Retrieval of Speech in MPEG-7 Audio

Philip N. Garner

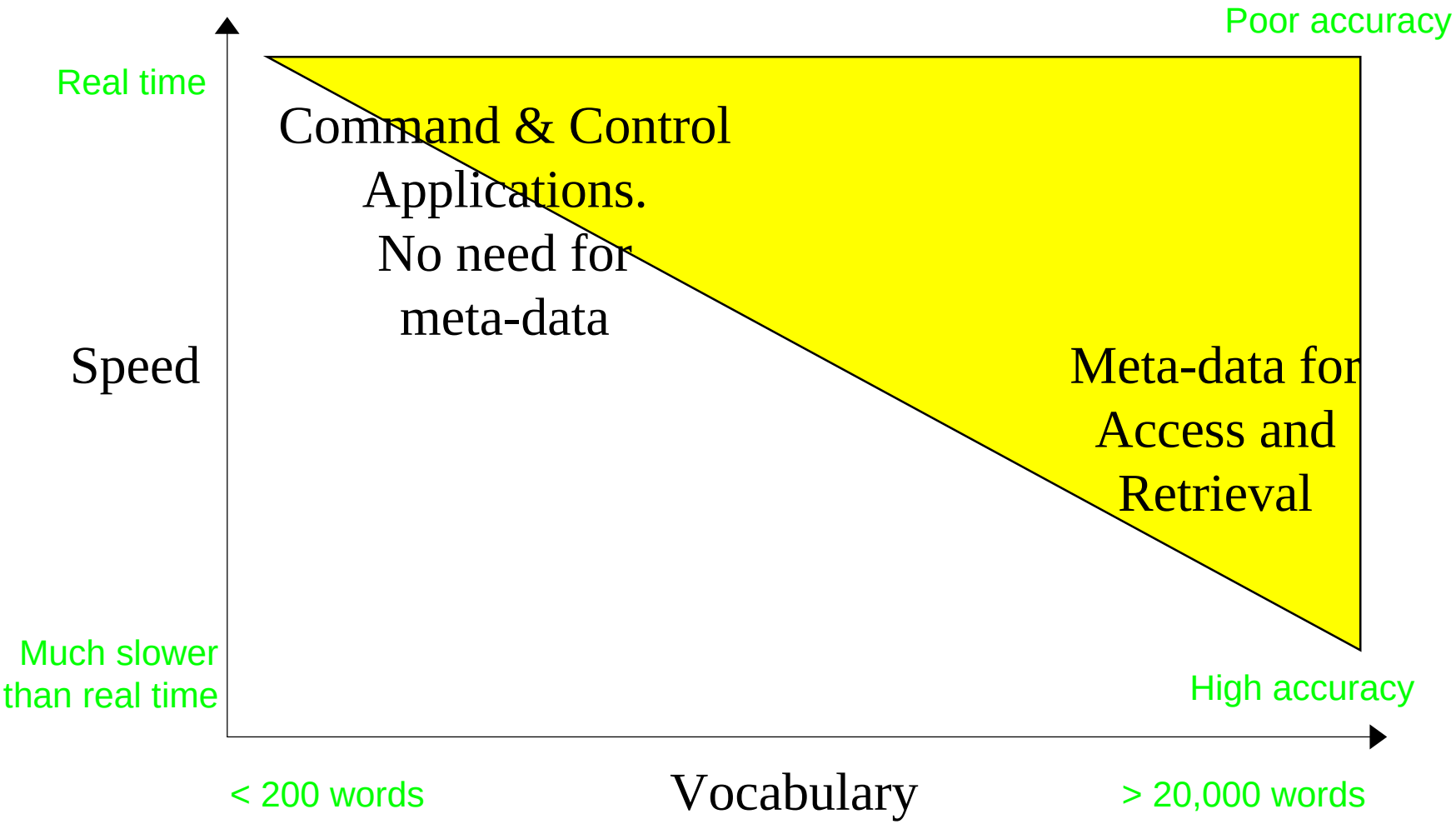
Canon

Why Speech?

- **Automatic Speech Recognition (ASR) has matured over the past 30 years**
  - Virtually all based on Hidden Markov Models
- **Several commercial packages available, e.g.**
  - IBM (ViaVoice)
  - Dragon (acquired by L&H)
  - Microsoft (Free SAPI engine)
- **Much other work . Behind The Scenes. :**
  - Most telephone companies
  - Most consumer electronics companies
  - Many universities

Canon

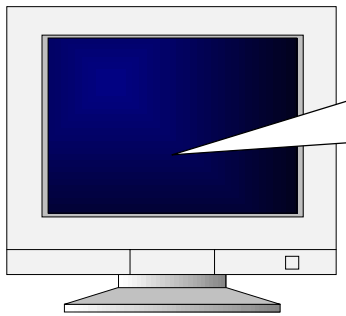
Speech recognition technology



Canon

Accuracy is not enough

Recognise speech!



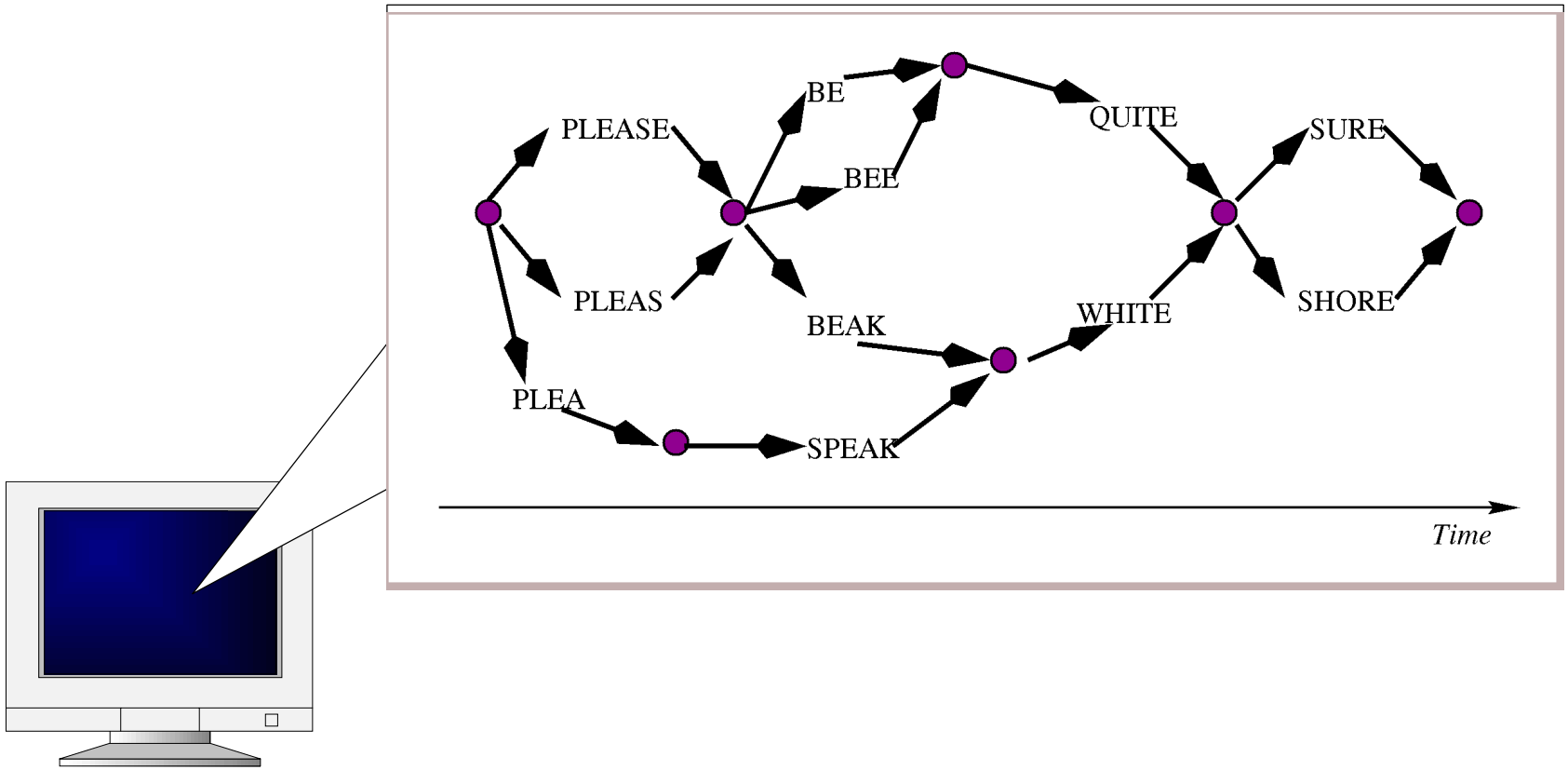
...wreck a nice beach...

Recognize speech!

Canon

A common solution

Please be quite sure!

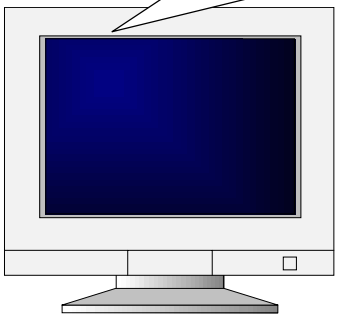


Canon

So, lattices are enough, right?

Let’s tell Juergen Herre

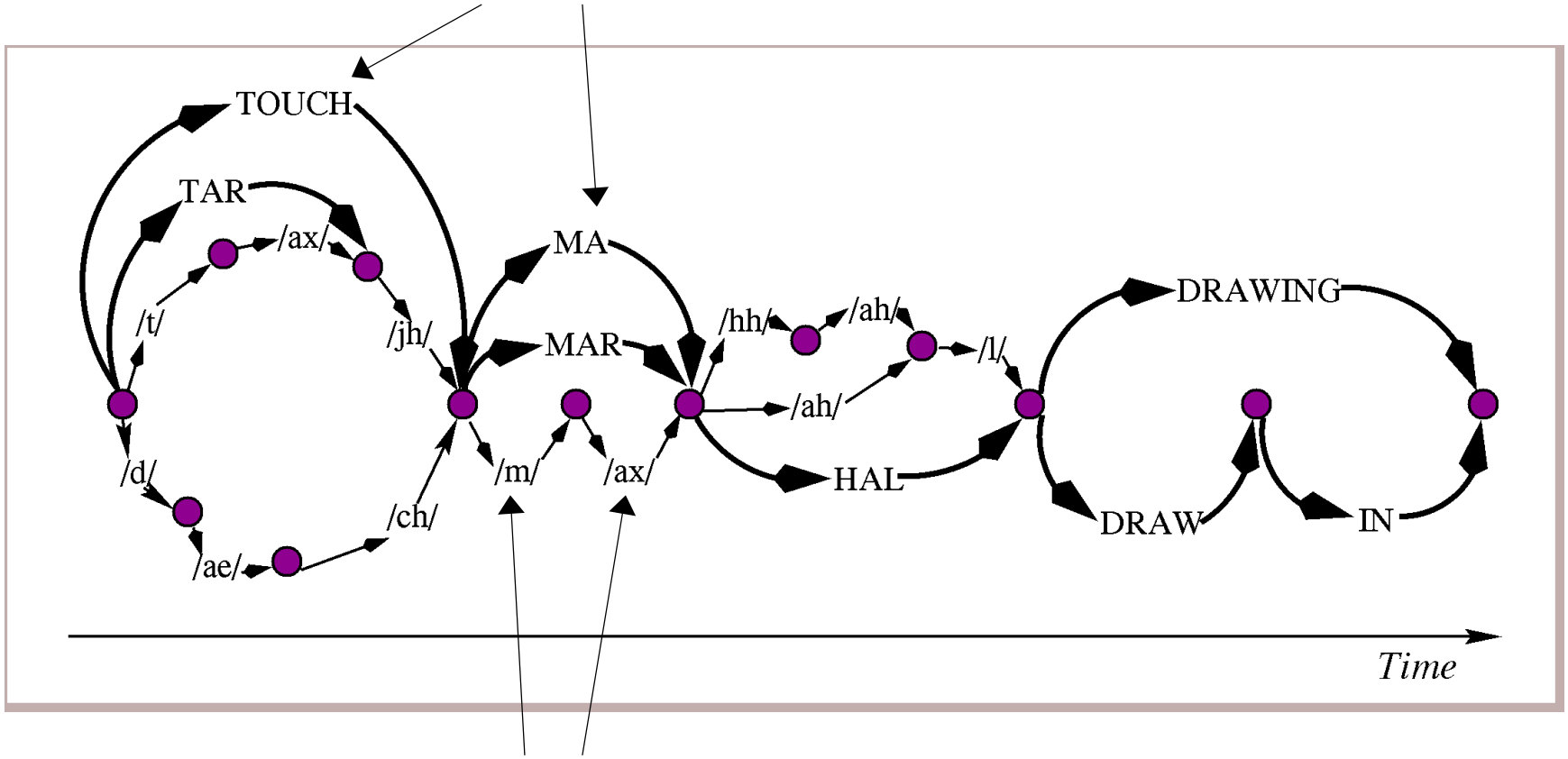
...your can hear her...



Canon

Proper names and places

Out Of Vocabulary words cannot be decoded correctly



Phones are always defined, but less reliable

Canon

Spoken Content Summary

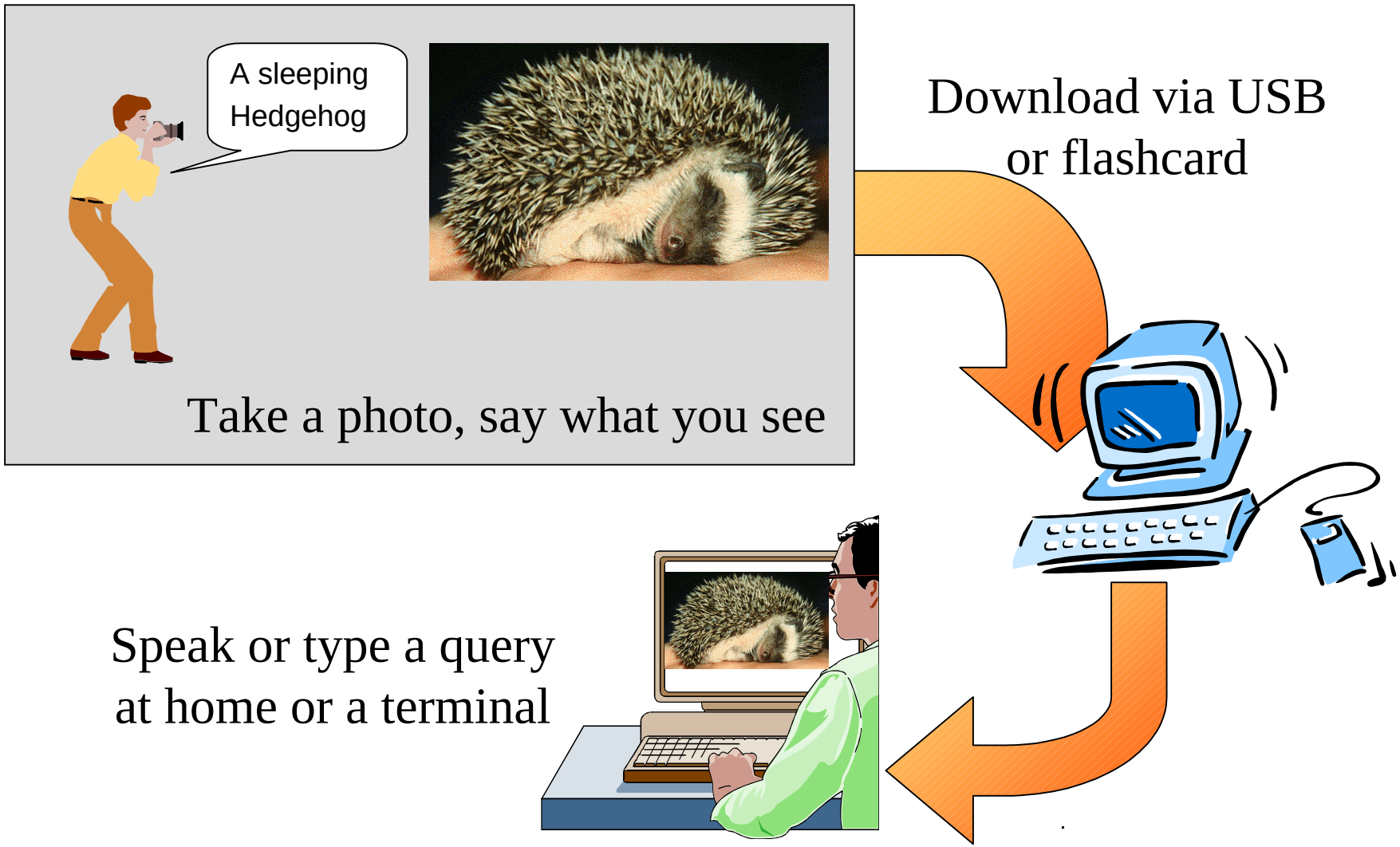
- **The MPEG-7 Spoken Content DS is designed around state of the art speech recogniser capabilities:**

- **Speech recognisers do not produce plain text**
  - They produce lattices
  - Spoken Content DS stores these lattices rather than plain text
  - Spoken Content DS hence allows retrieval of ambiguities

- **Speech recognisers do not have complete vocabularies**
  - They can decode phones, however
  - Spoken Content DS stores phones too
  - Allows retrieval of words which were unknown at the time of annotation

Canon

An image retrieval application



Canon