The Seventeenth Annual Interactive Audio Conference
PROJECT BAR-B-Q 2012

Group Report: Smart and Connected Microphones and Speakers

Participants: *A.K.A. "Doppler Chickens"*
Roderick Hogan, Texas Instruments
Yuval Weinreb, WAVES
Larry Przywara, Tensilica
Alex Westner, iZotope
David Roach, Optimal Sound
Facilitator: Doug Peeler, Dell
Jeremiha Douglas, Dolby Labs
Joel Susal, Dolby Labs
Moshe Sheier, CEVA
Mikko Suvanto, Akustica
Medford Dyer, Harman/MWM

Problem Statement

Consumer Devices are becoming smaller and thinner all the time. These factors, along with a trend toward reduced cost, introduce obstacles that make it increasingly difficult to render a good audio experience.

This difficulty applies to both speakers and microphones. Yet the possibility to marry these two components more tightly can provide opportunities to create a more coordinated audio system that can significantly improve audio experiences.

Key problems with speakers:

- Poor frequency response
- Lack of “low” frequencies
- Insufficient SPL
- Size
- Competing experience requirements (voice vs. music)
- Cost focus is impacting the audio experience

Key problems with microphones:

- Not quiet enough
- Distorts easily at high levels
- Unmatched (arrays)
- Size
- Microphone implementations don’t match usage models
- Mounting techniques not well understood by OEMs/ODMs
- Little opportunity for real-time playback / monitoring - you don’t know it’s crappy until later
- Acoustical and electrical noises
    - Chassis noises
    - Rotational noises: fan and hard drive noise
    - Keyboard clicks; finger thumps

Solutions

"Smart" Applications

What are “smart” applications?
Smart applications seek and use external input/data to improve the capture and/or rendering of audio and voice streams. Smart Applications include the following:

- *Wake on Word*: always on, low power voice trigger function to wake up the device upon a pre-configured keyword identification so that a higher level voice recognition engine can kick in
- *Aggregate Microphone*: an aggregate microphone can use the speaker for capture if the mic input is too loud
- *Thermal Management of Loudspeaker Transducer:* use the current sensor to monitor transducer temp. and reduce output power accordingly
- *Loudspeaker Excursion Maximization:* understanding the limits of the cone travel of a transducer based on the current sensor; used to protect the transducer
- *Energy Redistribution:* use the mic input to monitor ambient noise and reduce energy of noisy frequencies by not rendering frequencies which will not be heard anyway
- *Loudspeaker Linearization:* maintain loudspeaker linear output response
- *Cone of Silence:* creating an area of “silence” around the receiver to eliminate noise in the end-user’s ear
- *Smart Volume*: automatic volume based on ambient noise (uses mic input)
- *Smart EQ:* automatic equalization
- *Gesture Detection:* uses ultrasonic speaker waves and related microphone inputs to detect user hand gestures (to signal volume up/down, answer/hangup etc.)

What do we need to know about speakers that will factor into our applications?

- Frequency rolloff
- Power handling at different frequencies
- SPL, given digital signal level
- Real time speaker displacement
    - Can be used to calculate acoustical displacement and assist with echo cancellation
- Speaker degradation
- Component variation
- Acoustical leaks
- Voice coil temperature
    - or air temp, if voice coil temp n/a
- System and device capability constraints
- Configuration
    - # of channels
    - Geometry
    - Orientation
- Ambient sound levels & spectral content
- Rest of system telemetry
- Interchannel communication

- Where is the end user relative to the device
- Power budget
- Real-time amp output voltage or acoustic pressure
- When additional transducers are added to the system in real-time

What do we need to know about microphones that will factor into our applications?

- How many
- Configuration
    - Orientation
    - Locations
- Phase & amplitude
- Tolerance
- Power budget
- Dynamic range & Acoustic Overload
- Noise floor
- SPL, given digital signal level
- Long term degradation
    - Can be measured and inferred via processing
- Frequency response
- Pattern

Additional high-level concepts

“Feedback vs Feed Forward”
In the context of a speaker amp and transducer combination, feedback typically implies a low-latency feedback path of the current and voltage which can be monitored to determine the speaker excursion and the voice coil temperature. This data can then be tightly coupled with a local processing block to dynamically modify the output in real-time. In this same speaker amp and transducer combination, feed forward will use a pre-built model of the expected performance of that transducer across content, and - with less accuracy than a feedback system - will predict the speaker excursion and voice coil temperature, and dynamically modify the output based on this predicted data. This type of processing can be done in a higher level processing block and is less sensitive to latency.

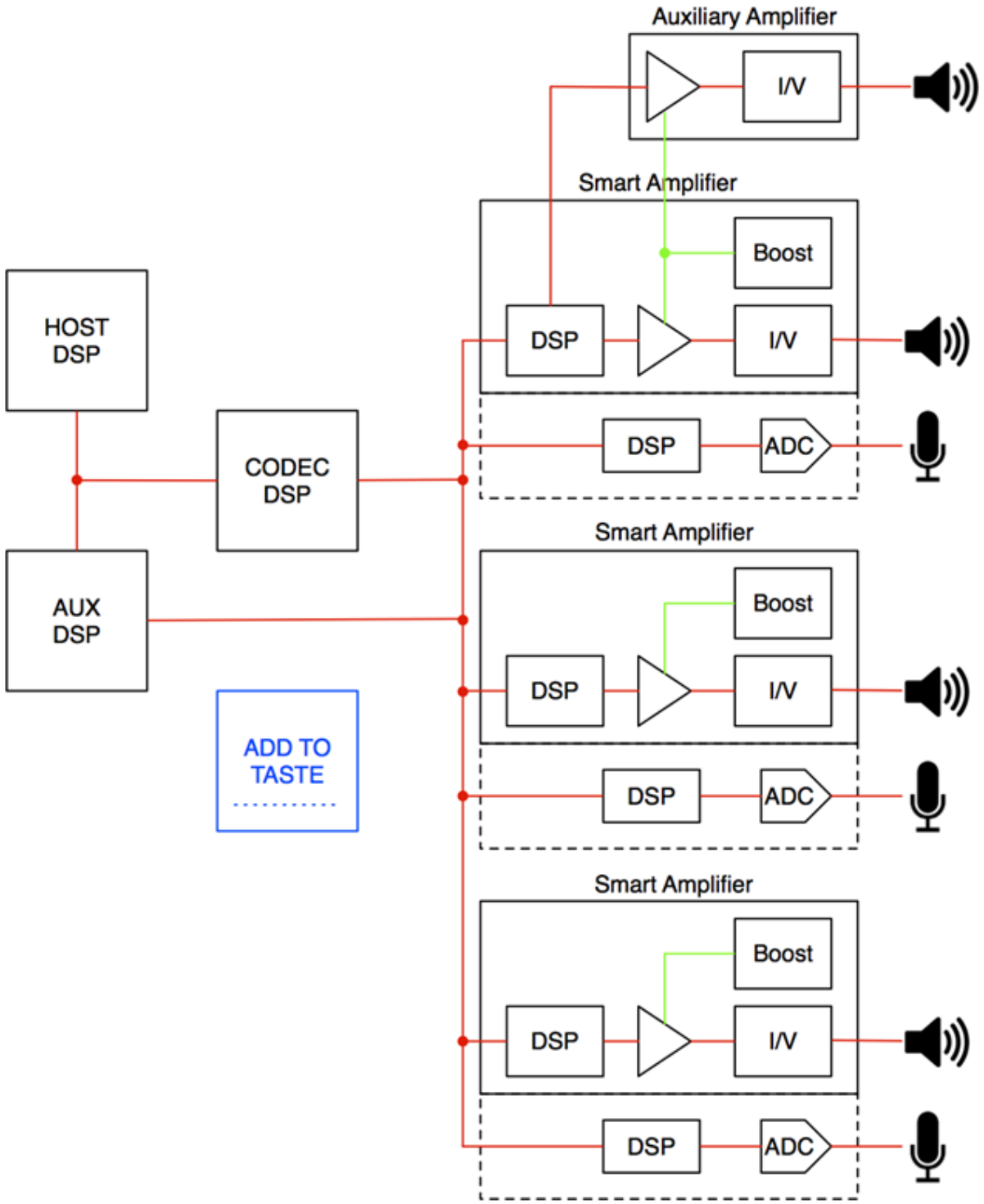
“Seven Levels Needed for Voice Identification”
In order for a system to recognize a human voice, there are several steps that must be followed:

1. Energy - Identify that energy is present (digital or analog).
2. VAD – Identify that the energy is human (digital or analog).
3. SAD – Identify that the human is using speech.
4. Diarization - The ability to immediately identify that a new person is speaking.
5. Speaker identification using a known text or phrase
    a. I know what the word is
    b. I know a small number of users
    c. Talker identity
6. Speaker identification using an unknown text or phrase
    a. Just talking, saying anything, no idea about language and words
    b. Talker identity
7. Verification - the computer has verified that the individual speaking is the individual with privileges on

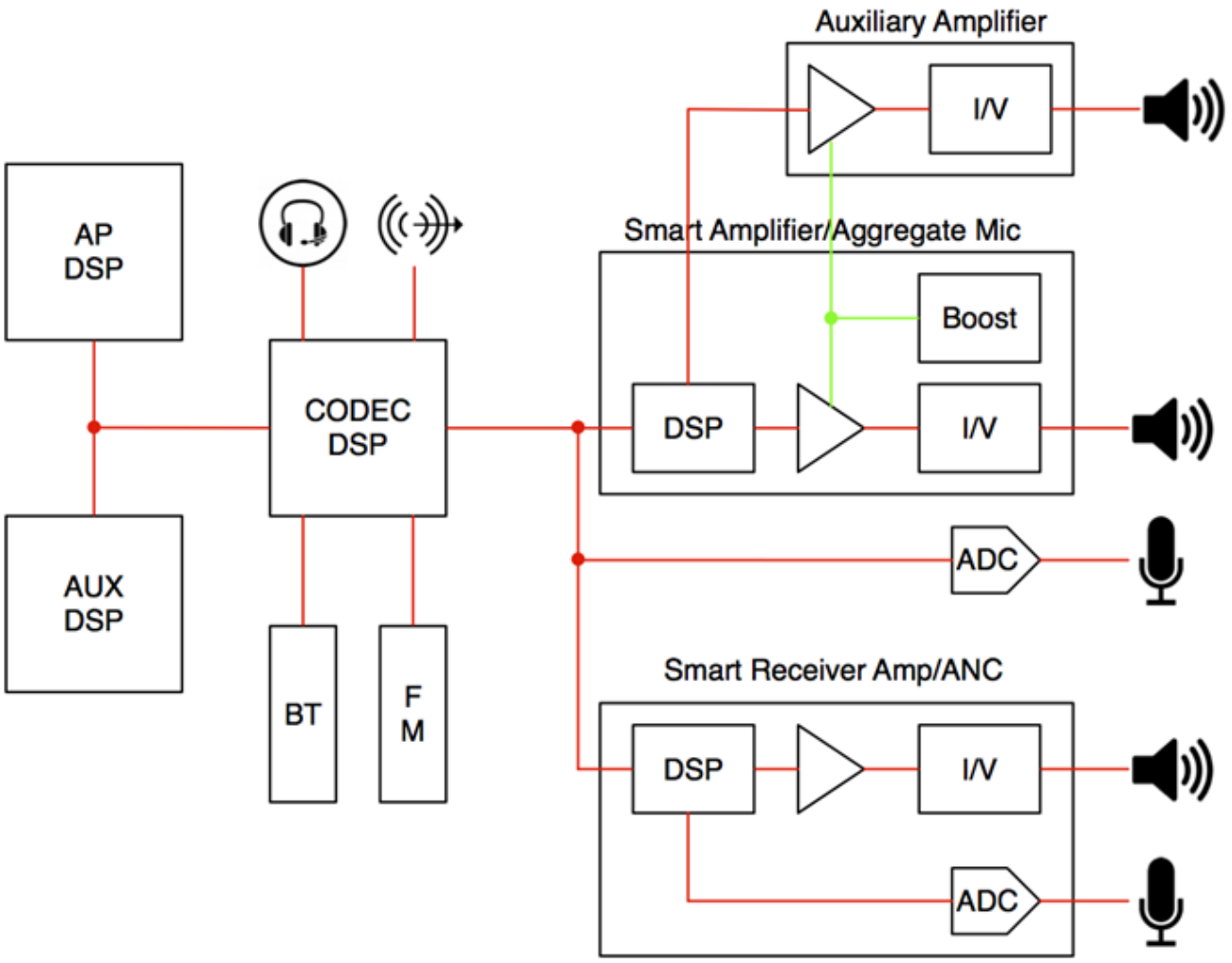
that device.

“Metadata Sharing”
It is important that metadata be able to travel with the audio data in both directions on any of these pipes.

2.1-Speaker, 3 Microphone Superset Block Diagram



SmartPhone Block Diagram



Smart and Dumb Applications using DSP with Microphones and Speakers

Applications	DSP Location	Capture	Render	Render w/ Capture	Current Sensor	Proximity Sensor	Ultrasonics	Power Sensitive	Latency Sensitive
"Smart" Applications									
Wake on word	mic	x				x		x	
Aggregate microphone	all	x							
Thermal management of loudspeaker transducer	spkr		x		x				
Loudpeaker excursion maximization	spkr			x	x				x
Energy redistribution	all			x					
Loudspeaker linearization	all			x	x				x
"Bose-style" noise cancellation (ANC)	mic/spkr			x		x			x
Cone of silence	mic/spkr			x	x	x	x		
Smart Volume	all	x	x	x	x	x			
Smart EQ	all	x	x	x	x				
Gesture detection	all	x					x		
AGC	all	x							
"Dumb" Applications									
Capture classification	C/A/H	x							
Media classification	C/A/H		x						
Snap, crackle and pop removal	C/A/H	x							
Key click suppression	C/A/H	x							
Clipping detection	C/A/H	x	x						
EQ	C/A/H	x	x						
Dynamic Range Compression	C/A/H	x	x						
Spectral Shaping	C/A/H		x						
Feedback elimination	C/A/H			x					
Voice Recognition	C/A/H	x							
Talker ID	C/A/H	x							
Talker Verification	C/A/H	x							
Voice vs. Music ID & Tuning	C/A/H	x	x						
Upmix / downmix / virtualization	C/A/H		x						
Encode/decode voice	C/A/H	x	x						
Encode/decode music	C/A/H	x	x						
Adaptive Noise Reduction	C/A/H		x						
Speech intelligibility	C/A/H	x							
AEC	C/A/H	x	x	x					
Beamforming	C/A/H	x	x						
Key	H = Host								
	A = Aux								
	C = Codec								

Industry Solution Next Steps

Cooperation required!

- **Doppler Chickens members will evangelize back in our home companies**
- We need to let transducer manufacturers know what they can do to improve physical characteristics
- In general, communication & engagement between transducer manufacturers, semiconductor, systems, algorithms, OEMs
- Software & chip guys can subscribe to Voice Coil Magazine and/or join an organization like ALMA to get their learn on about the challenges of speaker and transducer design and manufacturing

- Invite relevant folk to Project Bar-B-Q!
- Get branded algorithm guys more interested in this “closely coupled to transducers” space
- Can we plan a meeting around this ecosystem (@ CES?)
- Talk to MIPI Alliance
- Work with "Sounds like Chicken" and "Screaming Monkeys" groups
- Software people should talk to hardware people about desired applications

Prototyping of more tools in the toolbox