Speech Technologies

New Media & Language Technologies
Jozef Stefan International Postgraduate School
November 2005

Jerneja Žganec Gros
jerneja@alpineon.com
Lecture structure

- Speech technologies
  - research topics, market drivers
  - useful links: information, organizations, resources
  - speech technologies in Slovenia: groups, resources, apps
- How to build a TTS system?
  - principles of TTS design
  - designing and annotating a speech corpus
- Speech technology applications
  - platforms
  - sample applications
  - demos
Human speech

Spoken language interaction:

- plays a fundamental role in almost all human activities

- from simple gossip -> commercial transactions, sensitive business negotiations…

- a highly evolved social communication system capable of linking two or more people in a joint endeavour
Human speech

“Spoken language processing is the most sophisticated behaviour of the most complex organism in the known universe.”

(Dawkins, 1991)
Speech technologies - topics

- **speech synthesis**
  - Demo of AT&T TTS system
  - The Bell Labs Text-to-Speech system, with a demo.
  - Elan text to speech

- **speech recognition**
  - Tifaq Speech Recognition FAQ
  - Voice Recognition Forums
  - 21st Century Eloquence Speech Recognition Specialists
  - COST249 ASR demo's for various languages
  - How can I build a simple speech recogniser?
  - MDT
Speech technologies - areas

- speaker verification (biometrics, security)
  - Speaker recognition demo of CAVE (CAller VErification)
  - Keyware's Voice Verification (VOICEGUARDIAN)
  - Speaker recognition demo of CAVE (CAller VErification)
- spoken dialogue systems
- speech-to-speech translation
- speech prosody: emotional speech
- audio-visual speech (broadcast news, talking heads)
- multimodal applications
- multilinguality
Speech technologies - markets

- automotive
- broadcast
- consumer
- defense
- disabled
- education
- legal
- medical
- telephony...
Future market drivers

FP6 IST Call, ‘ambient intelligence’:

‘computers and networks will be integrated into everyday environment, rendering accessible a multitude of services and applications through easy-to-use interfaces’
Community opinion

- ASRU 1997 and 2003 surveys

Figure 4: Responses to the statement “Most telephone Interactive Voice Response systems accept speech input (and more than just digits)”.

Source: Moore, Interspeech 2005
Community opinion

- ASRU 1997 and 2003 surveys

Figure 6: Responses to the statement “Voice recognition is commonly available at home (e.g. interactive TV, control of home appliances and home management systems).”

Source: Moore, Interspeech 2005
Community opinion

- ASRU 2003 survey

*Figure 10: Responses to the statement “Translating telephones allow two people across the globe to speak to each other even if they do not speak the same language.”*
Community opinion

- ASRU 2003 survey

*Figure 9: Responses to the statement “Most routine business transactions take place between a human and a virtual personality (including an animated visual presence that looks like a human face)”.

Source: Moore, Interspeech 2005
Why never?

‘The industry has yet to bridge the gap’ between what people want and what it can deliver. Reducing the \textit{ASR error rate} remains the greatest challenge.’

X.D. Huang, MS, 2002

‘After 60 yrs of intensive R&D in TTS, our gadgets, toys and appliances still do not speak to us intelligently.’

C. Henton, 2002

\textit{Source:} Moore, Interspeech 2005
Useful links

- ELSNET [http://www.elsnet.org]
- EUROMAP, HLTCentral [http://www.hltcentral.org/]
- ISCA [http://www.isca-speech.org/]
- ISCA Special Interest Groups (SIG):
  - Speech Synthesis - SynSig
  - Audio Visual Speech - AVISA
  - Speech And Language Technology for MInority Languages - SALTNIL
  - Integration of Speech Technology in (Language) Learning - InSTIL
  - SPEaker and Language Characterization - SpLC
  - Speech Prosody - SProSIG
  - Dialogue Processing - SigDial
Info, publications

- free speech journal
- Computer Speech & Language
- International Journal of Speech Technology
- Language Learning & Technology
- Phonetica, International Journal of Speech Science
- Spoken Language Input, by Ron Cole & Victor Zue
- Speech Analysis, by Tony Robinson
- EAGLES database of terminology
Useful links - Slovenia

- The Slovenian Language Technologies Society
  http://nl.ijs.si/sdjt/index-en.html

  HLT in Slovenia
  HLT-related events in Slovenia
  SDJT members
  SDJT mailing list
Speech groups - Slovenia

Laboratorij za arhitekturo in procesiranje signalov
FRI, University of Ljubljana

Laboratorij za digitalno procesiranje signalov
FERI, University of Maribor

Laboratorij za umetno zaznavanje, sisteme in kibernetiko
FE, University of Ljubljana

Oddelek za prevajanje in tolmačenje
FF, University of Ljubljana

Odsek za inteligentne sisteme, IJS Ljubljana

Alpineon, d.o.o., Ljubljana

Amebis d.o.o., Kamnik

Hermes Softlab, Ljubljana
Spoken language resources

- LDC [http://www.ldc.upenn.edu/](http://www.ldc.upenn.edu/)

**Phonetic transcription** – phonetic alphabet:
- IPA [http://www2.arts.gla.ac.uk/IPA/ipa.html](http://www2.arts.gla.ac.uk/IPA/ipa.html)
- SAMPA: machine-readable phonetic symbols [http://www.phon.ucl.ac.uk/home/sampa/slovenian.htm](http://www.phon.ucl.ac.uk/home/sampa/slovenian.htm)
Slovenian speech corpora

**FERI, University of Maribor**
- SNABI (Kačič 2002), *SpeechDat II* (Kačič 2002), PoliDat (Kačič 2002) in emotional speech corpus (Hozjan 2002), broadcast news corpus (Žgank 2004), LC-STAR (Verdonik 2004)

**FE, University of Ljubljana**
- **Mobiluz** (Dobrišek 1998, Gros 2000, Mihelič 2003), **K211D** (Dobrišek 2001), weather forecasts VNTV and VNRAD (Žibert 2000), broadcast news (Žibert 2004) and 4 specialized speech corpora: diphone corpus (Gros 2000), **VINDAT** (Škrlj 2001), multi speech rate corpus (Gros 2000), speaker ID corpus (Kranjc 2001)
Slovenian speech corpora

Other Slovenian speech corpora:

- FRI: Števke (Rozman 2000)
- IJS: diphone corpus (Šef 2001), MULTEXT-East (Erjavec 1998)
- Masterpoint: diphone corpus (Mihelič 2002)
- Hermes Softlab (Šket 2002)
- Center za slovenski jezik (Zemljarič, 2004): sample spontaneous speech corpus
Slovenia: speech applications

- TTS system Govorec
- HOMER: voice-driven web browser for the blind
- M Vstopnica, Mobitel: IVR application (ASR)
- Glasovna pošta, Mobitel: e-mail reader (TTS)
- Glasovni SMS, Mobitel: SMS2Voice (TTS)
- Letinfo, FERI: IVR application (ASR)