Phonetic knowledge in speech technology
Joaquim Llistrerri
Grup de Fonètica
Departament de Filologia Espanyola
Universitat Autònoma de Barcelona
http://liceu.uab.cat/~joaquim
Phonetic knowledge in speech technology

✔ The crucial question
✔ Early day vs. present day approaches
✔ Some obstacles to the dialogue
✔ Some consequences of the lack of dialogue
✔ Towards integration
Phonetic knowledge in speech technology

✔ The crucial question
✔ Early day vs. present day approaches
✔ Some obstacles to the dialogue
✔ Some consequences of the lack of dialogue
✔ Towards integration
Phonetic knowledge in speech technology

• The crucial question
  • Human expert knowledge-based approach: phonetics
  • Machine-learning statistical approach: engineering

• Is a combination of both approaches possible?
“Can we leave it to the computer to learn about speech or shall we insist on developing our own insights in the many dimensions of the speech code?”

(Fant, 1983: 17)
Phonetic knowledge in speech technology

✔ The crucial question
✔ Early day vs. present day approaches
✔ Some obstacles to the dialogue
✔ Some consequences of the lack of dialogue
✔ Towards integration
Phonetic knowledge in speech technology

- Early day approaches
  - Based on phonetic knowledge
  - Knowledge formalised as rules
- The phonetician as a provider of knowledge about the speech code
The Pattern Playback
Haskins Laboratories

Four hours of steady work faced us.

http://www.haskins.yale.edu/featured/patplay.html

Joaquim Llistèrri, Grup de Fonètica, Departament de Filologia Espanyola
MITalk
Allen, Hunnicut & Klatt (1979)

http://www.speech.kth.se/~scheri/
http://www.ling.su.se/fon/phoneticians/Gubbar.html
“in the last decades we have witnessed a decrease in the amount of phonetic knowledge used in ASR and TTS”

(Strik, 2005: 168)
“the linguistic approach soon lost terrain, in recognition applications at least, to (nonlinguistically oriented) engineers who were less concerned with formal linguistic insights, treating the signal as a pattern just like any other, and this with outstanding success”

(Barry et al., 2005: 1)
“Data-driven approaches overcame the problem of the extraction of manual rules by employing either statistical methods or artificial neural network (ANN) based techniques which automatically produce phonetic rules and construct [...] models from large speech corpora. Their main advantage is that this process is automatic and thus does not need contribution from linguists.”

(Submitted paper, 2009)
Phonetic knowledge in speech technology

• A “corpus-centric” perspective on spoken language

“Manually annotated material forms the basis for statistical characterization of speech, as well as for training systems to perform automatic labeling for speech recognition. Currently, most manual annotation focuses on the lexical level and seeks to derive labels and segmentation for the lower tiers (particularly segments) via automatic methods (…)” (Greenberg, 2005: 111)
Phonetic knowledge in speech technology

- Present day approaches
  - Based on information automatically acquired from corpora: machine learning
  - Knowledge inside the corpus
  - The phonetician as a manual labeller of corpora
Phonetic knowledge in speech technology

✔ The crucial question
✔ Early day vs. present day approaches
✔ Some obstacles to the dialogue
✔ Some consequences of the lack of dialogue
✔ Towards integration
Phonetic knowledge in speech technology

- Some obstacles to the dialogue between linguistics (phonetics) and technology (engineering)
  - The nature of current phonetic knowledge
  - The abstract nature of linguistic models
  - Lack of interdisciplinary training
Some obstacles to the dialogue

- The nature of current phonetic knowledge
  - Lack of detail
  - Lack of quantification
  - Lack of formalisation

  “phonetics does not provide ready-made quantitative models that can be plugged directly into a system”

  (Strik, 2005: 177)
Some obstacles to the dialogue

• The nature of current phonetic knowledge
  • Mostly based on “laboratory speech” due to the need to control variables in an experimental setting
  • Availability of large corpora of spontaneous speech recorded in naturalistic or application related environments?
Some obstacles to the dialogue

• The abstract nature of linguistic models

“Phonological systems like the ToBI approach only introduce a quantisation error: the whole variety of $F_0$ values available in acoustics is reduced to a mere binary opposition L vs. H, and to some few additional, diacritic distinctions”

(Batliner & Möbius, 2005: 26)
Some obstacles to the dialogue

- Lack of interdisciplinary training
  “a successful phonetician working on a spoken language system will need some knowledge of computers, algorithms, statistics and signal processing” (Acero, 1995: 175)

- The “linguist speech-technology engineer” and the “speech technology linguist” (Barry et al., 2005: 1)
Some obstacles to the dialogue

• Lack of interdisciplinary training

“The features related to these factors can be extracted from several levels of information, such as the phonetic, the prosodic, the linguistic and the syntactic level”

(Submitted paper, 2009)
Phonetic knowledge in speech technology

 ✓ The crucial question
 ✓ Early day vs. present day approaches
 ✓ Some obstacles to the dialogue
 ✓ Some consequences of the lack of dialogue
 ✓ Towards integration
Some consequences of the lack of dialogue

• The lack of dialogue has some consequences
  • Different focuses and interests
  • Difficulties in understanding the needs of “the other”
  • “Imperfect” technology
Some consequences of the lack of dialogue

“the phonetics community has not focused on questions most relevant for speech technology while the speech technology community has not developed algorithms and data structures that are optimally receptive for the incorporation of phonetic knowledge”

(van Santen, 2005: 149)
Consequences of the lack of dialogue

“The majority of the difficulties I have had in the past when cooperating with linguists stemmed from the fact that they gave me suggestions that were either very hard to incorporate in a computer program, or would probably not make any impact in overall system’s performance.”

(Acero, 1995: 174)
“Speech technology can proudly point to its apparent success with speech recognition and concatenative synthesis in defense of its machine-learning-centric approach…”

(Greenberg, 2005: 111)
Consequences of the lack of dialogue

“imperfect science is capable of providing an effective foundation for technology - as long as the demands of the market are not exceedingly stringent or profound”

(Greenberg, 2005: 111)
Demands of the market

• Natural and expressive speech synthesis
• Prosodic modelling
• Expression of emotions

“In the current age of corpus-based TTS systems that rely on searching a large inventory of speech for variable-length units to concatenate, with relatively little system modification of prosodic features, effective and realizable prosodic assignment represents a major problem.”

(Hirschberg, 2006: 55)
Demands of the market

- Spontaneous speech recognition
  - Heavily dependent on the availability of training data: spontaneous speech corpora annotated at all levels
- Task, domain and speaker independence
- Robustness
“If you permit me to give a caricature of present-day recognizers, then these machines are trained with all the speech, speaker, textual and environmental variability that may occur under the application in mind, thus giving the system a lot of global knowledge without understanding all the inter-relations. [...] Subsequently the recognizer performs rather well on average behavior and does poorly on any type of outlier, be it an unknown word, or a non-native speaker, or a fast speaker, or one with a cold, or a whispered input.”

(Pols, 1999: 4)
Demands of the market

- Pragmatically adequate spoken language systems
  - Speech acts (dialogue acts) and prosody
  - Emotions or user states ... but

  “the extra information that is called for is not that of raw emotional expression; rather it is the socially-relevant interpersonal information that signals speaker-listener relations, and speaker-attitude and affect, and discourse intentions”

  (Campbell, 2004: 10)
Phonetic knowledge in speech technology

✔ The crucial question
✔ Early day vs. present day approaches
✔ Some obstacles to the dialogue
✔ Some consequences of the lack of dialogue
✔ Towards integration
Phonetic knowledge in speech technology

• Is the combination of knowledge-based and “corpus-centric” approaches possible?

“Over the coming decades this tension is likely to dissolve into a collaborative relationship melding linguistic knowledge with machine-learning and statistical methods as a means of developing mature science and technology.”  

(Greenberg, 2005: 129)
... but, changes are needed

- Interdisciplinary training of phoneticians and engineers
- Academic structures should promote interdisciplinary collaboration and work with industry
- Companies should consider investing in basic long-term R&D
- Changes in “mentality” and “culture”
Phonetic knowledge in speech technology

“Computing power can not substitute crucial knowledge”
(Fant, 2004: 11)

http://www.speech.kth.se/~gunnar/
References


Batliner, A., & Möbius, B. (2005), Prosodic models, automatic speech understanding, and speech synthesis: Towards the common ground?, in *The integration of phonetic knowledge in speech*
References


van Santen, J. P. H. (2005), Phonetic knowledge in text-to-speech synthesis, in *The integration of phonetic knowledge in speech technology*. (W. J. Barry & W. A. van Dommelen, editors), Dordrecht: Springer, 149-166.
Il dialogo tra linguistica e tecnologia
6° Convegno Nazionale dell’Associazione Italiana di Scienze della Voce
Napoli, 4 febbraio 2010

http://liceu.uab.cat/~joaquim/speech_technology/AISV_10/AISV_10.html