A Standard Audio Encapsulation Method

Homayoon Beigi
Beigi@RecognitionTechnologies.com
http://www.RecognitionTechnologies.com
of
Recognition Technologies, Inc.
300 Hamilton Avenue
White Plains, NY, U.S.A.

and

Judith Markowitz
Judith@JMarkowitz.com
http://www.JMarkowitz.com
of
J. Markowitz Consultants
5801 N. Sheridan Road
Chicago, IL, U.S.A.

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Large-Scale Speaker Recognition

- **Large Government Applications**
  - Social Security Eligibility Verification, Border Crossing, etc. – *millions of participants*
  - Forensic Applications
  - Verification of Life Status for remote citizens – *e.g. Pension plans*
- **Financial Applications** – *Fraud Protection, Account Access, etc.*
- **Large Health Insurance Memberships** – *Access to Medical Records, etc.*
- **Large Corporation VoiceMail Applications**
- **Telephone Order Credit Card Charges** – *Verify buyers in place of signature*
- **Remote Test Proctoring** – *Requires continuous verification*
- **Other System-Wide Applications** – *Requiring Remote Authentication or Customization*

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Starting Question to Ask

- **What Should be Standardized at This Stage of Development in Speaker Recognition?**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Starting Question to Ask

- **What Should be Standardized at This Stage of Development in Speaker Recognition?**
  - **Audio Format?**
  - **Speaker Models?**
  - **Results of Recognition?**
  - **Interaction with Engines?**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Starting Question to Ask

- **What Should be Standardized at This Stage of Development in Speaker Recognition?**
  - **Audio Format?** Definitely
  - **Speaker Models?**
  - **Results of Recognition?**
  - **Interaction with Engines?**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Starting Question to Ask

- **What Should be Standardized at This Stage of Development in Speaker Recognition?**
  - **Audio Format?** Definitely
  - **Speaker Models?** Not Yet
  - **Results of Recognition?**
  - **Interaction with Engines?**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Starting Question to Ask

- **What Should be Standardized at This Stage of Development in Speaker Recognition?**
  - **Audio Format?** Definitely
  - **Speaker Models?** Not Yet
  - **Results of Recognition?** Yes
  - **Interaction with Engines?**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Starting Question to Ask

- **What Should be Standardized at This Stage of Development in Speaker Recognition?**
  - **Audio Format?** Definitely
  - **Speaker Models?** Not Yet
  - **Results of Recognition?** Yes
  - **Interaction with Engines?** Yes

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Proposal and Status

- **An Audio Encapsulation Standard**
- **Meeting Specific Requirements** – *Discussed Later*
- **Currently Considered by ANSI/INCITS for standardization**
  - **Public Review Period has been Completed**
- **Being Considered by ISO/JTC1 SC37**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Goals (Audio Format Only)

- **A Basic List of Audio Formats Meeting All Interchange Requirements**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Goals (Audio Format Only)

- **A Basic List of Audio Formats Meeting All Interchange Requirements**
  - **With Minimal Redundancy for the Sake of Clarity, Simplicity, and Compactness**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Goals (Audio Format Only)

- **A Basic List of Audio Formats Meeting All Interchange Requirements**
  - **With Minimal Redundancy for the Sake of Clarity, Simplicity, and Compactness**
- **Preference Given to Open-Source and Royalty-Free Formats**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Goals (Audio Format Only)

- **A Basic List of Audio Formats Meeting All Interchange Requirements**
  - **With Minimal Redundancy for the Sake of Clarity, Simplicity, and Compactness**
- **Preference Given to Open-Source and Royalty-Free Formats**
- **Ease of Adoption**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Goals (Audio Format Only)

- **A Basic List of Audio Formats Meeting All Interchange Requirements**
  - **With Minimal Redundancy for the Sake of Clarity, Simplicity, and Compactness**
- **Preference Given to Open-Source and Royalty-Free Formats**
- **Ease of Adoption**
- **Stability of Implementation**

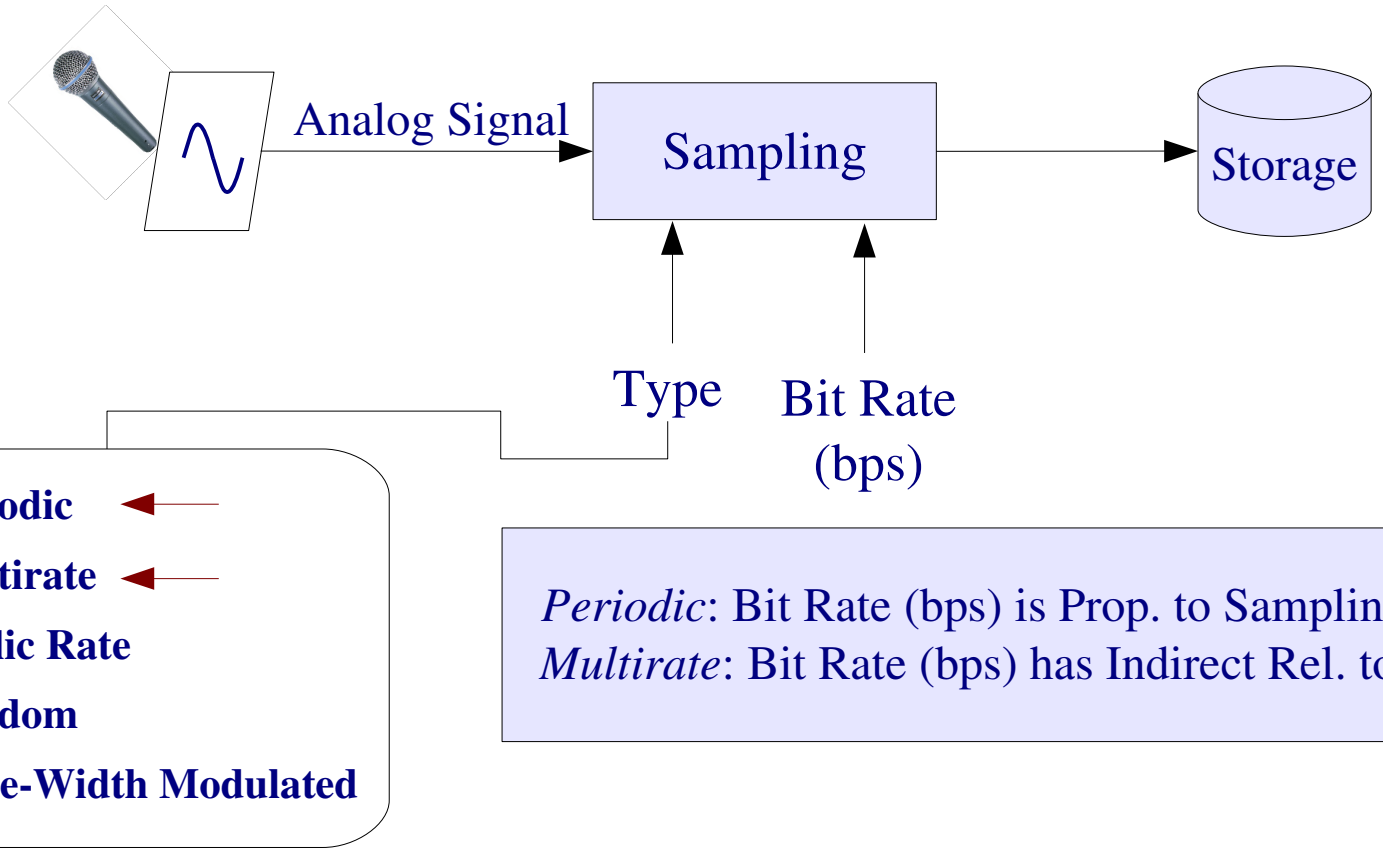
beigi@RecoTechnologies.com
judith@JMarkowitz.com

Goals (Audio Format Only)

- **A Basic List of Audio Formats Meeting All Interchange Requirements**
  - **With Minimal Redundancy for the Sake of Clarity, Simplicity, and Compactness**
- **Preference Given to Open-Source and Royalty-Free Formats** – *as Suggested by Kazuyuki*
- **Ease of Adoption**
- **Stability of Implementation**
- **Relative Quality – Compared to Contenders**

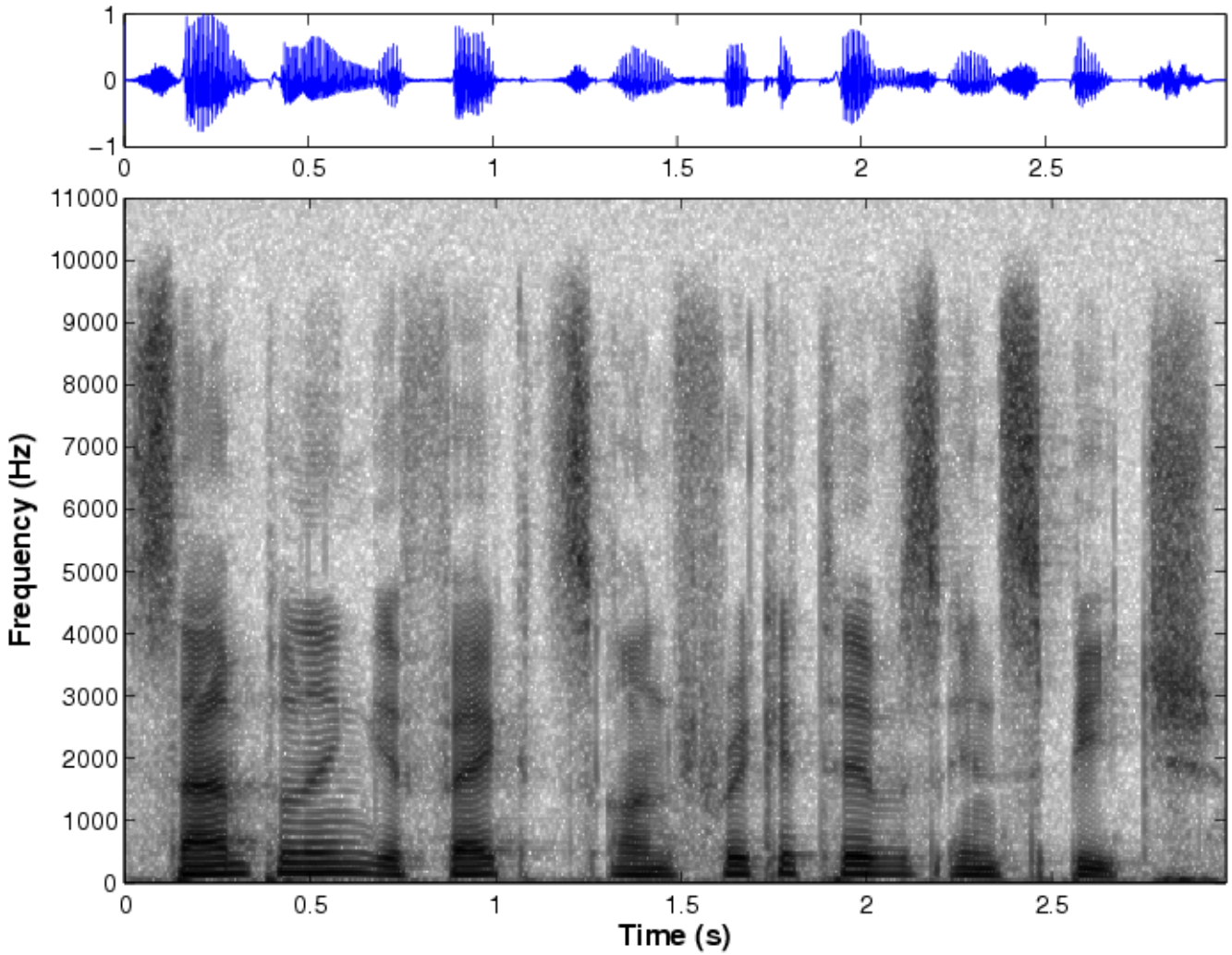
beigi@RecoTechnologies.com
judith@JMarkowitz.com

Sampling Process



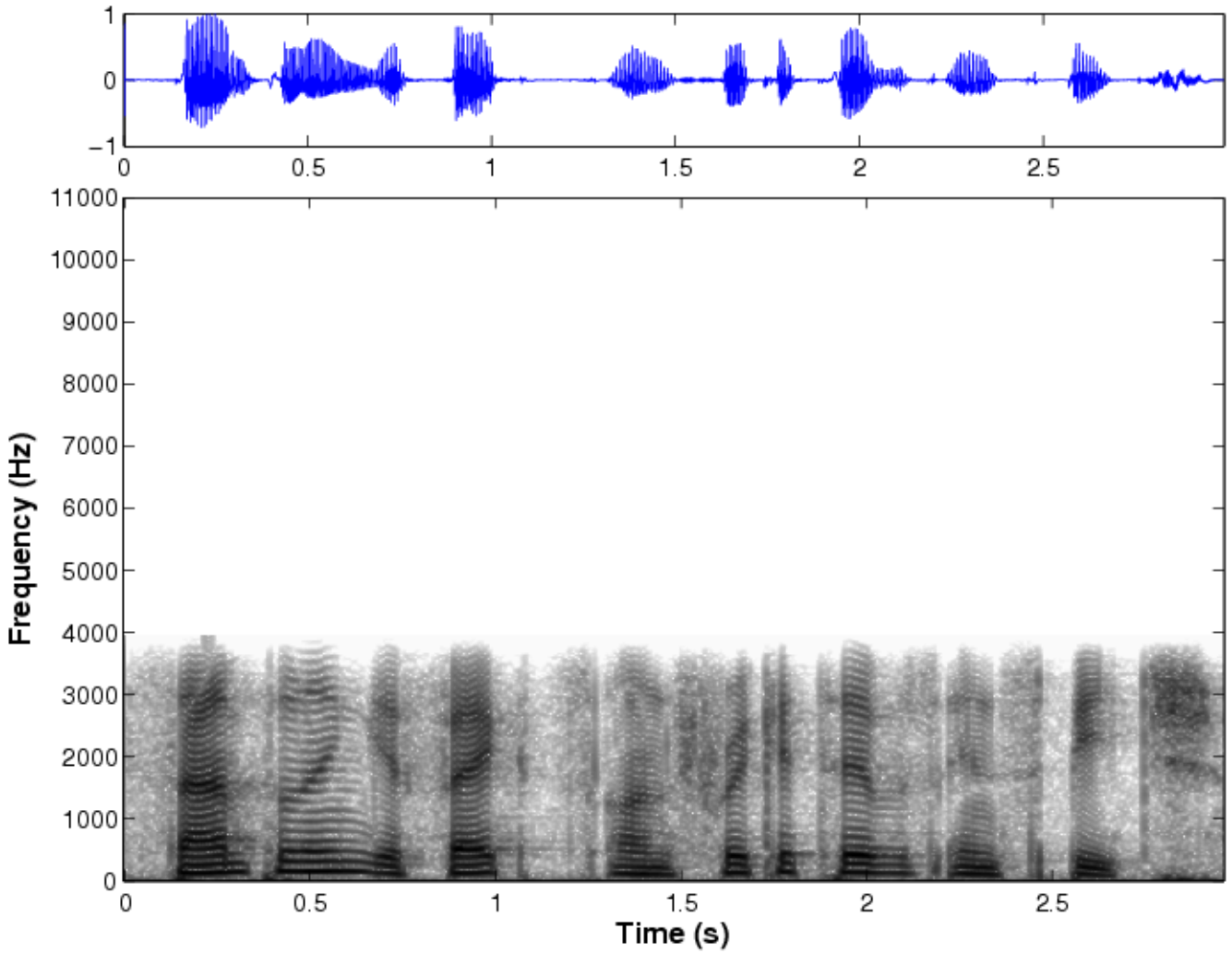
beigi@RecoTechnologies.com
judith@JMarkowitz.com

22kHz Sampling Rate



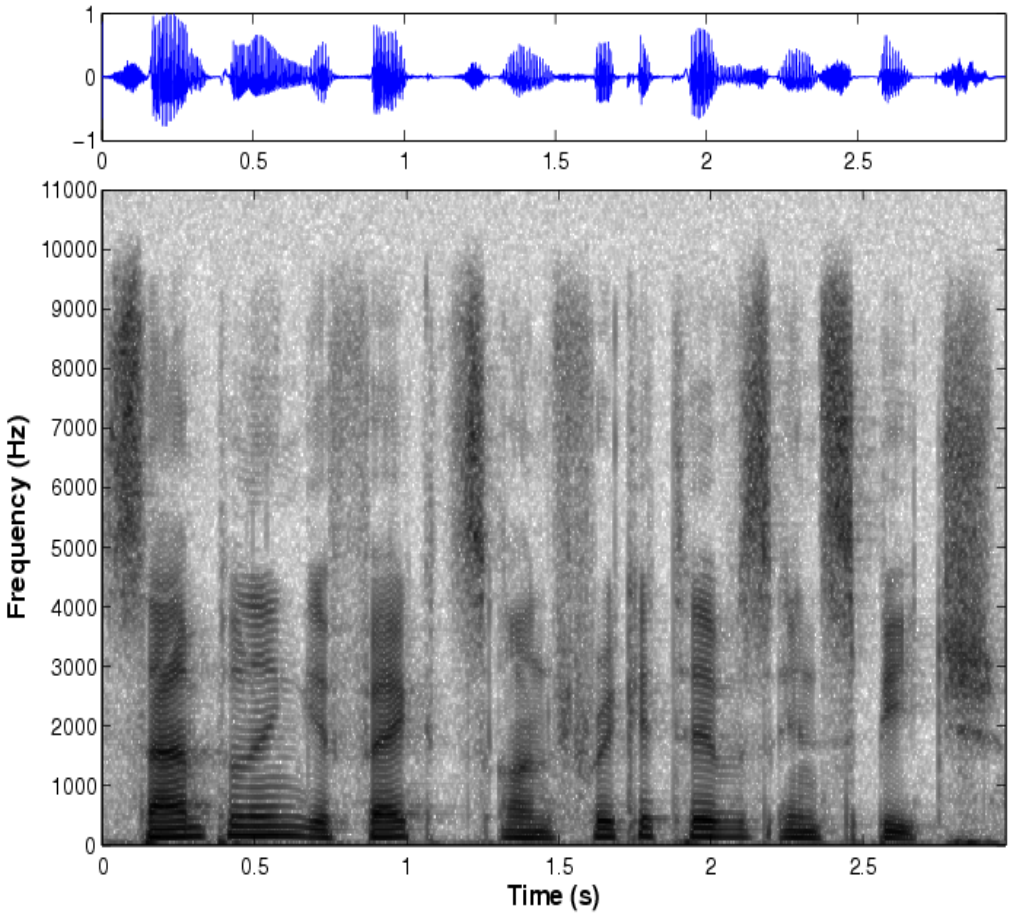
beigi@RecoTechnologies.com
judith@JMarkowitz.com

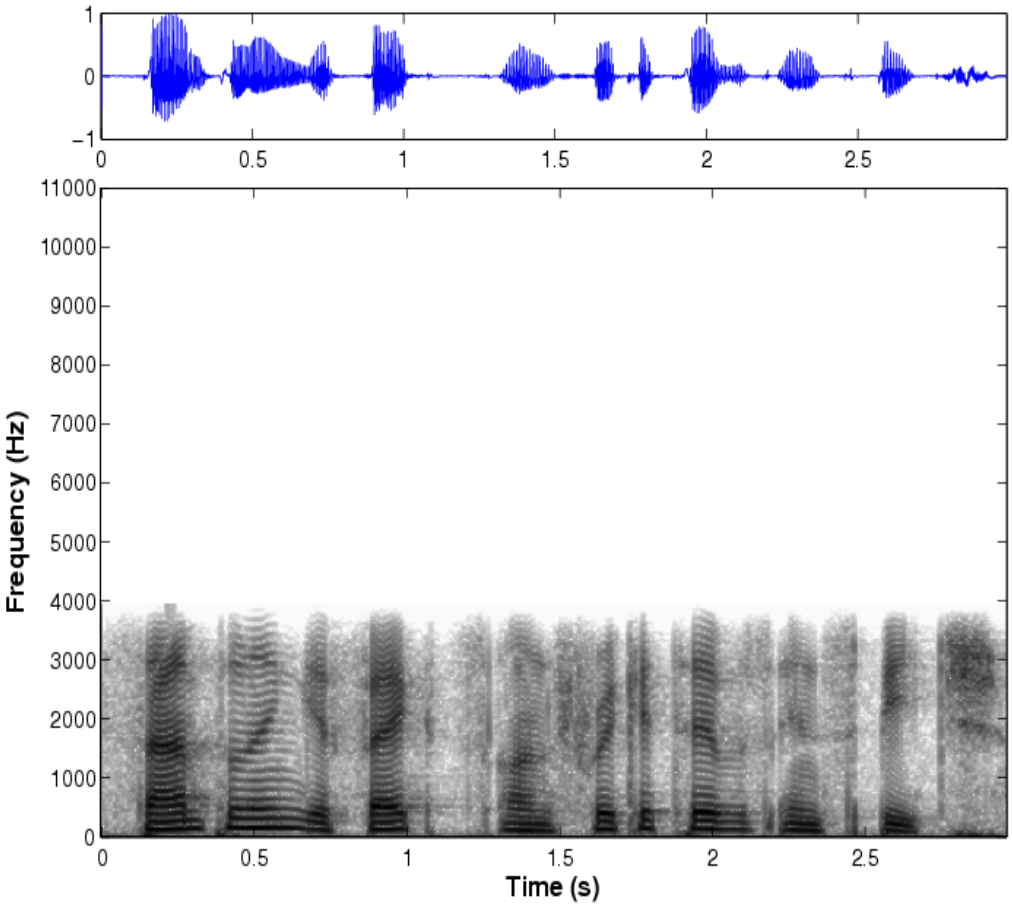
Band Limitation – 8kHz Sampling Rate



beigi@RecoTechnologies.com
judith@JMarkowitz.com

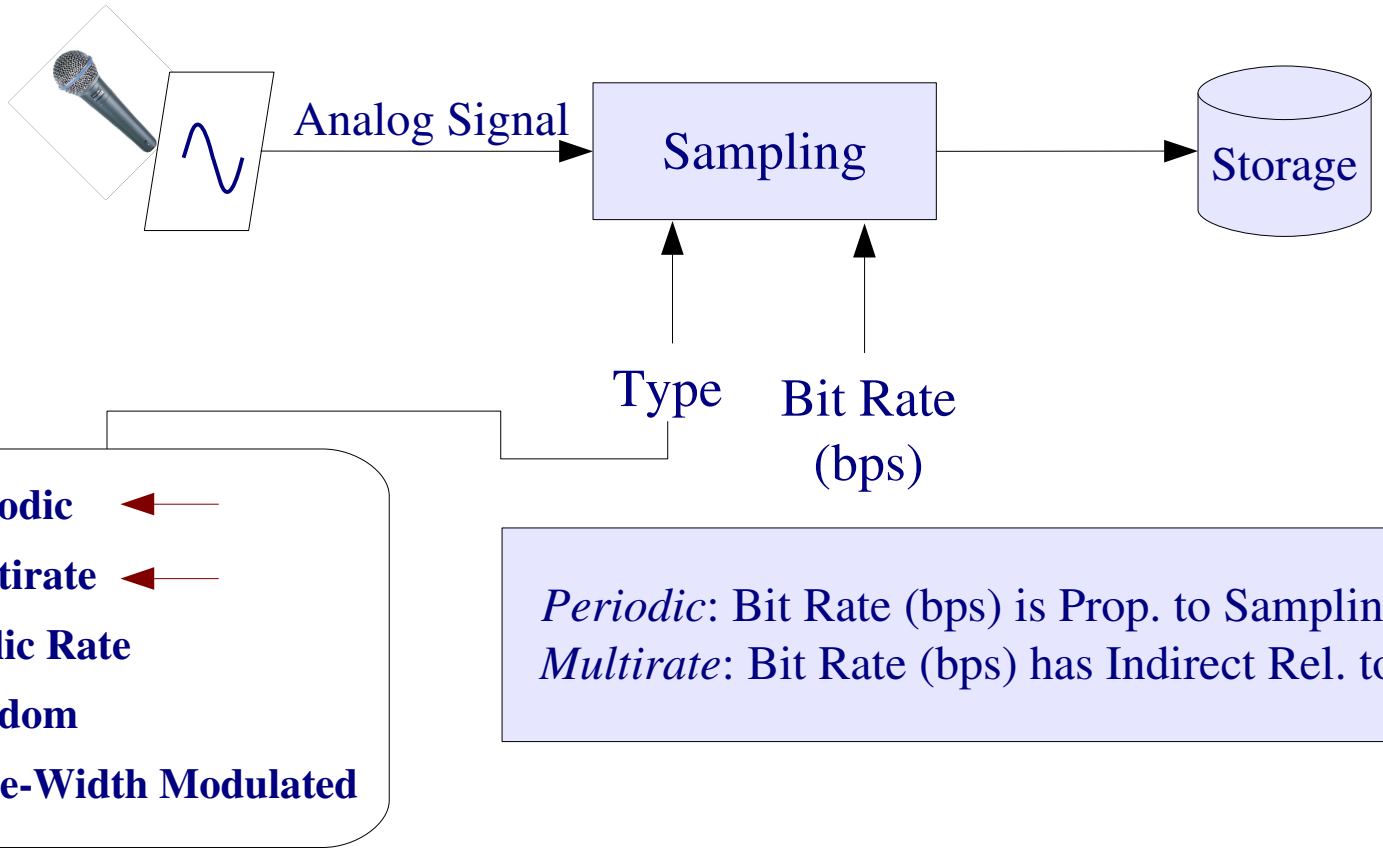
Band Limitation – Telephony (Landline)





beigi@RecoTechnologies.com
judith@JMarkowitz.com

Sampling Process



beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Coding Scenarios

- **Lossless Representation** – *Amplitude and Frequency are Unchanged*

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Coding Scenarios

- **Lossless Representation** – *Amplitude and Frequency are Unchanged*
- **Amplitude Compression** – *Freq. Stays the Same, Amplitude is Represented Nonlinearly*

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Coding Scenarios

- **Lossless Representation** – *Amplitude and Frequency are Unchanged*
- **Amplitude Compression** – *Freq. Stays the Same, Amplitude is Represented Nonlinearly*
- **Multirate Sampling** – *Aggressive Variable Bitrate Compression*

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Coding Scenarios

- **Lossless Representation** – *Amplitude and Frequency are Unchanged*
- **Amplitude Compression** – *Freq. Stays the Same, Amplitude is Represented Nonlinearly*
- **Multirate Sampling** – *Aggressive Variable Bitrate Compression*
- **Streaming** – *Usually includes multirate sampling and streaming*

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Interchange Scenarios

- **Lossless Representation**
  - ~~**Microsoft WAV Comes to Mind** – *A Wrapper which includes over 104 codecs*~~
  - **LPCM offers all that is needed** – *Just need to code the header information*

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Interchange Scenarios

- **Lossless Representation**
  - **LPCM offers all that is needed** – *Just need to code the header information*
- **Amplitud Compression**
  - **G.711 and G.711.1 ITU-T define PCMU and PCMA for 64, 80, and 96kbps**
  - ~~**ADPCM was considered, but it has many flavors and is not open source**~~

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Interchange Scenarios

- **Lossless Representation**
  - **LPCM offers all that is needed** – *Just need to code the header information*
- **Amplitud Compression**
  - **G.711 and G.711.1 ITU-T define PCMU and PCMA for 64, 80, and 96kbps**
- **Multirate Sampling**
  - ~~**MP3 comes to mind** – *Patent driven and certainly not an open standard*~~
  - **OGG Vorbis** – *Open Source and better quality as MP3 for the same bit rate*

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Interchange Scenarios

- **Lossless Representation**
  - **LPCM offers all that is needed** – *Just need to code the header information*
- **Amplitud Compression**
  - **G.711 and G.711.1 ITU-T define PCMU and PCMA for 64, 80, and 96kbps**
- **Multirate Sampling**
  - **OGG Vorbis** – *Open Source and better quality as MP3 for the same bit rate*
- **Streaming** – *Usually includes multirate sampling and streaming*
  - **OGG Media Stream** – *Open Source with capability of streaming different audio types*

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Interchange Scenarios

Quality	Format
Lossless	Linear PCM (LPCM)
Amplitude Compression	μ-law (PCMU) and A-law (PCMA)
Aggressive variable bit-rate compression	OGG Vorbis
Streaming	OGG Media Stream

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Format Header

Type	Variable	Description
U16	ByteOrder	The value is 0xFF00 and it is set by the audio file producer
U16	HeaderSize	Size of the header in bytes
Boolean	Streaming	This will 0 for non-streaming and 1 for streaming. This boolean variable is redundant since the AF_FORMAT for streaming audio is greater than 0x0FFF. However, it is used for convenience.
U64	FileLengthInBytes	In Bytes not including the header
U64	FileLengthInSamples	In Number of samples
U16	AudioFormat	See AF_FORMAT macros
U16	NumberOfChannels	Number of channels, *N.B.*, Channel data alternates
U32	SamplingRate	Sampling rate in samples per second – This is the audio sampling rate and not necessarily the sampling rate of the carrier which may be variable.
U64	AudioFullSecondsOf	It is the truncated number of seconds of audio
U32	AudioRemainderSamples	This is the number of samples of audio in the remainder which was truncated by the above variable
U16	BitsPerSample	Number of bits per sample, may be 0 for formats which use variable bits

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Audio Interchange Scenarios

Macro	Value
AF_FORMAT_UNKNOWN	0x0000
AF_FORMAT_LINEAR_PCM	0x0001
AF_FORMAT_MULAW	0x0002
AF_FORMAT_ALAW	0x0003
AF_FORMAT_OGG_VORBIS	0x0004
AF_FORMAT_OGG_STREAM	0x1000

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Conclusion

- **Are There any Interchange Requirements Not Covered?**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Conclusion

- **Are There any Interchange Requirements Not Covered?**
- **Are There any Important Features Missing in General?**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Conclusion

- **Are There any Interchange Requirements Not Covered?**
- **Are There any Important Features Missing in General?**
- **Are There any Formats that will Lose Important Features when Converted?**

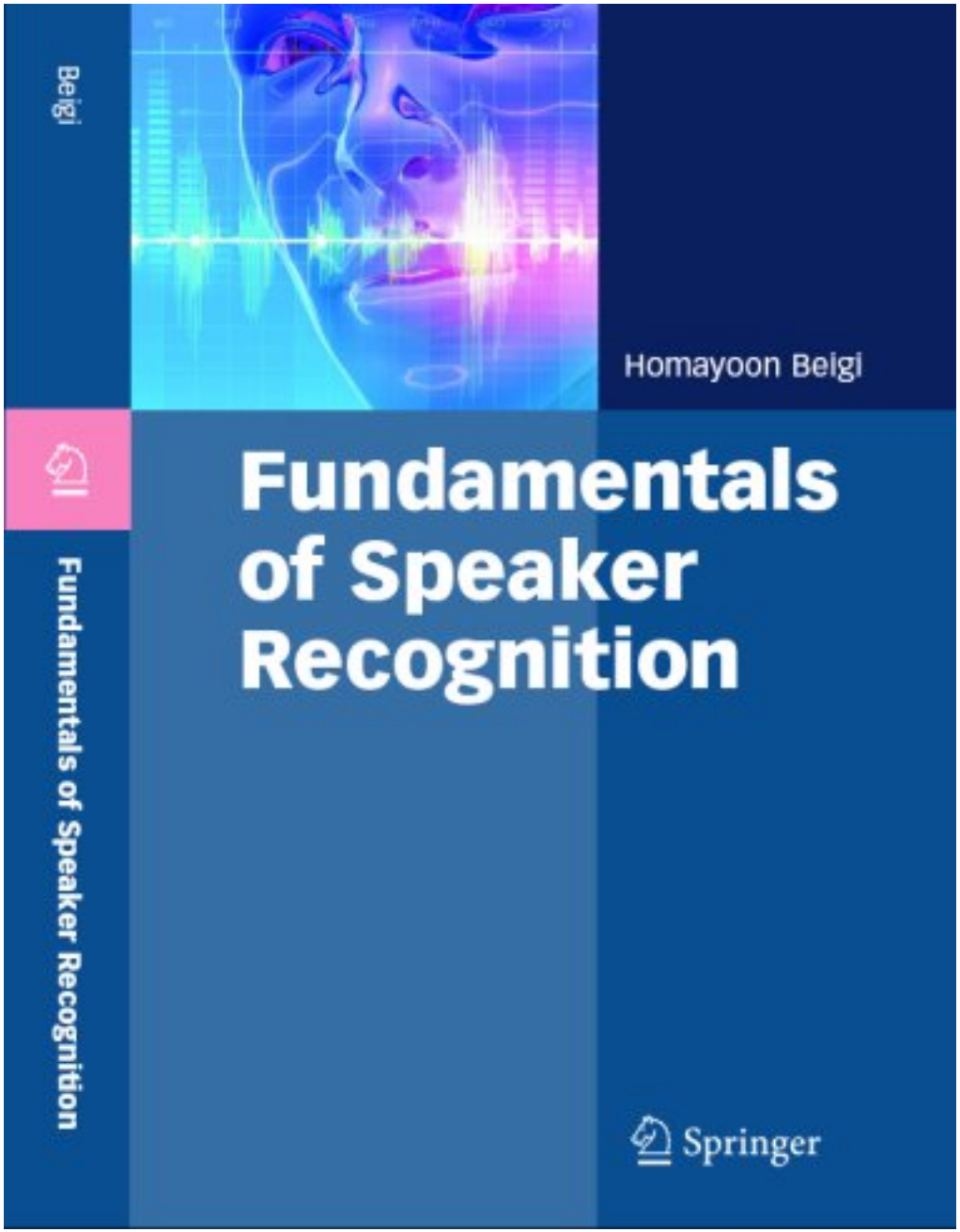
beigi@RecoTechnologies.com
judith@JMarkowitz.com

Conclusion

- **Are There any Interchange Requirements Not Covered?**
- **Are There any Important Features Missing in General?**
- **Are There any Formats that will Lose Important Features when Converted?**
- **Any other Compelling Reasons to Add more Formats to the Supported List?**
  - **Please! "Popularity" is no Reason!**

beigi@RecoTechnologies.com
judith@JMarkowitz.com

Fundamentals of Speaker Recognition
Out in April 2009 – Springer



A Standard Audio Encapsulation Method

Homayoon Beigi
Beigi@RecognitionTechnologies.com
http://www.RecognitionTechnologies.com
of
Recognition Technologies, Inc.
300 Hamilton Avenue
White Plains, NY, U.S.A.

and

Judith Markowitz
Judith@JMarkowitz.com
http://www.JMarkowitz.com
of
J. Markowitz Consultants
5801 N. Sheridan Road
Chicago, IL, U.S.A.