

Contents

- TTS : What for?
- Challenges
- Available Technologies
 - DSP
 - NLP
- **Currently Available Systems**
 - Non-commercial
 - **Commercial**
- A Bright (Corpus-based) Future?
- Conclusion



AT&T's NextGen TTS

<http://www.research.att.com/projects/tts/>

- « Our research group's charter is to increase the naturalness of speech synthesis significantly while maintaining good intelligibility. Next-Generation TTS was introduced in 1998 and marked a dramatic leap in naturalness. Next-Gen has continued to improve markedly each year since, and is broadening the range of applications in which TTS can be deployed. »



AT&T's NextGen TTS

<http://www.research.att.com/projects/tts/>

- NLP
 - Bell Labs « front end » (30 years of upgrades for rule-based preprocessing and phonetization)
 - N-gram based contextual analysis
 - Tone-based intonation using corpus-based F0 synthesis, CART-based durations
- DSP
 - **Non Uniform Unit** Synthesis, using half-diphones as speech units
 - **ALMOST NO SIGNAL PROCESSING (!!!)**
 - 10 hours of speech (10 GB!!!), for Am. English ONLY
 - VERY careful choice of original voice
- Sold by **SpeechWorks** (Speechify, TM), Boston



and AT&T



Demo



Scansoft's RealSpeak

<http://www.scansoft.com/realspeak/>

- Previously Lernout and Hauspie's Realspeak
- Massively corpus-based, NLP/DSP
 - based on large phonetized and tagged databases in many languages
- NUU-based
 - units = diphones
 - target and concatenation costs are weighted in a phonetic class-dependent way
 - ex : higher for vowels
 - costs are passed through non linear functions
 - if lower than perceptual threshold : forced to 0
 - linear between threshold and acceptability threshold : forced to very high values



Scansoft's RealSpeak

<http://www.scansoft.com/realspeak/>

- 14 languages
- « The L&H RealSpeak engine generates speech that is almost indistinguishable from human speech. L&H RealSpeak is based on concatenation algorithms, where actual human voice segments are stored and used to convert any text into speech. In-depth language specific linguistic knowledge provides intelligent pronunciation of a wide range of variable input. »



RealSpeak
COMPACT



Babel Technologies' Babil

<http://www.babeltech.com>

- Massively Corpus-based
- NLP
 - Generic regular grammar for preprocessing
 - Dictionary-based tagging
 - CART-based phonetization
 - CART-based duration generation
 - Corpus-based intonation (non uniform intonative units+Viterbi)
- DSP
 - Diphone-based MBROLA (26 languages)
 - 2002 : NUU selection (French, ...)
- Compact and portable (Linux embedded apps)



Loquendo's TTS

<http://www.loquendo.com>

- Previously CSELT (Telecom Italia R&D)
- Massively corpus-based, NLP/DSP
 - based on large phonetized and tagged databases in many languages
- NUU-based, but little information...
- 9 languages
- Mostly integrated in hardware for call centers and telecom operators



Mi chiamo Valentina. La mia voce è prodotta dal sistema di sintesi di Loquendo. Scrivi una frase in italiano, e io te la leggerò.



Babel-Infovox's TTS

<http://www.babeltech.com>

- One of the oldest companies for multilingual TTS
- NLP
 - Exclusively rule-based : RULESYS (1982)
 - Regular rules, on a *single* tier (input=output), which gets increasingly complex as processing goes by.
 - Solid code ☺
 - 13 languages
- DSP
 - Formant rule-based synthesis (up to 1998) (VERY COMPACT!)
 - MBROLA-based (new system) using compressed databases (typ. 2 Mb/voice)



(ELAN)'s TTS

http://194.98.105.37/ttsdemo/default_fr.asp

- Diphone-based PSOLA synthesis
- 9 languages
- Not much additional information available (used to be massively rule-based)
- « Vous écoutez une démonstration du système de synthèse vocale d'Elan Informatique. Ma voix est produite grâce à la technologie TD-PSOLA. »



Contents

- TTS : What for?
- Challenges
- Currently Used Technologies
 - DSP
 - NLP
- Currently Available Systems
 - Non-commercial
 - Commercial
- **Conclusion :**
A Bright (Corpus-based) Future?

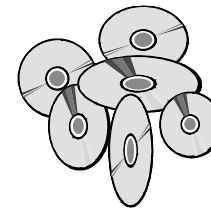
Other important players

- Nuance Communications (CA, spinoff of SRI)
- SpeechWorks, Boston
- (BT : closed)
- Lucent Technologies (Bell Labs)

NB : Most other companies do not sell *their* technology : they act as *integrators*.

Towards corpus-based techniques

1995-?: The *database* years

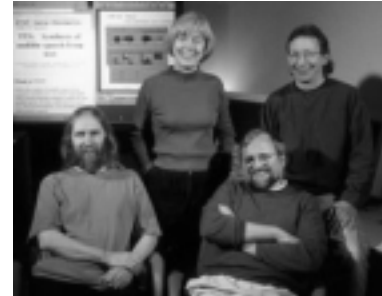


- For automatic phonetization
- For automatic generation of intonation and phoneme duration
- **For automatic selection of units for concatenative synthesis (ATR, Univ. Edinburgh, AT&T, FPMs?)**

On the importance of large text and speech corpora

- Tagged **text corpora** required for training phonetizers and taggers for TTS
- Phonetically labeled **speech corpora** needed for TTS (single speaker, 1-10 hours)
- **ELRA** (European Language Resource Agency) and **LDC** (Language Data Consortium) collect and distribute databases
- **From expert-based systems to corpus-based systems**

Example : AT&T NextGen Team



- Juergen Schroeter (head, formerly articulatory synthesis)
- Alistair Conkie (computer scientist experiences in TTS)
- Mark Beutnagel (Computer scientist experience in TTS)
- Ann Syrdal (linguist, resp. for dba preparation and assessment)

The only engineer is the dept head...

Speech Science?

This time is over

- planes do not flap their wings
- replace experts by corpora

cf. Jelinek 's «Each time I fire a linguist my recognition rate goes 1% higher»

1. Future milestones in speech processing will come from labs with strong commitment to solid, portable, and extensible code;
2. Speech scientists and software engineers will soon be the same people.

Speech Engineering!

**I don't believe
in
Computer "Science"**

**from R. Feynman's talk
on Quantum Computers
Bell Labs, 1985**

Where to go now?

- Books

- » *Traitement de la Parole*, R. Boite, H. Bourlard, T. Dutoit, J.Hancq, H. Leich, Presses Polytechniques Romandes, 2000
- » *An Introduction to Text-to-Speech Synthesis*, Thierry Dutoit, Kluwer Academic Publishers (Dordrecht), 1997
- » *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, R. Sproat (Ed.), Kluwer Academic Publishers, 1997.
- » *Data-Driven Techniques in Speech Synthesis*, R.I. Damper, ed., Kluwer Academic Publishers (Dordrecht), 2001

- On the web

- » speech.comp newsgroup
- » the unescapable speech FAQ :
<http://svr-www.eng.cam.ac.uk/comp.speech/>
- » LDC web-based TTS comparison :
<http://morph ldc.upenn.edu/cgi-bin/lts/list>
- » My speech course web site
<http://tcts.fpms.ac.be/cours/1005-08/speech/>



T. Dutoit © 1997