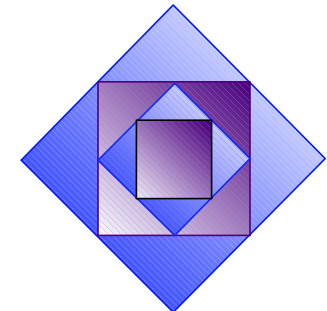
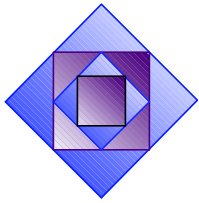


Recognition Compatible Voice Coder (RECOVC)

Multimedia & Signal Processing Group

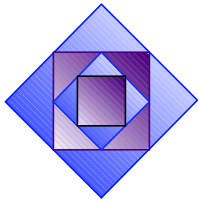
IBM Research Lab in Haifa





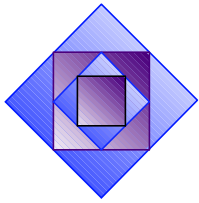
The Application

- **A low-complexity (possibly mobile) pervasive device captures voice**
- **Voice needs to be converted to text by an Automatic Speech Recognition (ASR) Server**
- **Voice needs to be compressed in order to reduce storage space or transmission bandwidth**
- **Playback of the voice on the pervasive device may be required (e.g. digital voice recorders)**
- **Playback of the voice on the server may be required**
 - Interactive Voice Response (IVR) services of "sensitive type", e.g. banking and brokerage transactions (playback is a legal requirement)
 - Human verification of collected speech databases, used to retrain the engine and improve recognition rates
- **Other applications: real-time voice communication with ASR-based computer monitoring (e.g. key word spotting)**



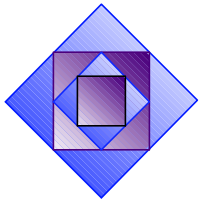
The Problem

- **Low bit rate speech compression methods available today are tuned for the human listener**
- **ASR performance substantially degrades when processing compressed speech**
- **Especially true for:**
 - Large vocabulary tasks,
 - Continuous speech recognition,
 - Noisy recording environments,
 - Channels prone to bit-errors and packet loss.



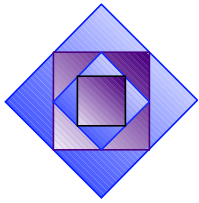
Possible Solutions

- **Do not use low bit rate compression when ASR is to be performed (e.g. use G.711 at 64 kbps for telephony bandwidth speech)**
 - Not applicable for most applications
- **Develop new ASR engines that are tuned for specific low bit rate speech coders (e.g. AMR-GSM)**
 - Requires a new engine when the network coder changes
 - Questionable performance for large vocabulary tasks, noisy recordings and channels prone to bit-errors and packet loss
 - Not extendable to wideband speech



Distributed Speech Recognition (DSR)

- ASR features are calculated on the pervasive device (ASR "front-end")
- ASR features are compressed at low bit rate such that the performance of the speech recognizer does not degrade
- Compressed ASR features are packed with error detection and recovery bits, and transmitted over a data channel to the Server
- ASR "back-end" performed on the server
- DSR Standardization Activity:
 - ETSI/Aurora
 - ITU-T SG16 (new activity)
- **The problem: Speech playback capability is lost !**
 - Transmission of an additional compressed voice stream for playback purposes only increases complexity & bandwidth



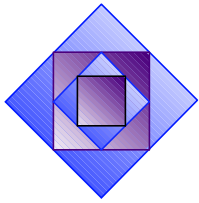
What is RECOVC ?

■ IBM proprietary technology for DSR

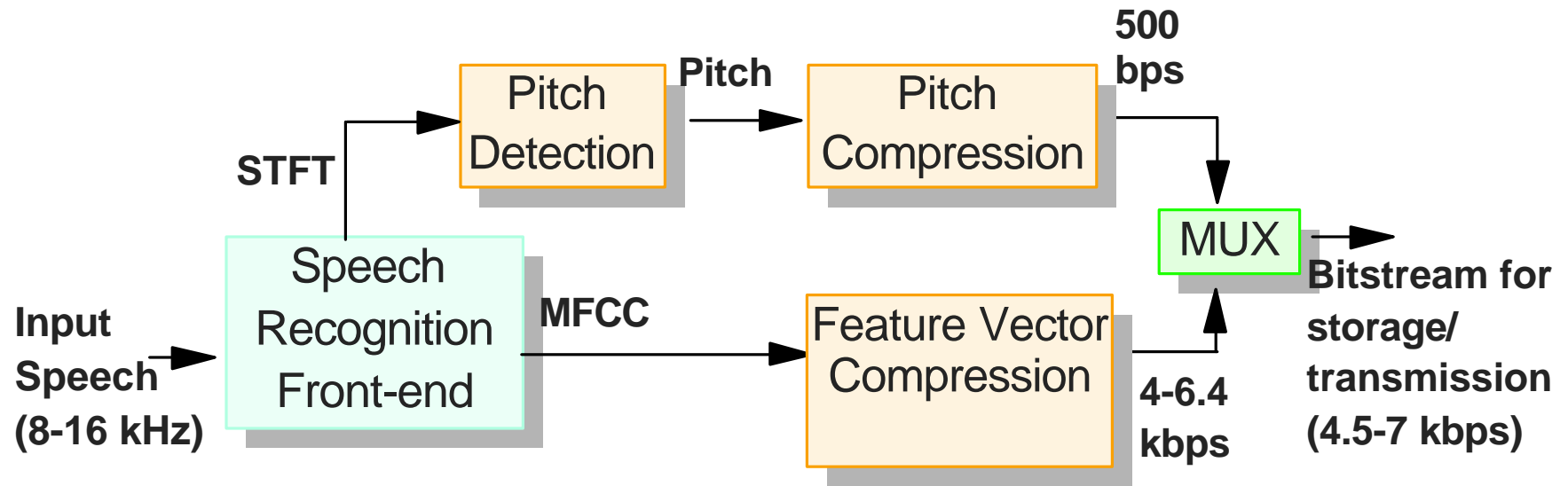
- Suitable for large-vocabulary, continuous speech recognition (as well as small vocabulary, command & control type applications)
- Interoperable with the IBM ViaVoice recognition engine (8 - 16 kHz speech)
- Enables playback of speech on the pervasive device and on the Server, with minimal overhead

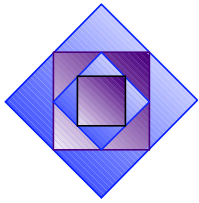
■ Includes:

- Extraction of Mel-Frequency Cepstral Coefficient (MFCC) on the pervasive device (common ASR features)
- Low complexity fundamental-frequency ("pitch") extraction on the pervasive device
- Compression of ASR features such that recognition rates do not degrade, and compression of fundamental-frequency
- Transmission format suitable for packet-based data channels
- Speech reconstruction from the decoded ASR and fundamental-frequency, for playback purposes

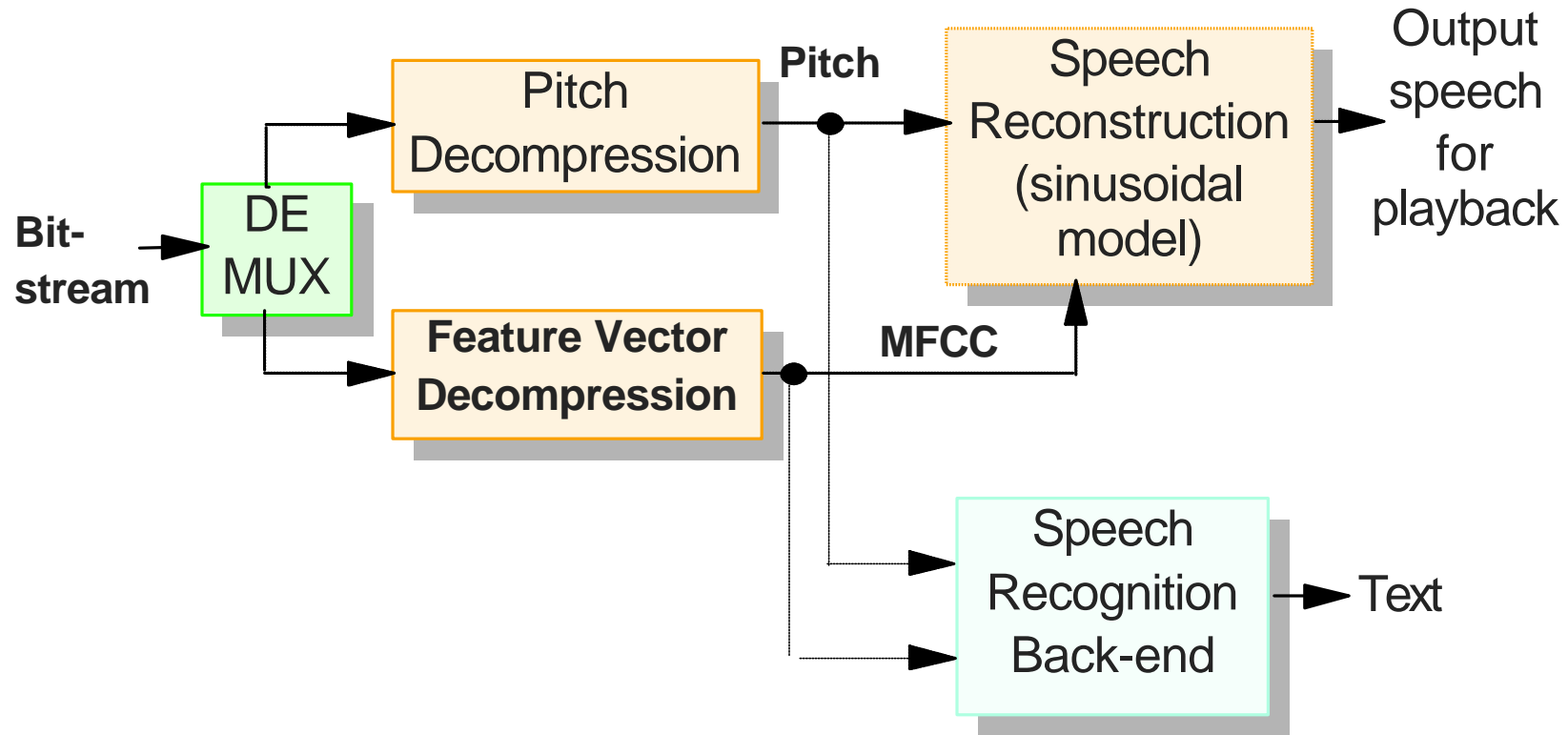


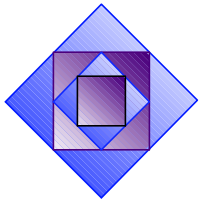
RECOVC Encoder (Client Side)



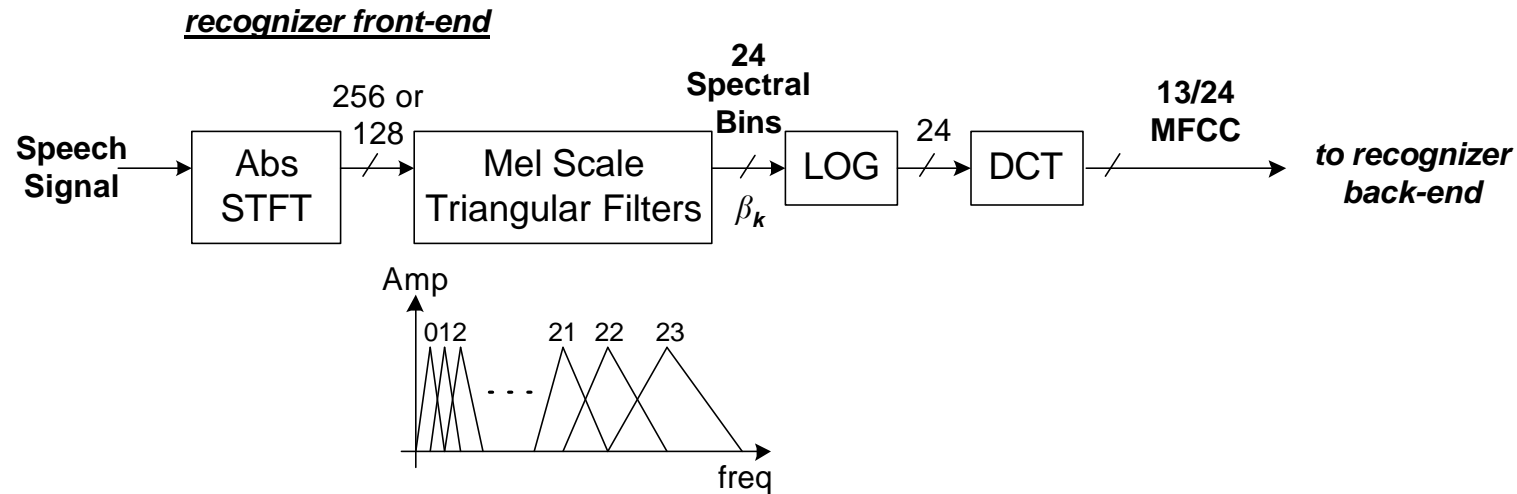


RECOVC Decoder (Server Side)





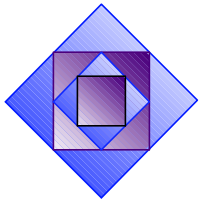
Mel-Frequency Cepstral Coefficients (MFCC)



- ★ Phase lost while taking the ABS of STFT
- ★ Spectral resolution lost while integrating
- ★ Only part of the MFCCs are maintained (13 out of 24)

Q: Can intelligible speech be regenerated from MFCC ?

A: Yes, if the pitch and voicing information is present



Implementation Results

■ Bit rates:

	MFCC only:	MFCC+pitch*:
13 MFCCs:	4 kbps	4.5 kbps
24 MFCCs:	6.4 kbps	6.9 kbps

* Reconstruction enabled

■ Floating point run-time*:

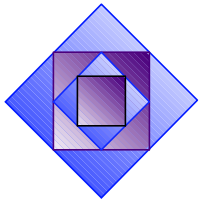
MFCC Compression/ Decompression	Full RECOVC Encoder	Full RECOVC Decoder
~3.5 % -MFCC calculation (24 dim) -MFCC compression & decompression	~9 % - Pitch Detection - MFCC calculation - MFCC & Pitch compression	~4 % (8 kHz) ~6.5 % (22 kHz) - MFCC & Pitch decompression - Speech reconstruction

* Real-time percentage on Pentium II 266 Mhz running on Windows NT 4.0

■ 32 bit fixed point implementation:





Mode	CPU (MIPS*)	ROM (KB)	RAM (KB)
Full RECOVC	15	90	40
MFCC Comp./decomp. only	4	70	10

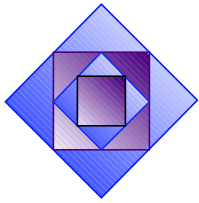
* Estimated, encoder plus decoder



Demonstration

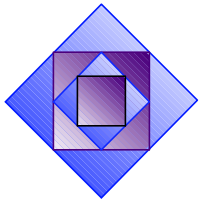
- Signals are sampled at 8 kHz

Original	Reconstructed (6.9 kbps)
	
	



ETSI/Aurora Activity

- **Enabling DSR services for mobile devices**
- **2.5 & 3G handsets, VoIP (RTP), ...**
- **Members:**
 - IBM, Nokia, Motorola, Intel, Alcatel, Ericsson, Qualcomm, TI, FT, HP, Siemens, SpeechWorks, ...
- **April 2000 - First front-end standard for MFCC feature extraction & compression (ETSI ES 201 108)**
- **1Q 2002 - Advanced front-end standard (noise robust)**
- **New activity - "extension of front-end for tonal-language recognition and speech reconstruction"**
 - IBM is a major contributor to this activity (standardization of RECOVC technologies)



For Further Information

- **RECOVC Web Site:** <http://www.haifa.il.ibm.com/recovc/>
 - Overview
 - Presentation
 - Demonstration
 - Conference papers
- **IBM Conference Papers:**
 - D.Chazan, G. Cohen and R. Hoory, "Efficient Periodicity Extraction Based on Sine-Wave Representation and its Application to Pitch Determination of Speech Signals", EUROSPEECH 2001
 - D. Chazan, G. Cohen, R. Hoory and M. Zibulski, "Low bit rate speech compression for playback in speech recognition systems", EUSIPCO 2000.
 - D. Chazan, G. Cohen, R. Hoory and M. Zibulski, "Speech reconstruction from mel-frequency cepstral coefficients and pitch frequency", ICASSP 2000
 - G. N. Ramaswamy and P.S. Gopalakrishnan, "Compression of acoustic features for speech recognition in network environments", ICASSP 1998.
- **Contact person: Gilad Cohen (giladc@il.ibm.com)**