

Interactive Media Research & Speech Recognition: usable transcriptions of webcast lectures

Cosmin Munteanu

Computer Science Department, University of Toronto



Our research goal: to make archives of webcast lectures & presentations the digital equivalent of traditional libraries

Focus: improve the usefulness and usability of automatically-generated transcripts of webcast lectures and presentations

Research themes:

- measuring user needs for transcription accuracy
- engaging users to collaborate (wiki) on editing imperfect computer-generated transcripts
- better transcripts for lectures by using external knowledge sources (WWW, lecture slides)
- using the partial wiki corrections to improve the automatic speech recognition
- developing a user-motivated measure of transcript quality
- developing computationally-feasible higher-level textual representation of lectures
- establishing collaborative models for enhanced usability of interactive media archives

The screenshot shows a webcast player interface. At the top left is a logo for 'University of Toronto' and 'Powered by ePresence Interactive Media'. Below it is a video player showing a lecturer. A 'Play/Pause' button is visible. To the right of the video is a table of contents with 20 items, including 'Mental Models' at item 14. Below the table of contents are 'Select Chapter:' and 'Select Slide:' dropdown menus. On the right side of the slide, there are navigation buttons: 'next', 'previous', 'last slide', and 'first slide SLIDE 13 of 27'. The slide content is as follows:

Mental Models

- Definition of mental models (Carroll, 1984):
 - "...structures and processes imputed to a person's mind in order to account for that person's behaviour and experience."
- More generally (Carroll & Olson, 1988):
 - "...all of what a user knows about using a particular piece of software, including *how to use it*, and how it works."
- Mental models allow a user:
 - to understand a system
 - to predict effects of actions
 - to interpret the results
- Role of mental model: to answer questions like:
 - What is X?
 - What happens when you do Y?
 - Why is happening when I see Z?

Design of Interactive Computational Media ©1992-2004, Ronald M. Baecker Fall 2004 Slide 5.16

of of metaphor
so i start
now entity ban an all models
kahn
their couple of
fairly abstract f. issues the mental models
on jack hero back in nineteen eighty four
sad
structure sin process c.'s imputed for person's my
he nor door top for that person's be here
an experts
ok what is
real wet this reminds is of
is the we real we really and some level have now half so notion of what seen
peoples minds
am it our first a look into some
some these nine
radio in crashes distinctly can report army
yes i understand that yes i can see
that
this desktop is organized in a guy files and folders on the
nineteen track something into the trash in of i empty the trash

Approach:

combine research in Human-Computer Interaction and Automatic Speech Recognition to integrate automated transcripts of lectures into webcast archives.

Research team:

- Cosmin Munteanu, www.cs.toronto.edu/~mcosmin
- Ron Baecker, www.kmdi.utoronto.ca/rmb
- Gerald Penn, www.cs.toronto.edu/~gpenn
- For references and further information, please see the back page

Selected publications:

- Munteanu, C., Baecker, R., Penn, G. **Collaborative Editing for Improved Usefulness and Usability of Transcript-Enhanced Webcasts.** In Proceedings of the *ACM SIGCHI Conference on Human Factors in Computing Systems - CHI*, Florence, Italy, 2008
- Munteanu, C., Baecker, R., Penn, G. **Usable speech recognition: toward improved access to webcast lectures.** In Proceedings of the *ACM SIGCHI Conference on Human Factors in Computing Systems - CHI - Workshop on Usable Artificial Intelligence*, Florence, Italy, 2008
- Munteanu, C., Penn, G., Baecker, R. **Web-Based Language Modelling for Automatic Lecture Transcription.** In Proceedings of the *Tenth European Conference on Speech Communication and Technology - EuroSpeech / INTERSPEECH*, Antwerp, Belgium, 2007.
- Munteanu, C., Penn, G., Baecker, R., Zhang, Y. **Automatic Speech Recognition for Webcasts: How Good is Good Enough and What to Do When it Isn't.** In Proceedings of the *Eight International Conference on Multimodal Interfaces - ICMI*, Banff, Canada, 2006
- Munteanu, C., Zhang, Y., Baecker, R., Penn, G. **Wiki-like Editing of Imperfect Computer-Generated Webcast Transcripts.** In the Demo track of the *20th ACM Conference on Computer Supported Cooperative Work - CSCW*, Banff, Canada, 2006.
- Munteanu, C., Baecker, R., Penn, G., Toms, E., James, D. **Measuring the Acceptable Word Error Rate of Machine-Generated Webcast Transcripts.** In Proceedings of the *Ninth International Conference on Spoken Language Processing - INTERSPEECH*, Pittsburgh, USA, 2006
- Munteanu, C., Baecker, R., Penn, G., Toms, E., and James, D. **The Effect of Speech Recognition Accuracy Rates on the Usefulness and Usability of Webcast Archives.** In Proceedings of the *ACM SIGCHI Conference on Human Factors in Computing Systems - CHI*, Montreal, Canada, 2006 (Best of CHI Nominee)

Further information:

- **Cosmin Munteanu** is a Ph.D. candidate in the Department of Computer Science, University of Toronto, primarily interested in the areas of Human Computer Interaction and Automatic Speech Recognition. His work is focused on interactive interfaces and automatic speech recognition techniques for transforming collections of webcast lectures & presentations into the digital equivalent of traditional libraries.. For more information, contact Cosmin at mcosmin@cs.toronto.edu or visit www.cs.toronto.edu/~mcosmin
- This research is supported by the NSERC Canada Network for Effective Collaboration Technologies through Advanced Research (NECTAR).
- ePresence Interactive media is the world's first Open Source Software for Interactive Webcasting and Conferencing, www.epresence.tv, and serves as a testbed for this research.

