Handoff with DSP Support: Enabling Seamless Voice Communications across Heterogeneous Telephony Systems on Dual-Mode Mobile Devices

Hung-Yun Hsieh (謝宏昀)

Department of Electrical Engineering & Graduate Institute of Communication Engineering
National Taiwan University
October 8, 2008
Outline

- Dual-mode mobile handsets
- Research challenges
- DSP support for vertical handoff
- Performance evaluation
- Summary
Heterogeneous Wireless Networks

- Multiplicity of wireless communication systems
  - 3G/3.XG, WiFi, WiMAX, mesh networks, ...

- Disparity of wireless networks
  - Access technologies, network architecture, service provisioning, ...
Multi-Mode Mobile Devices

- Providing multiple modes of communication in one device
  - Dopod CHT 9100, HTC Touch Diamond, Apple iPhone, ...
  - GSM+WiFi, 3G+WiFi, 3G+WiMAX, ...
  - Richer set of services
  - Always best connected
  - Resource aggregation
  - ...

E-commerce
Teleconferencing
Music on demand
On-line TV
Surveillance
Tour guide
Heterogeneous Communication Modes

- Not just a multi-band mobile handset

- Disparate network protocol stacks
  - GSM/3G/HSPA mode
    - Circuit-switched voice service
    - Packet data service
  - WLAN/WiMAX mode
    - IP-based infrastructure

- Mismatched network characteristics
  - Bandwidth, latency, jitter, loss rates, ...
Dual-Mode Handset Application Scenarios
Research Challenges

- Data access through multiple modes with heterogeneous characteristics
  - Mobility management
  - Bandwidth aggregation
  - Congestion control

3. An End-to-End Approach for Transparent Mobility across Heterogeneous Wireless Networks [MONET 2004]
4. On Transport Layer Adaptation in Heterogeneous Wireless Data Networks [IWQoS 2005]
Research Challenges (cont.)

- One in **voice** mode and one in **data** mode
  - Circuit-switched voice service vs. packet-switched data service
  - Protocol stack and network infrastructure are distinctly different for the two paths

  ❖ Traditional IP-based approaches cannot be used in this context
  ❖ New solution paradigms are needed
Ubiquitous Voice Communication

- Seamless voice continuity between packet-switched and circuit-switched voice streams
  - Failures of packet-based approaches
    - Time-scale modification of speech waveforms

GSM voice call while using VoIP (over WLAN) opportunistically → ABC: always best-connected
Heterogeneous Teleconferencing

- IP video atop an existing audio conference with legacy devices for enhanced user experience

- Synchronization of the packet-switched video to the circuit-switched voice mixture
  - Failures of packet-based approaches
  - Source separation based on speech sparsity
Case Study

Ubiquitous voice communication across heterogeneous telephony systems

IEEE Transactions on Mobile Computing (to appear)
Elsevier Computer Networks (September 2008)
Ubiquitous Voice Communication

- GSM/3G voice ↔ VoWLAN (voice over WLAN)
Mismatch of End-to-End Delay

![Graph showing end-to-end delay for different networks]

- Wide-area WLAN
- GSM
- Metro WLAN
- Campus WLAN
- Local WLAN

End-to-End Delay (ms) vs. Experiment Count
Time Alignment

- Synchronization of GSM and VoWLAN voice streams with varying delay characteristics
  - Similarity measure of speech waveforms
    - Cross-correlation or cross-AMDF
  - From GSM to VoWLAN
    - New stream is ahead of time (hold in the buffer)
  - From VoWLAN to GSM
    - New stream is lagging
Temporal Discontinuity

- Gap in the received voice stream
  - For VoWLAN=50ms and GSM=350ms, a gap of about 300ms is introduced

- Temporal discontinuity impairments
  - Degradation of speech quality (proportional to the magnitude of the discontinuity)
  - More objectionable and annoying to native listeners

- Potential solutions
  - Dejitter (playout) buffer?
  - Wait for handoff until the silence period?
Time-Scale Modification

- **Goal**
  - Slow down the leading voice stream to mitigate the audio gap
  - Explore human’s insensitivity to minor modulations in the speed of a speech signal

- **Requirements**
  - Modify only the timing attribute (e.g. speaking rate)
  - No modification on the perceived frequency attribute (e.g. pitch)
  - Online modification of speech (time-domain operations preferred)
Time-Frequency Representation

\[ X(\omega, m) = \sum_{n=-\infty}^{+\infty} x(n + m)w(n)e^{-j\omega n} \]

- Synthesize \( y(n) \) such that its STFT \( Y(w, n) \) is maximally close to \( \hat{Y}(w, n) \) in the least-square sense

\[ E = \sum_k \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left| \hat{Y}(\omega, k) - Y(\omega, k) \right|^2 d\omega \]
Time-Domain Operation

- Parseval’s theorem

\[ E = \sum_{k} \sum_{m=-\infty}^{+\infty} (\hat{y}_w(m,k) - y(m+k)w(m))^2 \]

\[ \frac{\partial E}{\partial y(n)} = -2 \sum_{k} (\hat{y}_w(n-k,k) - y(n)w(n-k)) \times w(n-k) = 0 \]

\[ y(n) = \frac{\sum_k w(n-k)\hat{y}_w(n-k,k)}{\sum_k w^2(n-k)} \]

where

\[ \hat{y}_w(n-k,k) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \hat{Y}(\omega,k) e^{i\omega(n-k)} d\omega \]

- Different methods exist in approximating \( \hat{y}_w(n-k, k) \) from operations of \( x(n) \)
Overlap-Add Synthesis (OLA)

\[ y(n) = \frac{\sum_k w^2(n - kS)x(n - kS + \tau^{-1}(kS))}{\sum_k w^2(n - kS)} \]

- Serious phase discontinuities may occur at segments joints with a straightforward application of the overlap-add (OLA) method

- Introduce tolerance \( \Delta_k \) in the synthesis of \( y(n) \)

\[ y(n) = \frac{\sum_k w^2(n - kS + \Delta_k)x(n - kS + \tau^{-1}(kS) + \Delta_k)}{\sum_k w^2(n - kS + \Delta_k)} \]
Waveform-Similarity OLA (WSOLA)

- Waveform similarity
  - Maintain maximum local similarity to the original waveform in all joints
    \[ \forall m : y(n + m)w(n)(=)x(n + \tau^{-1}(m) + \Delta_m)w(n) \]
    \[ \forall m : \hat{Y}(\omega, m)(=)X(\omega, \tau^{-1}(m) + \Delta_m) \]

- Use the Hann window to remove the denominator
  \[ y(n) = \sum_k w^2(n - kS)x(n + \tau^{-1}(kS) - kS + \Delta_k) \]
  \[ \Delta_k \in [-\Delta_{\text{max}} \ldots \Delta_{\text{max}}] \]
Testbed Setup
Dual-Mode Handset

- Implementation on an O₂ handset
  - Windows Mobile 5.0 Platform
  - Intel PXA272 CPU (400MHz) & 64MB RAM
Audio Mixing Block

Audio Out

Time Scale Modification

\( \tau > 0 ? \)

Delay

Mismatch Profiling

Buffer

Gain Control

Old Stream

New Stream

Audio Mixing
Handoff Control

- Make-before-break soft handoff
Performance Metric: PESQ

- Perceptual evaluation of speech quality
- An ITU standard
Optimization of Parameters

- Computation complexity
  - Window length, search range, search granularity

![Graphs showing the relationship between window length and quality, and delta divisor and complexity](image-url)
Optimization of Parameters (cont.)

- Tradeoffs between computational complexity and speech quality
A Closer Look at the Waveform
Subjective Quality Test

- Experiment with 30 users
- Time-scale modification: 3s - 4.5s
Summary

- Research challenges for dual-mode mobile devices over heterogeneous wireless networks
- Case study for seamless vertical handoff based on digital speech processing algorithms
  - Computational complexity is acceptable on existing dual-mode handsets
- Cross-disciplinary research between signal processing and networking
Questions and Comments