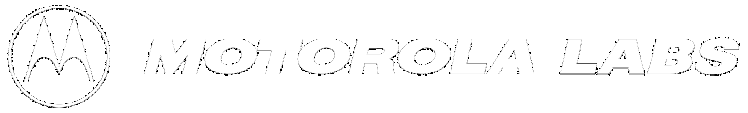
RTP Payload Format for Distributed Speech Recognition

draft-meunier-avt-rtp-dsr-00.txt

Jeff Meunier

Human Interface Laboratory

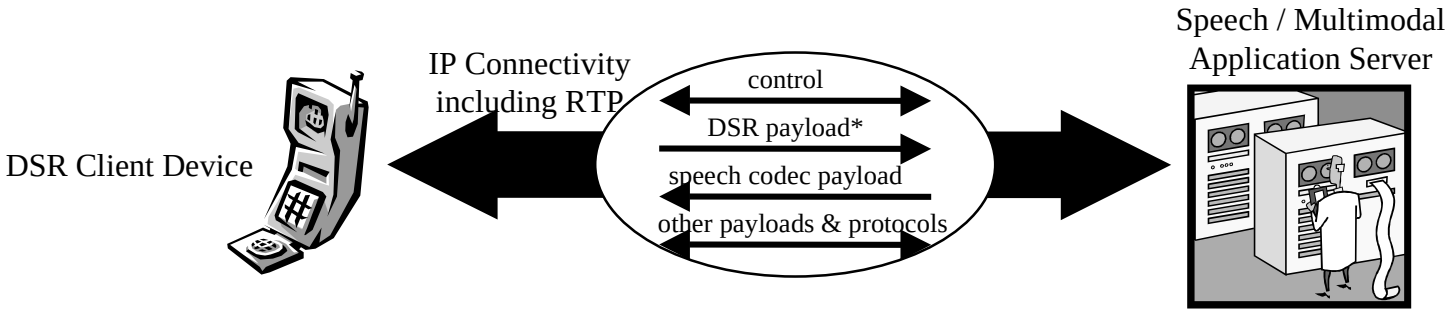


Presentation Goals & Overview

- **Provide background on DSR concept**
  - **motivate interest with potential applications of payload**
  - **provide technical summary**
- **Illustrate need for DSR payload**
  - **describe benefits over standard speech codecs**
- **Outline protocols concerns and DSR deployment issues**
  - **motivation to use existing protocols**
- **Solicit feedback on the use of DSR payload within RTP**

Introduction to DSR: What is it?

- **Voice payload, designed for use in speech recognition**
- **Sent from client device (e.g., mobile handset, PDA) to recognition server (e.g., voice service platform)**
- **Would be paired with other data, likely including voice feedback, in a voice enabled application**



Introduction to DSR: Applications

- **Joint voice & data applications**
  - ***need a transport that can coexist with others and provides high recognition accuracy at a low bitrate***
  - **voice browsing (e.g., voice driven WAP)**
  - **short messaging & dictation**
  - **multimodal interfaces with speech input**
- **Interactive dialogue applications**
  - ***like a "conversation" with the remote server, requires low latency for best usability***
  - **interactive voice response (e.g., call centers, voice portals)**
  - **remote services (e.g., voicemail)**
- ***DSR payload an important enabler of these applications***
  - **low bitrate and latency, tuned for high performance speech recognition, accommodates errorful channels**

Introduction to DSR: Use of ETSI Payload

- **Frame-based payload**
  - **4.8 kbps for all sampling rates (8, 11, 16kHz)**
  - **mel-cepstral representation**
    - **as used in speech recognition**
    - **quantized**
  - **error concealment via forward-backward copy**
- **Multiframe header**
  - **sent for every 240ms of speech**
  - **contains payload specific config info**
- **Draft proposes use of ETSI payload for RTP**
- **92-bit “frame pair” ADU**
  - **represents 20ms of speech**

Payload Features & Speech Codec Limitations

- **DSR payload provides better speech recognition performance than low data rate codecs**
  - **optimized for speech recognition instead of audible quality**
  - **evidence of performance drop due to codec distortion: literature (e.g., Lilly & Paliwal ICSLP96), ETSI work**
  - **“direct” representation: since it is the preferred format for SR, requires minimal decode and can be used directly**
  - **error mitigation also optimized for recognition performance, FEC techniques possible as well**
- **DSR payload operates at less than half the bitrate of “tolerable” codecs**
  - **current codecs below ~13kbps introduce performance degradation unacceptable for many SR tasks**
  - **DSR payload at 4.8kbps is transparent, same performance as 64kbps raw speech**

Progressing DSR Protocols

- **Why leverage RTP for DSR?**
  - **DSR only one component in an end-to-end system, other more general VoIP components will already use RTP**
    - **encourages/simplifies adoption of DSR standard**
    - **reduces complexity of implementations by using common protocol stack**
  - **Lack of support for simultaneous voice & data in current mobile networks**
  - **Early work indicates direction (e.g., Nuance support of H.323)**
- **RTP features with utility to DSR payload**
  - **sequence number: packet reordering & loss detection**
  - **timestamp: sync of parallel data sources (e.g., optional speech data, related UI events, FEC)**
  - **marker bit: indicate presence of multiframe header**
  - **certainly more (RTCP, CRTP,…)**

Progressing DSR Protocols

- **ETSI Apps & Protocols sub-group**
  - **identifies next step in DSR deployment is finding protocol solution(s)**
  - **demonstrated preference for use of existing protocols where possible**
- **W3C Voice working group**
  - **interested in apps & protocols as it pertains to enabling voice browsing using DSR**

Feedback Needed From WG

- **Other relevant IETF activities, protocols, & proposals**
  - **header compression: use for low bitrate payloads, realistic usage scenarios**
  - **SDP/SIP: usage, gotchas**
  - **other activities?**
  - **public domain RTP implementations (esp. for mobile devices)**
- **Robustness to packet loss**
  - **considerations for RTP over IP vs. mobile IP (different?)**
  - **error concealment & FEC design issues**
- **Packetization of payload**
  - **accommodating “parallel” information: drawbacks to carrying in separate RTP streams**
  - **packet size**

Conclusion

- **Summary**
  - **Compelling multimodal applications motivate use of DSR on RTP**
  - **DSR payload provides significant gains over traditional speech codecs**
  - **Use of existing RTP protocol and VoIP call models seem best for easy & fast adoption of DSR payload**
  - **Feedback from working group on packetization, packet loss robustness, and other relevant activities would be extremely valuable**
- **Next steps?**

- **Contact information:**
  Jeff Meunier
  meunier@labs.mot.com
  +1-847-538-9479