5.0 Voice Interface for Human–network Interaction

5.1 Vision and Current Status

- **Ubiquitous Computing**
  - huge volumes of data disseminated across the globe by optical fiber networks
  - any time, from anywhere by wireless terminals
  - vehicular electronics, PDA, handset, home appliance, etc.

    new platforms accessing the global network information/services
  - traditional keyboard/mouse not adequate any longer

    size shrinkage, different user environment, etc.

    desired functionalities/human-network interactions increasing
  - voice interface will be one out of the few most important, natural, user friendly, attractive interface
  - examples: voice browser, voice portal, voice web

    voice-based web-user interaction

    voice-based web tools/Application Interfaces, etc.
  - voice interface is the only major “missing link” in the “semi-mature” technology chain

Core Technologies for Voice Interface

- **Speech Recognition**
  - word recognition

    voice command/instructions
  - keyword spotting

    identifying the keywords out of a pre-defined keyword set from input voice utterances
  - large vocabulary continuous speech recognition

    transforming input voice into corresponding full text

    entering longer texts
  - speaker dependent/independent/adaptive
  - acoustic reception/background noise/channel distortion
  - read/spontaneous/conversational speech
- **Text-to-speech Synthesis**
  - transforming any input text into corresponding speech signals
  - e-mail/Web page reading
  - prosodic modeling
  - basic voice units/rule-based, non-uniform units/corpus-based

Core Technologies for Voice Interface

- **Speech Understanding**
  - limited domains/tasks
  - spoken-dialogue-based interaction
  - identifying speaker's intention
  - grammatical approaches, partial parsing, lattice parsing
  - statistical approaches, corpus-based by training
  - word/phase analysis, semantics
- **Speaker Verification**
  - verifying the speaker as claimed
  - applications requiring verification
  - text dependent/independent
  - integrated with other verification schemes

Integrated Voice Interface Functionalities

- **Dictation/Transcription**
  - remote authoring/dictation
  - voice-mail transcription
  - meeting recording transcription
- **Spoken Dialogue Systems**
  - voice interface for all human-network interaction
  - voice portals, real-time information, personal database, electronic commerce, web page interface, etc.
  - speech recognition/understanding, text-to-speech synthesis, dialogue manager
  - system/user initiatives
  - human factors, reliability/efficiency, transaction success rate/average dialogue turns
- **Voice Indexing/Retrieval**
  - voice personal notebook, voice-mail retrieval, customer service analysis, etc.
  - integration with multi-media retrieval
  - indexing features/relevance evaluation
  - re-call/precision rates, average precision rate
- **Multi-lingual Functionalities**
  - globalized network with multi-lingual content/multi-lingual users
  - language dependent/independent technologies
  - major languages, dialects/accents
  - text/speech databases, lexicon for different languages for statistical approaches
  - shared acoustic units, plug-in language models

Distributed Speech Recognition (DSR) for Network Environment

- **Client-server**
  - feature parameters extracted at client

    transmitted via wireless links/networks to the server
  - voice processing primarily handled at servers
- **Network Response**
  - text-to-speech synthesis at server or client
  - data output directly