



Right information... Right place... Right time



Advanced Speech Encoding (ASE)

Program Overview

Proposer's Day Conference

7 June 2004



Advanced Speech Encoding (ASE)



Right information... Right place... Right time

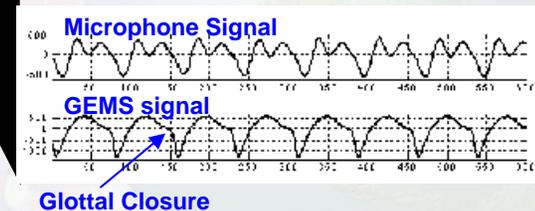
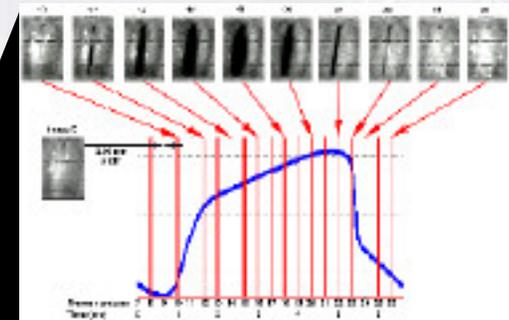
Program Goal:

A 300 bps vocoder with acceptable intelligibility, quality, and speaker recognizability in harsh military environments

Approach:

Main Thrust Area: Combine a multi-sensor approach, including noise-immune sensors that provide excitation information, with ultra-low-bit-rate coding approaches.

Second Thrust Area (Scientific Discovery): Explore the potential utility of subvocal speech as an alternative means of communication in acoustically challenging environments.



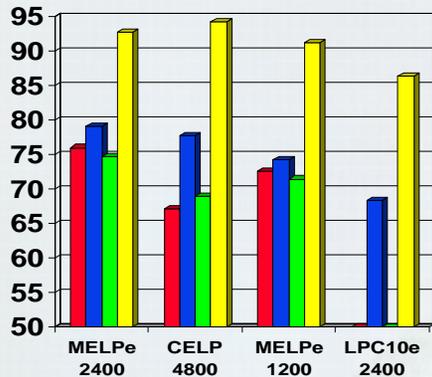


Big-Picture Programmatic Approach



Right information... Right place... Right time

2002 DDVPC* DRT Results

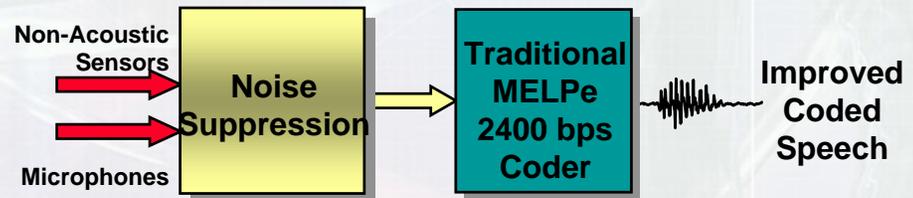


■ HMMWV
 ■ UH60
 ■ M2
 ■ Quiet

*DDVPC: U.S. Department of Defense Digital Voice Processor Consortium

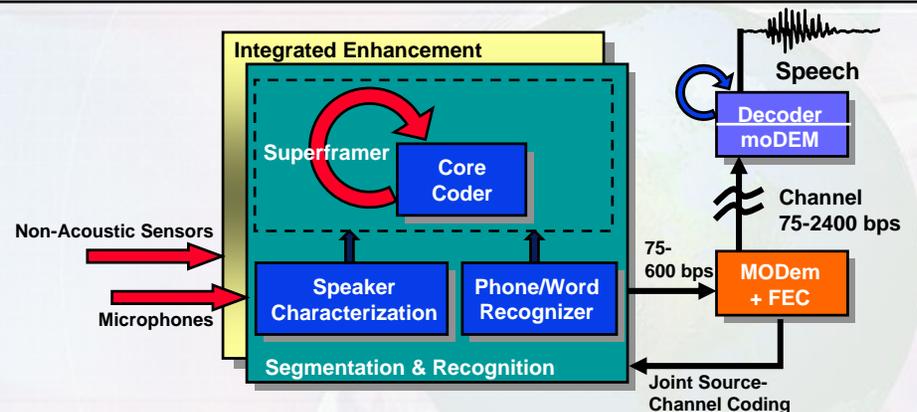
Phase 1

Bridge the intelligibility gap by using noise-immune sensors to enable high quality coded speech for current vocoders in harsh noise environments



Phase 2a

Shift the vocoding paradigm to achieve acceptable intelligibility, quality and recognizability at ultra-low (300 bps) bit rates



Phase 2b

Prototype demonstration of a multi-sensor system integrated with an ultra-low rate coder



ASE: Development Driven by Realistic Testing



Right information... Right place... Right time

- Objective measures (SNR, spectral distortion) do not capture speech intelligibility or quality
- ASE uses DoD/NATO subjective intelligibility and quality tests to demonstrate goals
 - Diagnostic Rhyme Test for intelligibility
 - Forced Choice quality testing for speech quality
- Program-sponsored effort to drive realistic development:
 - Developmental collections and final testing in high noise (UH60, M2, MOUT) environments

Acoustic Noise Conditions			
Condition	Field SPL dBC	Presentation SPL dBC	Presentation SPL dBA
Office Environment	55	56	45
MCE Environment	83	79	66
M2 Environment (High)	114	114	104
M2 Environment (Low)		74	64
MOUT Environment (High)	113	113	98
MOUT Environment (Low)		73	58
UH-60 Environment (High)	112	110	98
UH-60 Environment (Low)	95	70	65

1. Office at Standard SPL
2. MCE (Mobile Command Enclosure) at Standard SPL
3. M2 Bradley Fighting Vehicle at High Noise State SPL
4. M2 Bradley Fighting Vehicle at -40 dBC SPL from High Noise State SPL
5. MOUT (Military Operations on Urbanized Terrain) at High Noise SPL
6. MOUT at -40 dBC SPL from High Noise SPL
7. UH-60 Black Hawk Helicopter at Cruise SPL
8. UH-60 Black Hawk Helicopter at Idle (-40 dBC SPL from Cruise SPL)



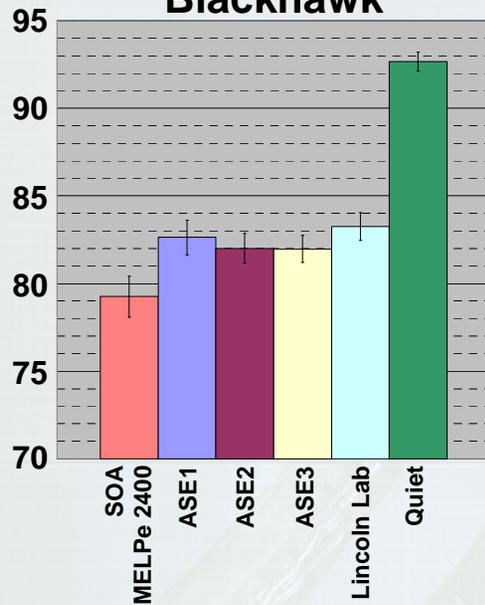
Phase 1 Test Results



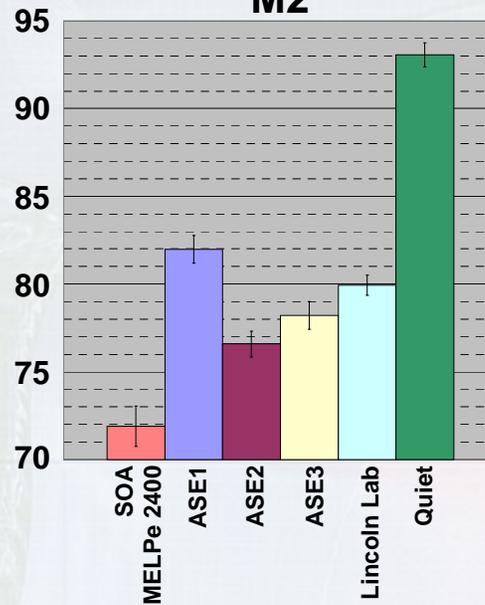
Right information... Right place... Right time

Intelligibility Test Results

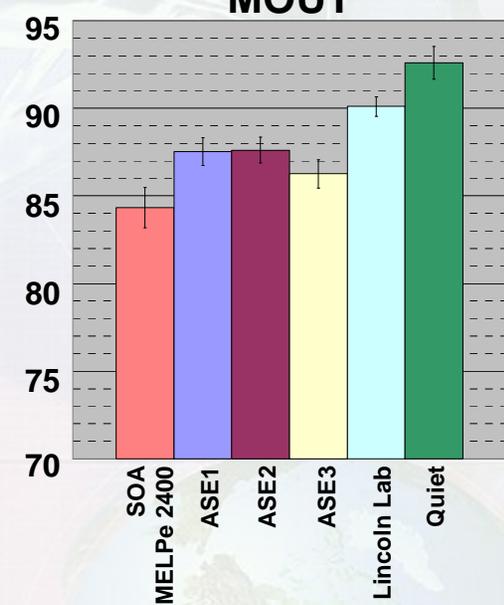
Blackhawk



M2



MOUT



Quality Test Results

Pair Comparison Results			
Percent Preference for Participant over CELP			
Environment	ASE1 (2400 bps)	ASE1 (1200 bps)	ASE3 (1500 bps)
M2	74.31%	70.49%	62.85%
Blackhawk	70.83%	60.76%	65.63%
MOUT	62.33%	54.86%	51.91%

Acoustic Conditions

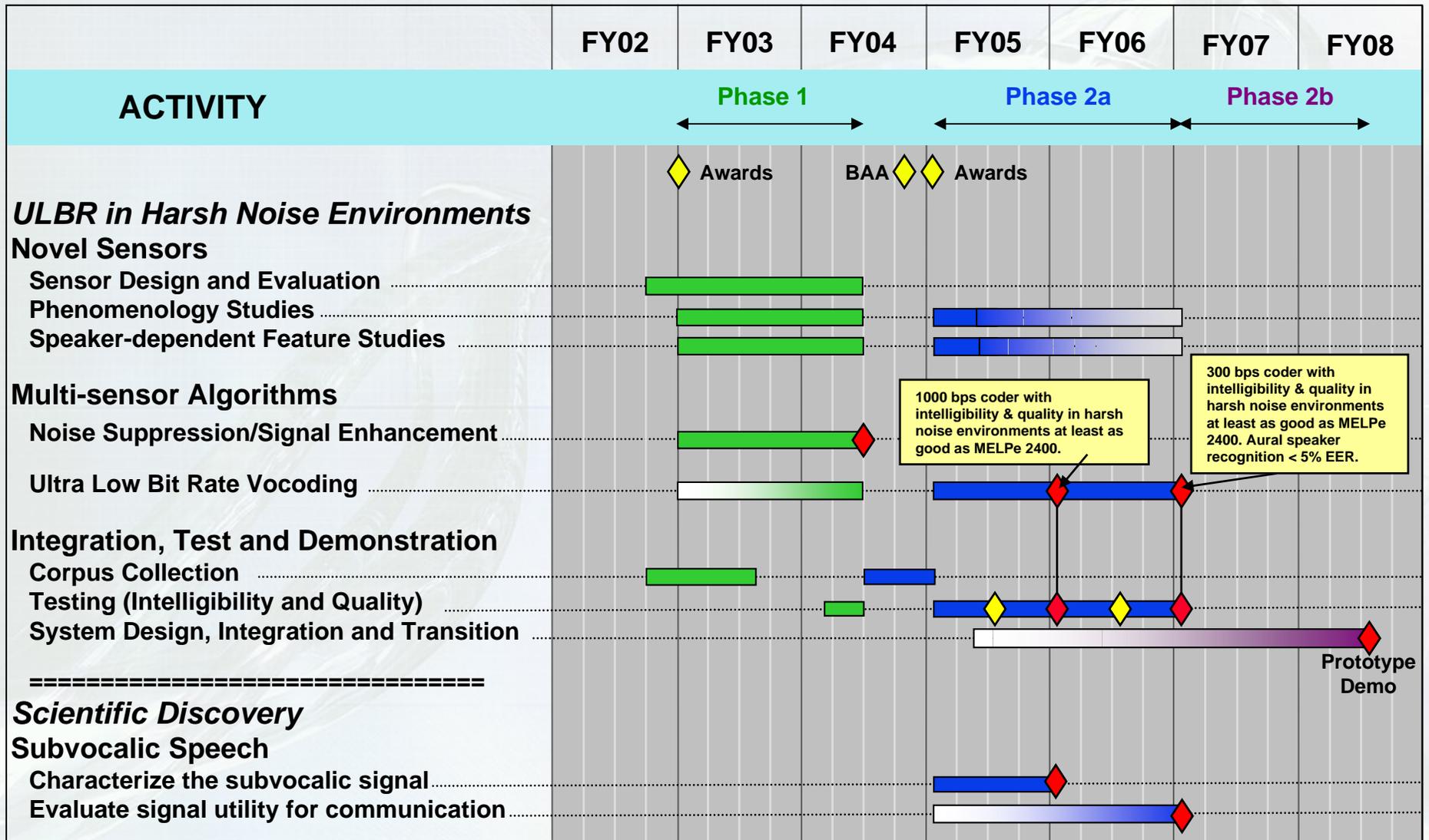
Maximum Noise Levels Achieved During Phase 1 Corpus Collection	
Environment	Noise Level (dBA)
M2	114
Blackhawk	110
MOUT	113



Schedule/Milestones



Right information... Right place... Right time





Ultra-low Bitrate Coding

Go/No-Go Decision Criteria



Right information... Right place... Right time

Milestone	Months After Program Kick-off	Criteria
1	12	Demonstrate = 1000 bps vocoder with intelligibility and quality in harsh noise environments that is at least as good as that of today's MELPe 2400 in harsh (> 110 dBC) military noise environments.
2	24	Demonstrate = 300 bps vocoder with intelligibility and quality in harsh (> 110 dBC) noise environments that is at least as good as that of today's MELPe 2400. Additionally, aural speaker recognition must achieve an EER < 5%.
3	42	Verify the performance attained in laboratory tests in a field demonstration of a prototype communications system



Ultra-low Bitrate Coding

Program Waypoints



Right information... Right place... Right time

Waypoint	Test/Data	Intent
6-month	Measure intelligibility and quality in harsh noise environments for a small speaker set at = 1000 bps	<ul style="list-style-type: none">- Strike a development baseline- Determine development emphasis- Provide input for potential data collection requirements
12-month	In addition to intelligibility and quality (the 12-month milestone), coder delay and aural speaker recognizability will be measured at = 1000 bps	<ul style="list-style-type: none">- Strike a development baseline for the 300 bps coder- Identify potential trades between bitrate and coder delay- Identify potential trades between bitrate and aural speaker recognizability
18-month	Measure intelligibility, quality, coder delay, and aural speaker recognizability in harsh noise environments for a small speaker set at 600 bps	<ul style="list-style-type: none">- Assess risk in scaling to 300 bps- Identify specific weaknesses



Subvocalic Voice Encoding

Go/No-Go Decision Criteria



Right information... Right place... Right time

Milestone	Months After Program Kick-off	Criteria
1	12	<ol style="list-style-type: none">1) What is the physiological source of the subvocal signals?2) How do the signals vary with speaker?3) What are the sources of noise for this signal (e.g., swallowing, sensor placement), and how do they impact the signal-to-noise ratio (SNR)?4) How does the signal vary with speaker state (e.g., stress, fatigue)?5) What is the impact of acoustic noise on the signal?
2	24	<ol style="list-style-type: none">6) Are the signals distinguishable for words that are acoustically similar? (e.g., DRT word pairs)7) Can the signals be differentiated at the phonemic level?8) Can the signals be extracted from continuous speech?9) What type of vocoding is most appropriate for these signals?10) Do these signals contain sufficient information to encode continuous speech, or do they need to be augmented by other information?
3	42	None (if answers to above questions are favorable, a separate program will spin off after 24 months)



Teaming



Right information... Right place... Right time

- **Thrust Area #1: ULBR vocoding:**
 - **Teaming is strongly encouraged. Teams should include expertise in:**
 - **Novel, noise-immune sensors**
 - **Novel, ultra-low bitrate algorithms**
 - **Integration of hardware and software into a prototype**
 - **DARPA is not looking for solutions that employ only acoustic sensors**
 - **DARPA is not interested in approaches that will not work in the real (noisy) world.**
- **Thrust Area #2: Scientific Discovery in Subvocalic Speech**
 - **Teaming is strongly encouraged. Teams should include expertise in:**
 - **The physiology of laryngeal and sublingual myoelectric signals**
 - **Robust sensors that can measure those signals**
 - **Extraction of robust, exploitable features from those signals**

An offeror may propose to both thrust areas



Tentative Solicitation Schedule



Right information... Right place... Right time

Proposers Day Conference	June 7, 2004
BAA Release	June 21, 2004
Proposals Due	August 13, 2004
Proposal Evaluation Completed	August 31, 2004
Source Selection Authority Brief	September 6, 2004
Contract Negotiation Period	
Contract Award(s)	mid-October, 2004



Right information... Right place... Right time

QUESTIONS?

After today, email questions to:

ase@darpa.mil

Answers to questions will be posted periodically
to the ASE website:

<http://www.darpa.mil/ato/solicit/ASE/index.htm>