Speech Compression for ASRs in Voice Search Applications

Veeru Ramaswamy, PhD

CTO, Vianix LLC

Email: veeru@vianix.com

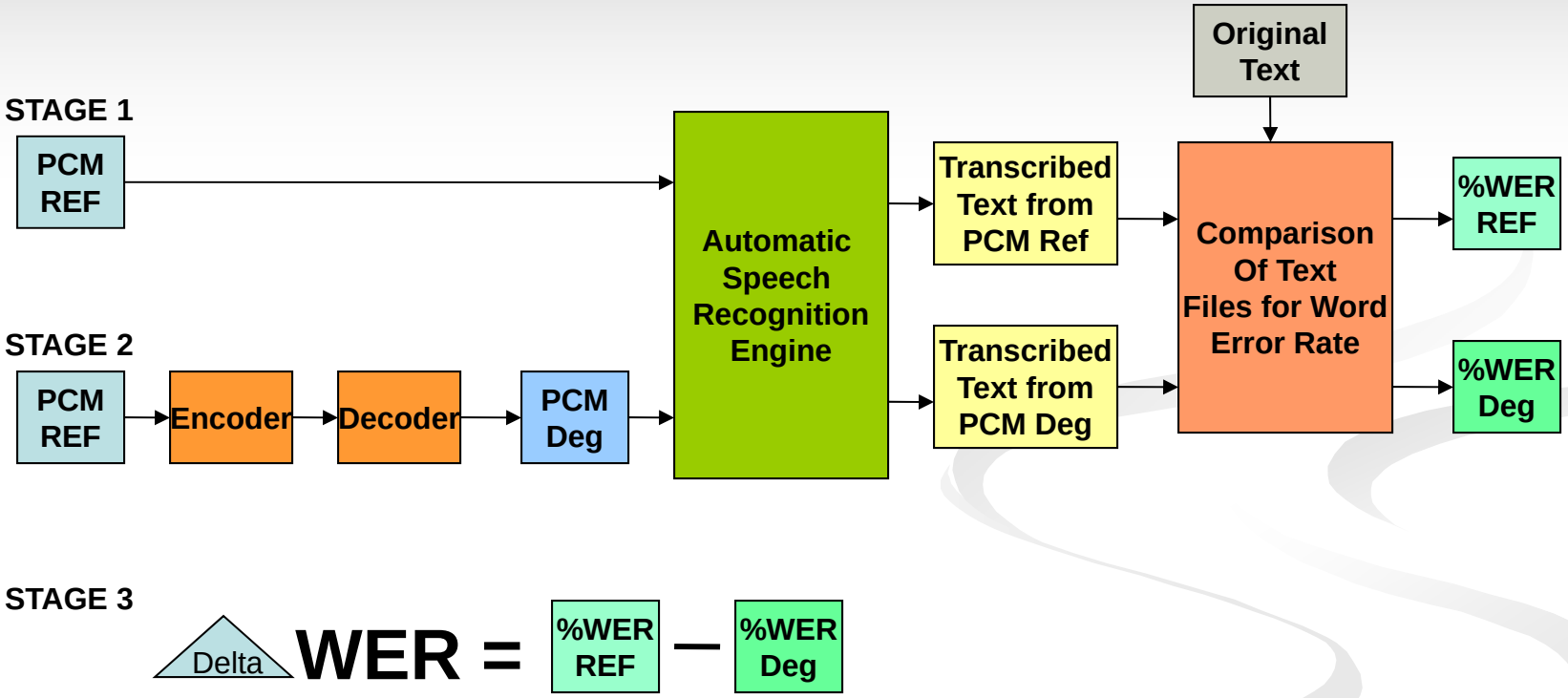
Vianix Background

- Fast-paced speech technology company with corporate headquarters located in Virginia Beach, Virginia.
- Vianix has developed, tested, proven and licensed MASC®
  - Managed Audio Sound Compression (MASC ®)
  - State-of-the-Art speech compression technology
    - High performance enabling voice technology
    - For a broad spectrum of healthcare, multimedia communications and enterprise applications

Metrics

- **Variable Bit-Rate:**
  - Bit-rates range from almost 5 kbps to 20 kbps.
- **MIPS:**
  - MIPS ranges from 20 to about 200 depending on the codec used.
- **PESQ** (Perceptual Evaluation of Speech Quality):
  - Defined as in ITU P.862
- **WER:**
  - A measure to compute the number of words in percentage that have NOT been correctly identified by an ASR.
- **DWER:**
  - difference in WER from the original uncompressed PCM samples to decompressed/decoded PCM samples.

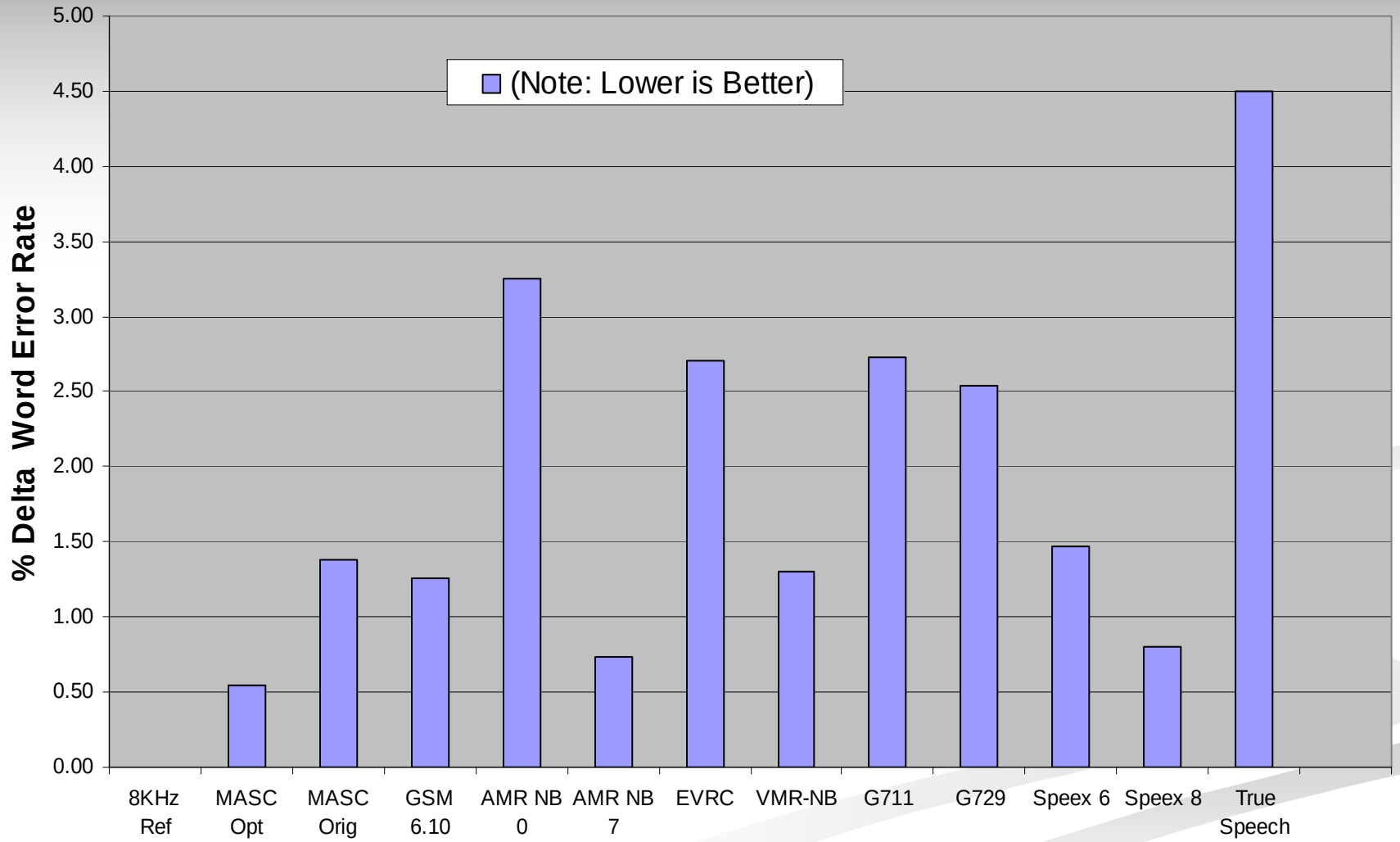
Signal Train for DWER Calculation



Comparison of 8KHz Codecs on ASR1

Codec	WER	Absolute DWER	PESQ	bit rate
8KHz Reference	14.9143		4.50	128.00
MASC Optimized	15.4569	0.5425	4.13	17.20
MASC Original	16.2902	1.3759	3.93	14.40
GSM 6.10	16.1723	1.2580	3.72	13.00
AMR NB setting 0	18.1710	3.2567	3.37	4.75
AMR NB setting 7	15.6474	0.7331	4.04	12.20
EVRC	17.6189	2.7045	3.55	8.00
VMR-NB	16.2121	1.2977	3.81	14.20
G711	17.6391	2.7248	3.44	64.00
G729	17.4506	2.5363	3.70	8.00
Speex setting 6	16.3804	1.4660	3.78	11.50
Speex setting 8	15.7137	0.7993	4.00	15.00
True Speech	19.4106	4.4962	3.41	8.00

Comparison of 8 KHz Codecs on ASR1

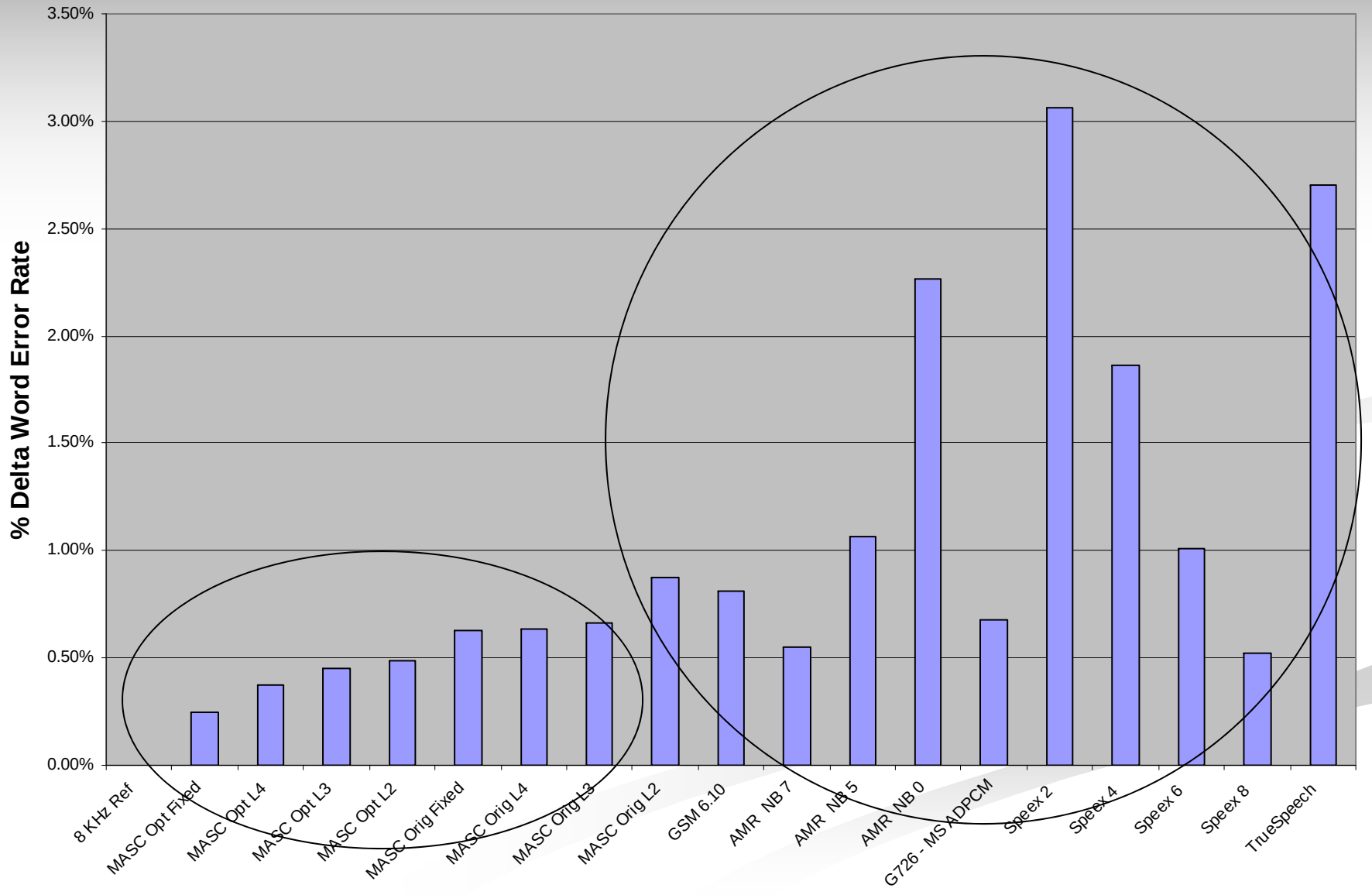


MASC is the only Codec that exists today at 8 KHz and at a ADWER in the 0.5 range

Comparison of 8KHz Codecs on ASR2

Codec	WER	Absolute DWER	Bit rate	PESQ
8 KHz Ref	7.93		128.0	4.50
MASC Opt Fixed	8.18	0.25	17.6	4.13
MASC Opt L4	8.30	0.38	15.8	3.97
MASC Opt L3	8.38	0.45	14.8	3.80
MASC Opt L2	8.41	0.49	12.7	3.65
MASC Original Fixed	8.55	0.62	14.4	3.98
MASC Original L4	8.56	0.64	12.9	3.93
MASC Original L3	8.59	0.66	12.2	3.82
MASC Original L2	8.80	0.87	10.5	3.61
GSM 6.10	8.74	0.81	13.0	3.72
AMR NB 7	8.48	0.55	12.4	4.04
AMR NB 5	8.99	1.06	8.4	3.79
AMR NB 0	10.19	2.26	5.2	3.37
G726 - MS ADPCM	8.60	0.67	32.8	3.68
Speex 2	10.99	3.06	6.4	3.23
Speex 4	9.79	1.86	8.9	3.48
Speex 6	8.94	1.01	11.7	3.78
Speex 8	8.45	0.52	15.7	4.00
TrueSpeech	10.63	2.70	8.5	3.41

Comparison of 8KHz Codecs on ASR2



Effect of High Quality Compression in Two Markets

- Speech Analytics:
  - Specific to Call / Contact Center recordings and speech analytics
- Mobile Voice Services
  - Applications such as Mobile Voice Search
    - Location based search
    - Multimedia (Audio/Video) based search
  - Mobile Voice based SMS/e-mail (mobile transcription)

Speech Analytics Solutions

- LVCSR: Compiles a list of commonly used key terms that allows users to find words more accurately
- Phonetics: Separate spoken words are split into phonemes and allow users to run a search for any word they want to find

MASC Benefits for Call Recording and Speech Analytics

- ASR Optimized MASC® Codec for G.711 or PCM Recordings
  - Virtually No Impact on Analytics Accuracy (FOM Scores) vs. Original PCM Uncompressed Speech / G.711
    - Verified with Industry Leaders
    - Other codecs degrade the accuracy of Speech analytics engines while MASC improves
- G.723/G.726/G.729 Codec Artifact Mitigation for Legacy/Existing Recordings
  - Improves Speech Analytics Accuracy
  - Solution: Combination of MASC and Proprietary Filtering Techniques
    - MASC Performs Background Noise Cancellation
    - Filtering Performs G.726 Artifact reduction
    - Various Parameters to Tune to a Specific Speech Analytics Engine
    - Additional Parameters to Tune Based on Length of Phonemes in Search
    - Yet to verify the G.723/G.729 solutions with more R&D

MASC on G.711/PCM

Customer

G.711

PSTN

G.711

CSR

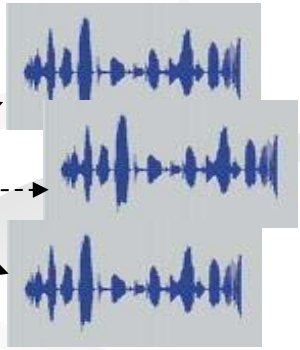
G.711

Call Recorder

MASC

Speech Analytics Engine
(LVCSR or Phonetics)

Voice Output

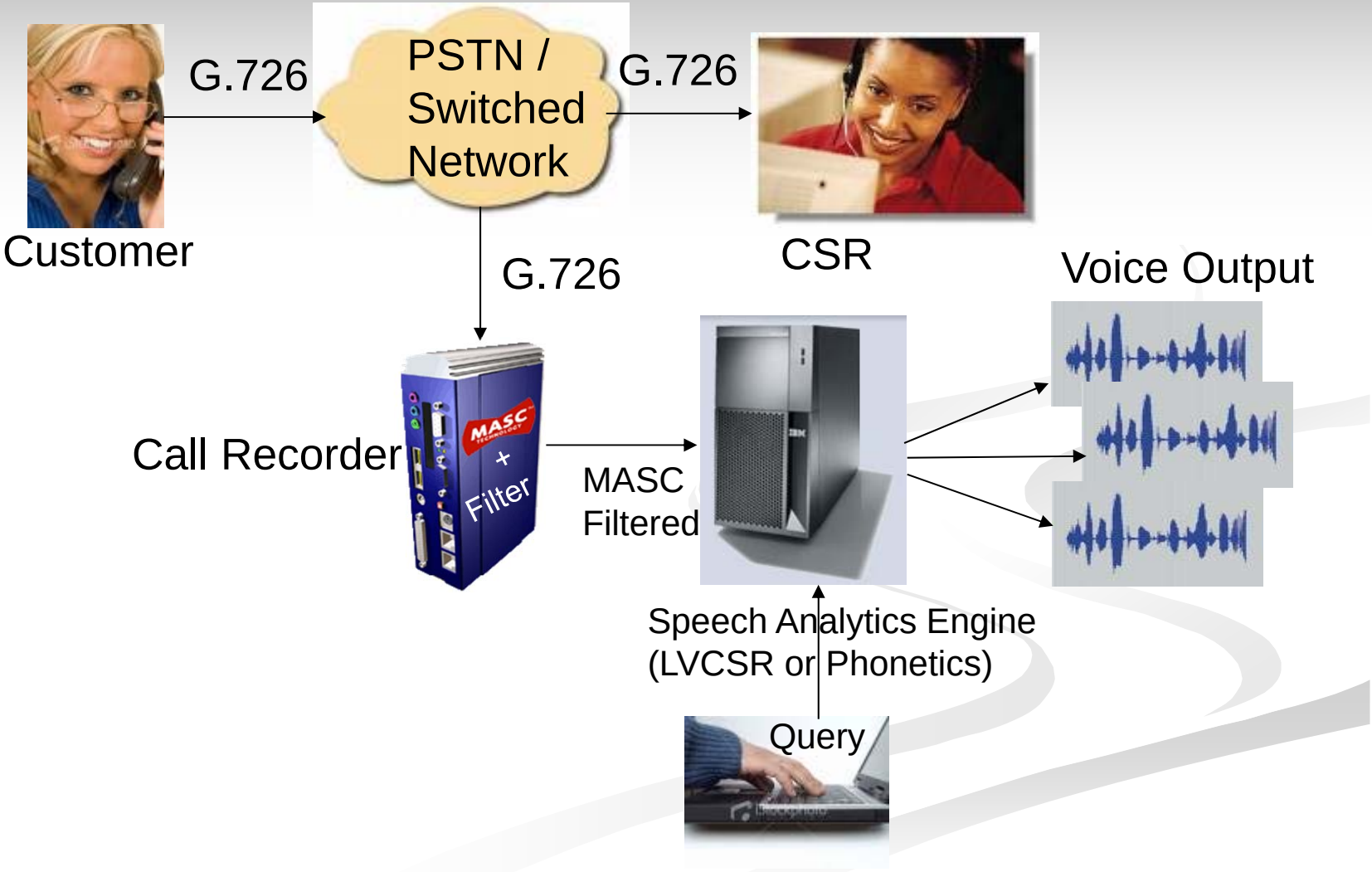


Query

------ Existing Markets without MASC

—— MASC improves Speech Analytics Accuracy

MASC with Filtering on G.726



Results of Codec Testing for Speech Analytics

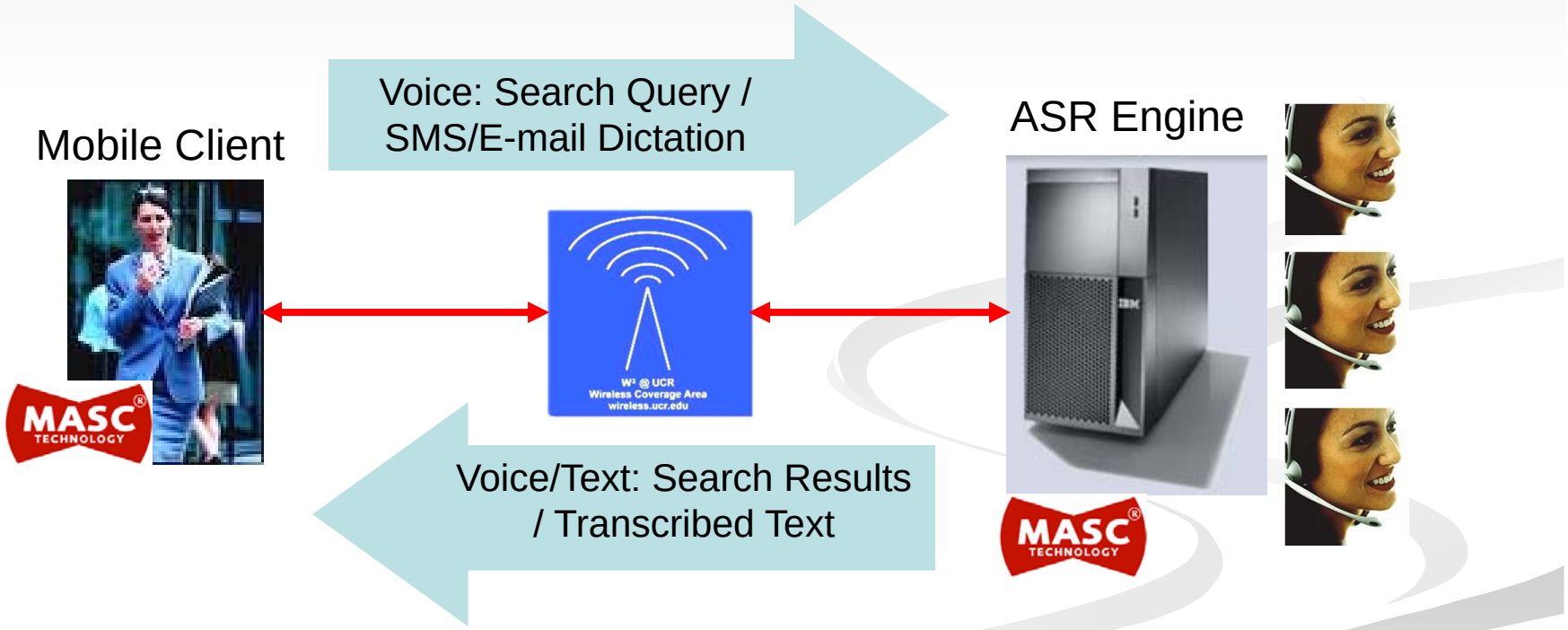
Test Condition	FOM 7-11 (%)	FOM 12-16 (%)	FOM 17-21 (%)
Hub4 8kHz data (for comparison)	**35.96**	**64.95**	**82.69**
Hub4 G726 (for comparison)	**23.49**	**50.97**	**70.24**
MASC-Filter10_3-10-32	21.72	55.09	71.48
MASC-Filter10_3-10-64	22.97	56.04	73.22
MASC-Filter10_3-10-96	22.98	55.67	74.16
MASC-Filter10_10-10-32	22.07	54.72	72.76
MASC-Filter10_10-10-64	22.86	54.98	74.02
MASC-Filter10_10-10-96	**22.84**	**56.47**	**76.02**

- Produces Dramatic Increase in Figure Of Merit (FOM) Scores about 10% for Speech Analytics Engine In Evaluations with Industry Leaders
- Further Enhancements can be achieved with Specific Analytics Engine Adaptation with MASC-Filtered Signals.

Summary of the Results for Speech Analytics

- There were three phoneme sizes of speech data sets (small, medium and large) on which the testing was conducted.
- MASC processing on G.711/PCM signals can produce just about the same FOM scores as the G.711/PCM.
- For G.726 signals, observed that neither MASC alone nor filtering alone improved the FOM scores.
- The MASC+Filtering (in that order) scheme helped to improve the FOM score significantly over the G.726 signals.
- On the smaller phoneme set, the combination of MASC+Filtering did not improve as high as the Medium and Large phoneme query sets.

Mobile Voice Service (MVS)



MVS Market

- Global mobile search market will generate $11 billion of revenue by EOY 2008 (Piper Jaffray)

- Segments
  - Voice Channel
  - Data Channel

- Usage
  - Voice dialing accounts for 44 percent of usage, SMS and email dictation represent nearly 25 percent of usage, and Web Search (including finding a business, stock quotes, weather, Web navigation and search) represents 22 percent

MASC Contrasted with Alternatives for MVS

	No Compression	Native Codec	DSR	MASC
Response Time	Slow	Fast	Fast	**Fast**
Accuracy	Good	Poor	Good	**Best**
Intelligibility	Good	Poor	None	**Best**
QA/Tunability	Good	Poor	None	**Best**

Smaller files => quicker response times
Low DWER => better ASR accuracy

High voice quality => high intelligibility
Retain voice => QA and tunable

Results of Codec Testing for MVS

	Bit Rate (Kbps)	Compress Test only Rel (WERI vs baseline)	Compress adaptation data & test Rel (WERI vs basline)
Test case : High-noise test (30667 words)			
Codec A - MASC	12.3	4.5	1.9
Codec B		8.8	9.2
Codec C		5.3	6.6
Codec D		7.1	6.1
Test case : Med-noise test (9128 words)			
Codec A - MASC	12.3	4.6	5.0
Codec B		5.4	6.6
Codec C		14.7	12.2
Codec D		14.4	11.6
Test case : Low-noise test (6179 words)			
Codec A - MASC	12.3	10.7	0.6
Codec B		22.6	17.9
Codec C		26.3	16.9
Codec D		22.2	13.7

Summary of the MVS Testing

- Evaluated 4 competitive codecs
  - Compared recognition accuracy of the audio passed through the codec Vs. Baseline audio
- Compress Test Only:
  - measure impact of compressing only test data
- Compress Adaptation and Test:
  - measure impact of compressing both enrollment and test data
- The difference between the two can be a measure of consistency, and could point to a potential impact reduction if entire training corpus is passed through the codec
- All codecs under test had a bit rate between 10 and 14 kbps, on 8kHz audio. In all tests, the audio was collected on mobile handsets as wideband high-quality audio recorded originally as ulaw or 16-bit linear.

Conclusion

- Traditional Codecs are good for communications but not good enough for ASR based engines such as Transcription, Speech Analytics, Mobile Voice Search or even Voice Biometrics.
- High quality reconstruction of voice signals are needed for high accuracy in ASR engines for these applications.
- MASC has been well positioned to exploit these markets with HQ performance over and above the traditional advantages of savings in bandwidth, storage etc., for communications.


MASC
TECHNOLOGY
Bringing your products
one step closer...
TO THE FUTURE