to correct pronunciation and improve reading skills [2], recognition of disordered speech [3] and voice reconstruction by synthesis [4]. The aim of CloudCAST is to make progress in this domain and to provide a freely-available platform for worldwide collaboration.

We aim to place CAST tools in the hands of professionals who deal with clients with speech and language difficulties, including therapists, pathologists, teachers, and assistive technology experts. We intend to do this by means of a free-of-charge (if possible), remotely-located, internet-based resource ‘in the cloud’ which will provide a set of software tools including personalised speech recognition, diagnosis and interactive spoken language learning. Following a user-centred design methodology, we will provide interfaces which will make these tools easy to use for professionals and their clients, who are not necessarily speech technology experts.

There are various models for user-centred design [5], among which the ISO standard 9241-210 [6] is prominent. This standard for human-centred design processes includes six guiding principles (P): P1. understand the user, the task and environ-

mental requirements; P2. encourage early and active involvement of users; P3. be driven and refined by user-centered evaluation; P4. include iteration of design solutions; P5. address the whole user experience; P6. encourage multi-disciplinary design.

The CloudCAST resources will also facilitate speech data collection necessary to inform the machine learning techniques which underpin this technology: we will be able to automatically collect data from systems which are already in use, as well as provide a database scheme for collecting and hosting databases related to this domain. Our 3-year aim is to create a self-sustaining CloudCAST community to manage future development beyond our current funding period.

While CloudCAST will build on previous work by its partners and others, we believe that it offers several ‘unique selling points’, including:

- The resource will be available worldwide, and free of charge.
- We will provide interfaces, resources and tools targeted at several kinds of users, including:
 - Developers, who want to embed CloudCAST technology into their own applications, for instance voice control of domestic robots,
 - Speech professionals, who want to use CloudCAST technology to work with their clients, for instance, to devise personalised therapy exercise programmes,
 - End users, for whom applications are developed, e.g., children learning to read,
 - Speech technologists, who are improving or adding to the CloudCAST technology itself.
- The technology will be based on open source toolkits such as Kaldi for automatic speech recognition and OpenHab for smart homes [7, 8].
- Subject to ethical constraints, we will collect speech data and metadata from every CloudCAST interaction. All this material will therefore be available for re-training the technology, and for analysis. In this way,
 - we will be able to personalise the technology for each End User,
 - by pooling the data, we will address the problem that for abnormal speech the large datasets needed for speech technology development are not available,
 - we will be able to underpin and evaluate improvements in analysis and classification of speech disorders.

2. Challenges for CloudCAST

CloudCAST's success requires meeting a number of technical, scientific and more general challenges:

- The technology will run remotely, but in many applications it must deliver results rapidly, within a few seconds.
- The technology should improve its performance as it is used, by adaptation to the data it is collecting.
- It will not be possible to control the conditions under which the tools are used to the extent that one might like. For example, diverse recording devices and recording conditions may make normalization challenging.
- There must be shared functionality of tools over applications. For instance, pronunciation tutors and reading tutors have much in common.
- There must be interfaces, and guides to these interfaces, which are suitable for each user-group listed above.
- There must be a scheme which protects the security and privacy of CloudCAST users and their data.
- There is understandable resistance to technology from some speech professionals, based on bad experiences.
- For this reason, and others, the technology must adapt to its user, rather than the other way round.
- There must be a strategy for developing a self-sustaining CloudCAST community.

Our intention is to commence with three exemplar applications: small vocabulary command-and-control with disordered speech, a literacy tutor and a computer aid for therapists. These are described after the next section, in which we introduce the common speech technology resource that will support them.

3. Speech technology resource

Several toolkits exist which provide core speech recognition facilities on which applications can be built, notably Speechmatics [9], Google's Web Speech API [10] and SoundHound [11]. Speechmatics provides a queue-based speech transcription service supporting multiple languages and audio formats, performs automatic punctuation, capitalisation and diarization (speaker separation) and supplies individual word timings and confidences. It's authors claim to achieve near real-time turnaround with very high accuracy. Google, through its proposed Web Speech API, provides both speech recognition and synthesis. The speech recognition service outputs the results in the form of multiple hypotheses of word-level transcriptions with associated confidence scores. SoundHound provides a speech-to-meaning service that performs simultaneous speech recognition and natural language understanding. This process outputs its results in the form of structured commands instead of plain text transcription.

For CloudCAST, these solutions fall short in terms of the types and details of the results they return, the flexibility of the recognition process, provisions for customisation of the speech models, and modes of interaction. The maximum level of detail provided in all these solutions is an N -best list of word-level transcriptions with associated confidences. In the case of Speechmatics, word-level time alignments are also available. However higher-level details such as phone time alignments are not accessible. Furthermore, other types of results such as decoding lattices and word confusion networks (WCN) [12] are not provided. The grammars (or language models) used in

these systems are fixed to general-domain dictation applications (in Google's Web Speech API, the introduction of a grammar specification function was discussed in 2012, but to the best of our knowledge it has not been concretized or implemented in Chrome). While Google's service does provide an interactive mode in which partial results of the decoding process are immediately available, this is not the case with the service provided by Speechmatics. None of the services provide any means of creating custom models using specific training material. This precludes targeting disordered speech or other niche cases.

The requirements of CloudCAST include providing an interactive speech recognition service where the client must be able to modify the grammar, the model, and other relevant parameters. The client should have instant feedback about the recognition process, such as partial decoding as well as access to fully detailed results such as phone-level alignments and posterior probabilities. Crucially, interactions of clients with CloudCAST should provide data resources to improve the recognition process and the training of future models.

The main architecture of CloudCAST (Figure 1) can be split into the exemplars, the frontend, and the backend. The exemplars are services using CloudCAST, for instance, webapps that perform literacy tutoring or command-and-control (see next section). The frontend is the visible CloudCAST website, from which users can manage their recordings, developers can obtain API keys, professionals can create models, and so on. Finally, the backend is the server which consumes audio from the exemplars and provides speech recognition results. The backend is also in charge of applying the parameter changes that the exemplars may request to the recognition process.

Both the frontend and the backend have access to a common storage space and database where they store models, recordings, and authentication details. The frontend and backend are both backed by worker processes, whose roles are to perform computationally intensive tasks, such as the training of the models and actual speech recognition, which may be run in separate devices. This split ensures the scalability of the system.

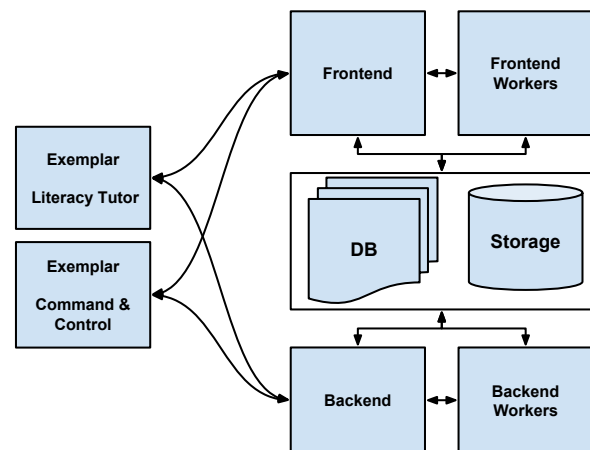


Figure 1: Architecture of the CloudCAST resource.

CloudCAST is developed with open source software. We have decided to create the frontend using Flask, a free software microframework for web development.

To implement the speech recognition service (the backend) we have decided to build on kaldi-gstreamer-server de-

veloped by Tanel Alume [13]. Kaldi-gstreamer-server is a distributed online speech-to-text system that features real-time speech recognition and a full-duplex user experience where the partially transcribed utterance is provided to the client. The system includes a simple client-server communication protocol and scalability to many concurrent sessions. The system is open source and based on free software and therefore serves as a starting point for building CloudCAST, allowing us to deploy recognisers developed at Sheffield within the CloudCAST framework [14]. It uses Kaldi [7] for speech recognition processing. Kaldi is a well-known free software library widely used in the research community partly due to its modular and flexible architecture.

To facilitate the creation of services using CloudCAST, we are also developing a speech recognition client in JavaScript based on the existing library `dictate.js`. The proposed client extends `dictate.js` with multiple types of interactions with the server, such as swapping grammars, models and other parameters, as well as interpreting the different results provided by the server.

4. Exemplars

4.1. Literacy tutor

Among the first set of exemplars to be developed will be an automated literacy tutor. In some respects the literacy tutor represents the most complex type of application that can be developed using the tools CloudCAST will make available. In addition to being a good showcase for the tools, the literacy tutor can be useful in bolstering current efforts to combat illiteracy as it will be a freely available, cloud-based resource that can be modified to meet the needs of individual users.

It has been shown that the use of speech-enabled literacy tutors can lead to significant improvement in their users ability to read [15, 16, 17]. Among the best known systems is Project LISTEN [18]. Project LISTEN, developed at Carnegie Mellon University, works as a tutor by listening to, as well as reading to the user. This system was one of the first to employ feedback that was able to effectively respond to readers when they encountered challenges or made mistakes. When the system was deployed in schools it was found that students using the reading tutor outperformed their peers who learned from regular classroom-based activities and compared favourably to students given one-to-one tutoring by human experts. Outside of the United States, the effectiveness of Project LISTEN to provide tutorial support for language learners has been tested (on very limited scales) in a number of countries, including Canada, Ghana and India [19, 20, 21]. Users were shown to make significant progress in literacy skills when they used the tutor.

Project LISTEN is available for research purposes, but it is not a commercial product; it is not openly accessible, nor is it cloud-based. One commercial, web-based reading tutor is the Reading Trainer component of the Ridinet online network. Ridinet is meant to provide practice and training in literacy and numeracy for Italian children diagnosed with autism spectrum disorder [22]. Reading Trainer was included in the system for the purpose of increasing reading fluency. The initial phase of the Reading Trainer is a speed test, where users are prompted to read a previously unseen text and the time it takes them to read it is used as their initial reading speed. The tool is customisable; it allows the user to select, among other parameters, reading speed, reading accuracy, reading unit and story length. The level of feedback can also be set to either prompt the user or praise

their performance and effort.

The basic functionality of the CloudCAST literacy tutor exemplar will be similar to the two tutors described above, but will differ in at least three important respects. Firstly, it will be freely available to anyone with an internet connection. Secondly, the tool will be further customisable: the user will be able to change language and upload new stories. Finally, the tutor will have an integrated reading age assessment tool to determine the reading level of the user and to act as a pre-test for potential learning challenges. The results of assessment and user performance for each session will be securely stored online for easy tracking of their progress.

4.2. Environmental control

The command and control exemplar will provide a service that will allow, for instance, manipulation of multiple devices in a smart home either directly with speech commands or through voice communication with assistive robots.

Current home automation systems and the increasingly popular Internet of Things (IoT) can provide great support to people with disabilities by improving their autonomy and safety in daily living activities.

There are several ways in which CloudCAST will improve on existing speech recognition solutions. Although user interfaces based on voice interaction are particularly suited for this type of application, current systems devised for assistive technology or for the mainstream market are unsuitable, in terms of performance, for many potential users. Common limitations are the inability to be completely hands-free and poor recognition performance.

Command and control systems are particularly useful for subjects with mobility issues. In many cases these people also experience speech disorders for which available speech recognition systems are not optimized. The possibility of using personalised speech models could greatly enhance the recognition accuracy and therefore the reliability of the system. Furthermore the speech material produced by such users will be of great value to improve future speech models for other users with similar issues.

The ability to define a customised grammar will render the system significantly more robust to speech disfluencies, environment noise and recognition ambiguity. Keeping grammars simple, with few word options (low perplexity) at each stage of the control sequence will make the system less prone to recognition errors.

Since many actual home automation fieldbuses can be easily connected to the internet and IoT devices are natively equipped with this property, cloud services developed within the CloudCAST project can be easily implemented, and will be flexible and customizable. The potential of the exemplars can be extended through the use of specific open source servers dedicated to the integration of home automation technologies and IoT solutions, such as Openhab [8].

4.3. Speech therapy

The ability to communicate is one of the most basic human needs. Many lose the ability to communicate, due to a range of health conditions which result in a speech impairment. Speech therapy helps improve communication ability and produces benefits in terms of quality of life and participation in society. Articulation therapy aims to improve the speech of people with speech impairment. It is however time-consuming, and patients

rarely receive sufficient therapy to maximise their communication potential [23, 24].

In articulation therapy speech therapists work with patients on the production of specific speech sounds and provide feedback on the quality of these speech sounds. This process helps the patient improve their production of these sounds thereby improving the overall intelligibility of speech. Our previous research shows that computer programs using speech recognition can improve outcomes of speech therapy for adults with speech difficulties [25, 26]

For our CloudCAST exemplar we intend to build on our past work to develop a web-based application. This demonstrator will enable therapists and clients to work together to specify speech exercises. These exercises could then be independently completed by the client between therapy sessions.

A big advantage of using technology over traditional practice will be that therapists can monitor and review the progress that their client has made. During the completion of the exercises, the speech produced by the client will be scored and then stored for review. This means that any difficulties that they encounter during the exercises can be identified and discussed with the therapist.

We have previously developed techniques for using automatic speech recognition to provide feedback to patients practising their speech [25, 27]. These approaches are based on using specially developed speech recognition software able to provide objective feedback, which acts as a substitute for the judgement of an expert listener, such as the speech and language therapist. This feedback can be given to patients when they are practising either with a therapist or on their own between therapy sessions [26]. We will use especially adapted recognisers available via the CloudCAST platform to generate this objective feedback. We will then make these approaches available in a range of motivational exercises.

5. Data collection and repository

CloudCAST will also serve as a data repository for the distribution of existing databases and for the acquisition of new databases, along with provided tools for that collection. Below we discuss the first database that will become freely available in CloudCAST, TORGO, and the database scheme we will use to represent future data collection

5.1. TORGO

TORGO consists of aligned acoustic and EMA measurements from individuals with and without cerebral palsy (CP), each of whom recorded 3 hours of data [28]. CP is one of the most prevalent causes of speech disorder, and is caused by disruptions in the neuro-motor interface [29] that do not affect comprehension of language, but distort motor commands to the speech articulators, resulting in relative unintelligibility [30]. The motor functions of each participant in TORGO were assessed according to the standardized Frenchay Dysarthria Assessment [31] by a speech-language pathologist affiliated with the Holland-Bloorview Kids Rehab hospital and the University of Toronto. Individual prompts were derived from non-words (e.g., /iy-p-ah/ [32]), short words (e.g., contrasting pairs from [33]), and restricted sentences (e.g., the sentence intelligibility section in the Yorkston-Beukelman Assessment [34], and sentences from MOCHA-TIMIT.

The EMA data in TORGO were collected using the three-dimensional Carstens Medizinelektronik AG500 system [35,

36]. Sensors were attached to three points on the surface of the tongue, namely tongue tip (TT – 1 cm behind the anatomical tongue tip), the tongue middle (TM – 3 cm behind the tongue tip coil), and tongue back (TB, approximately 2 cm behind the tongue middle coil). A sensor for tracking jaw movements (JA) was attached to a custom mould over the lower incisors [37]. Four additional coils were placed on the upper and lower lips (UL and LL) and the left and right corners of the mouth (LM and RM). Reference coils were placed on the subject's forehead, nose bridge, and behind each ear above the mastoid bone.

5.2. Future database scheme

New users of CloudCAST can immediately use our database framework for representing the data. To a large extent, this framework is designed to be generic to all speech recording tasks, and not all components need to be utilized. The database schema is broken down into three core sections: the subject, the task, and the session. A high-level overview of the data representation is shown in Figure 2.

The subject section generally involves aspects related to the speaker, including demographics, levels of permission to use the data, and factors affecting the subject's language quality, such as country of origin, country of residence, spoken languages, history of smoking, and education level. The task section specifies the language task (e.g., picture description, conversation, reading of text, repetition of audio) along with a bank of available task instances (e.g., pictures to be used in the picture description task). The system supports a variety of question and answer types, including text, speech, multiple-choice, and fill-in-the-blank, with the ability for easy extension to new types. Each task instance is optionally rated with a level of difficulty, measured across arbitrary dimensions (e.g., phonological complexity, syntactic complexity). Information related to automatic scoring of tasks is stored along with each task instance, where appropriate (e.g., the correct answer to a multiple-choice question). Each subject can be associated with a number of recording sessions, and each session can be associated with a number of task instances. The session section stores the subject responses to specific task instances every time they interact with the system. This includes their language data, as well as meta-data such as total amount of time spent on each task, and date of completion.

This database is designed to be extensible to future needs, and will be especially useful to streamline data organization to projects that otherwise have a more clinical focus. It enables (i) longitudinal subject assessments, due to the ability to accommodate multiple language task instances in order to avoid 'the learning effect' over time, (ii) dynamic variation of task instance difficulty and type based on subject performance, and (iii) automated scoring of subject performance where appropriate.

5.3. Ethics

As part of the CloudCAST initiative we will be seeking to collect speech data from individual participants. To do so we must ensure that we fully respect their personal data. As part of this process professionals who initiate a service through CloudCAST will need to first confirm that they are abiding by the local ethics and governance rules.

For individual participants making use of CloudCAST services we will follow a process approved by the University of Sheffield Research Ethics Committee. It is proposed that as part of this process we will first fully explain to each individual user when they register with CloudCAST the background to the

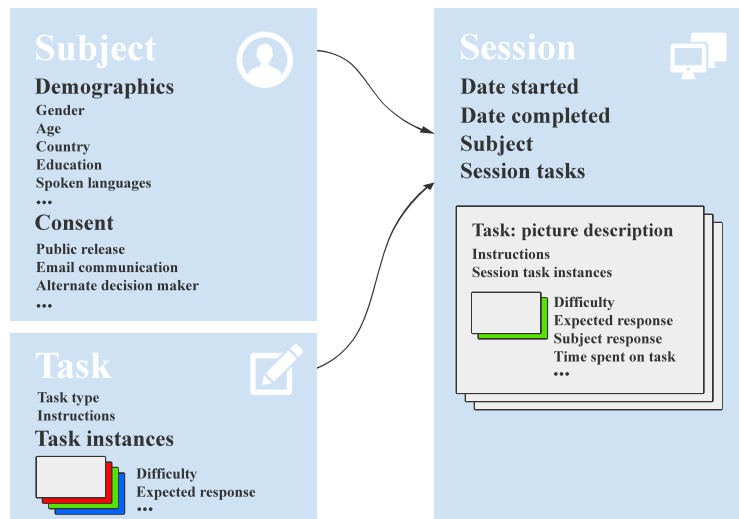


Figure 2: Simplified database schema, arranged into three core sections: Subject, Session, and Task.

project and how we intend to use their speech data. Participants will be able to opt-in to different levels of engagement with the CloudCAST initiative. At the most basic level, a participant will be able to make use of the CloudCAST services without their data being used for further research, or shared with other researchers. The second level of participation can be selected by the participant when they wish to allow the CloudCAST team to retain their data for further research. The final level of participation can be selected by participants when they wish their data to be retained and potentially distributed to other speech researchers.

As part of the on-going relationship with the participants, they will periodically be asked to re-confirm their consent for their data to be used in the way they chose.

6. Conclusions

CloudCAST aims to create a self-sustaining community of academic and speech professionals which will continue to grow after its 3 year funding period. It is our belief that only by collaborating in this way can we make the benefits of speech technology available to those who need it most and at the same time create the knowledge bases for further technical improvement. To attain critical mass we need to widen the participants beyond the initial partners. If you are interested, please contact us by registering on our website: <http://cloudcast.rweb.dcs.shef.ac.uk/>

7. Acknowledgements

CloudCAST is an International Network (IN-2014-003) funded by the Leverhulme Trust.

8. References

- [1] "PEAKS a system for the automatic evaluation of voice and speech disorders," *Speech Communication*, vol. 51, no. 5, pp. 425–437, 2009.
- [2] O. Saz, S.-C. Yin, E. Lleida, R. Rose, C. Vaquero, and W. R. Rodriguez, "Tools and technologies for computer-aided speech and language therapy," *Speech Communication*, vol. 51, no. 10, pp. 948–967, 2009.
- [3] H. Christensen, S. Cunningham, C. Fox, P. Green, and T. Hain, "A comparative study of adaptive, automatic recognition of disordered speech," in *Proc Interspeech 2012*, Portland, Oregon, US, Sep 2012.
- [4] C. Veaux, J. Yamagishi, and S. King, "Towards personalized synthesized voices for individuals with vocal disabilities: Voice banking and reconstruction," in *Proceedings of 4th Workshop on Speech and Language Processing for Assistive Technologies, SLPAT2013*, 2013, pp. 107–111.
- [5] S. Blackburn and P. Cudd, "An overview of user requirements specification in ICT product design," in *Proceedings of the AAATE workshop: The social model for AT Technology Transfer*, Sheffield, UK, 2010.
- [6] "Iso 9241-210: 2009. ergonomics of human system interaction - part 210: Human-centred design for interactive systems," Switzerland, 2009.
- [7] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, Dec. 2011, iEEE Catalog No.: CFP11SRW-USB.
- [8] "openHAB." [Online]. Available: <http://www.openhab.org>
- [9] "Speechmatics." [Online]. Available: <https://www.speechmatics.com/>
- [10] G. Shires and H. Wennborg, "Web speech api specification," W3C, Community Group Final Report, Oct. 2012, <https://dvc.w3.org/hg/speech-api/raw-file/9a0075d25326/speechapi.html>.
- [11] "Soundhound - houndify platform." [Online]. Available: <http://www.soundhound.com/houndify>
- [12] D. Hakkani-Tür, F. Béchet, G. Riccardi, and G. Tur, "Beyond asr 1-best: Using word confusion networks in spoken language understanding," *Computer Speech & Language*, vol. 20, no. 4, pp. 495–514, 2006.
- [13] T. Alumäe, "Full-duplex speech-to-text system for Estonian," Kaunas, Lithuania, 2014.

- [14] H. Christensen, I. Casanueva, S. Cunningham, P. Green, and T. Hain, "Automatic selection of speakers for improved acoustic modelling : Recognition of disordered speech with sparse data," in *Spoken Language Technology Workshop, SLT'14*, Lake Tahoe, Dec 2014.
- [15] J. Mostow, G. Aist, P. Burkhead, A. Corbett, A. Cuneo, S. Eitelman, C. Huang, B. Junker, M. B. Sklar, and B. Tobin, "Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction," *Journal of Educational Computing Research*, vol. 29, pp. 61–117, 2003.
- [16] M. J. Adams, "The promise of automatic speech recognition for fostering literacy growth in children and adults," in *Handbook of Literacy and Technology*, M. McKenna, L. Labbo, R. Kieffer, and D. Reinking, Eds. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 2005, vol. 2, pp. 109–128.
- [17] B. Wise, R. Cole, S. van Vuuren, S. Schwartz, L. Snyder, N. Ngampatipatong, J. Tuantranont, and B. Pellom, "The promise of automatic speech recognition for fostering literacy growth in children and adults," in *Interactive literacy education: Facilitating literacy environments through technology*, C. Kinzer and L. Verhoeven, Eds. Mahwah, New Jersey: Lawrence Erlbaum, 2005, vol. 2, pp. 31–76.
- [18] J. Mostow, S. Roth, A. Hauptmann, and M. Kane, "A prototype reading coach that listens," in *Association for the Advancement of Artificial Intelligence, AAAI-94*, Seattle, Washington, 1994.
- [19] T. Cunningham, "The effect of reading remediation software on the language and literacy skill development of ESL students," Master's thesis, University of Toronto, Toronto, Canada, 2006.
- [20] G. Korsah, J. Mostow, M. Dias, T. Sweet, S. Belousov, M. Dias, and H. Gong, "Improving child literacy in africa: Experiments with an automated reading tutor," *Information Technologies and International Development*, vol. 6, pp. 1–19, 2010.
- [21] F. Weber and K. Bali, "Enhancing esl education in india with a reading tutor that listens," in *Proceedings of First ACM Symposium on Computing for Development ACM*, vol. 20, London, UK, 2010, pp. 1–9.
- [22] S. Pinnelli, "Dyslexia and young adults. A case study: from assessment to intervention with reading trainer software," in *Proceedings of SIREM 2013*, Bari, Italy, 2014, pp. 84–94.
- [23] J. Law, Z. Garrett, and C. Nye, "Speech and language therapy interventions for children with primary speech and language delay or disorder," *Cochrane Database of Systematic Reviews*, no. 3, 2003.
- [24] P. Enderby and L. Emerson, *Does Speech and Language Therapy Work?* London: Singular, 1995.
- [25] R. Palmer, P. Enderby, and S. P. Cunningham, "Effect of three practice conditions on the consistency of chronic dysarthric speech," *Journal of Medical Speech-Language Pathology*, vol. 12, no. 4, pp. 183–188, 2004.
- [26] R. Palmer, P. Enderby, and M. Hawley, "Addressing the needs of speakers with longstanding dysarthria: computerized and traditional therapy compared," *International journal of language & communication disorders / Royal College of Speech & Language Therapists*, vol. 42 Suppl 1, pp. 61–79, Mar. 2007.
- [27] M. Parker, S. P. Cunningham, P. Enderby, M. S. Hawley, and P. D. Green, "Automatic speech recognition and training for severely dysarthric users of assistive technology: the STARDUST project," *Clinical Linguistics & Phonetics*, vol. 20, no. 2-3, pp. 149–156, 2006.
- [28] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria," *Language Resources and Evaluation*, vol. 46, no. 4, pp. 523–541, 2012.
- [29] R. D. Kent and K. Rosen, "Motor control perspectives on motor speech disorders," in *Speech Motor Control in Normal and Disordered Speech*, B. Maassen, R. Kent, H. Peters, P. V. Lieshout, and W. Hulstijn, Eds. Oxford: Oxford University Press, 2004, ch. 12, pp. 285–311.
- [30] R. D. Kent, "Research on speech motor control and its disorders: a review and prospective," *Journal of Communication Disorders*, vol. 33, no. 5, pp. 391–428, 2000.
- [31] P. M. Enderby, *Frenchay Dysarthria Assessment*. College Hill Press, 1983.
- [32] J. W. Bennett, P. Van Lieshout, and C. M. Steele, "Tongue control for speech and swallowing in healthy younger and older subjects," *International Journal of Orofacial Myology*, vol. 33, pp. 5–18, 2007.
- [33] R. D. Kent, G. Weismer, J. F. Kent, and J. C. Rosenbek, "Toward phonetic intelligibility testing in dysarthria," *Journal of Speech and Hearing Disorders*, vol. 54, pp. 482–499, 1989.
- [34] K. M. Yorkston and D. R. Beukelman, *Assessment of Intelligibility of Dysarthric Speech*. Tigard, Oregon: C.C. Publications Inc., 1981.
- [35] A. Zierdt, P. Hoole, and H. G. Tillmann, "Development of a system for three-dimensional fleshpoint measurement of speech movements," in *Proceedings of the 14th International Conference of Phonetic Sciences (ICPhS'99)*, San Francisco, USA, August 1999.
- [36] P. Van Lieshout, G. Merrick, and L. Goldstein, "An articulatory phonology perspective on rhotic articulation problems: A descriptive case study," *Asia Pacific Journal of Speech, Language, and Hearing*, vol. 11, no. 4, pp. 283–303, 2008.
- [37] P. H. Van Lieshout and W. Moussa, "The assessment of speech motor behavior using electromagnetic articulography," *The Phonetician*, vol. 81, pp. 9–22, 2000.