

# CONCLUSION

# Where are we now?

- Speech **recognition** is now reaching the its own market of **consumer products** (not really personal dictation systems);

Still a long way to go, though (robustness, language models, speech understanding)

- Speech **synthesis** is now living **exciting** times,  $\approx$  ASR in the mid 80's (automatic unit selection)

# On the importance of large text and speech corpora

- Tagged **text corpora** required for training language models for ASR, phonetizers and taggers for TTS
- Phonetically labeled **speech corpora** needed for ASR (multi-speaker, 100s hours) and TTS (single speaker, 1-5 hours)
- **ELRA** (European Language Resource Agency) and **LDC** (Language Data Consortium) collect and distribute databases
- **From expert-based systems to corpus-based systems**

# *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!*

# 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...**

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

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