The Essentials of Automotive Hands-Free Systems

Phil Hetherington, Ph.D., Sr. Director, Acoustics
Andrew Mohan, P.Eng., Sr. Engineer, Acoustics
QNX Software Systems Limited
phetherington@qnx.com, amohan@qnx.com

Abstract

This paper attempts to facilitate the task of choosing a hands-free system by describing key features and characteristics that contribute to system performance. After describing these features and characteristics, it presents a checklist which OEMs and Tier 1 manufacturers considering an automotive hands-free system for their vehicles can use to help them make their choices.

Introduction

As mobile phones evolve to take advantage of emerging LTE, 4G networks, the performance demands on in-vehicle hands-free systems will only increase. Where yesterday almost any more or less comprehensible hands-free communication was acceptable, today users expect a level of clarity and intelligibility that was unknown just a few years ago.

To deliver this performance, hands-free systems have become very sophisticated and complex, which makes evaluating and comparing different systems difficult and time-consuming. Five factors contribute to the success of hands-free deployment: performance, the supplier, the tool set that comes with the system, the feasibility and cost of implementation, and documentation.

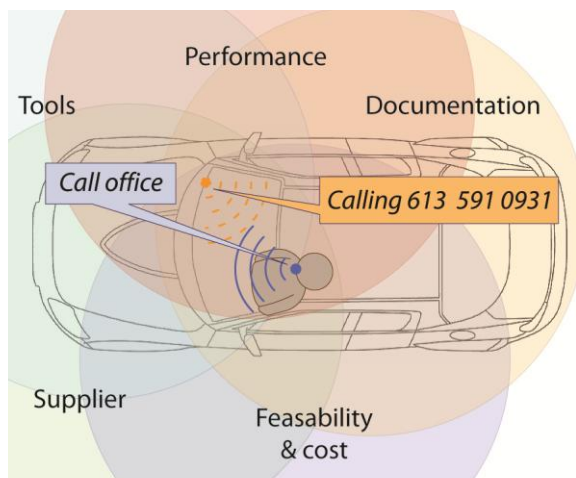


Figure 1. The five factors that contribute to the success of hands-free deployment. Ultimately, a hands-free system is judged by the quality of the useful sound it delivers to users.

Ultimately, however, a hands-free system is judged by the quality of the useful sound it delivers to users, so this document focuses on the technology used to deliver this sound.

Acoustic echo cancellation

Acoustic echo cancellation should handle double-talk and be stable, even under the harsh acoustic environment of the vehicle cabin and with the less than ideal transmission of mobile networks.

In an in-vehicle hands-free call the downlink speech signal with the far-end speaker's words is sent to the vehicle loudspeakers rather than to the mobile phone's speaker or earpiece. Since the sound is played over loudspeakers, it is picked up again by the near-end microphone (either on the mobile phone or in the vehicle) and echoed back to the far-end. This acoustic echo can seriously compromise the clarity and comprehensibility of a conversation.

Acoustic echo cancellation (AEC) prevents far-end sounds from echoing back to the far end. Removing these sounds so that they are not echoed is a complex and often computation-intensive procedure, as it must cancel only the unwanted sounds, and it must function effectively under many different and rapidly changing conditions[1].

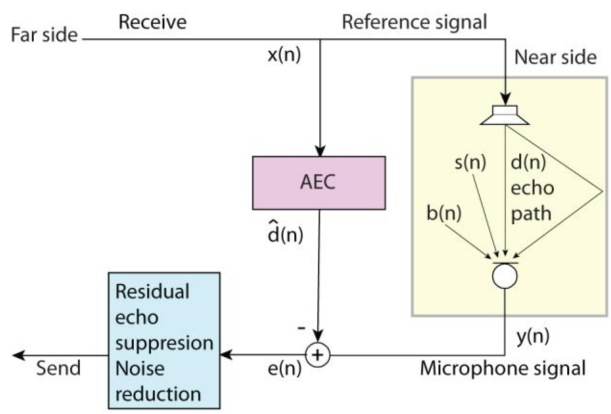


Figure 2.A high-level view of typical AEC system. Other designs are possible. Every AEC system uses its own algorithms and techniques to cancel echo without adversely affecting the useful part of the signal. Developing these systems is very specialized work and calls for a great deal of scientific expertise—as well as long stints in the lab proving and adjusting implementations.

In a vehicle, the AEC system must cancel the echo in conditions of high noise and high acoustic coupling, and during echo path variations when the vehicle occupants move, especially near the loudspeakers and microphones. Further, the AEC system must perform this cancellation without:

a) reducing the downlink level or suppressing the driver's speech when people at both the far-end and the near-end speak at the same time (double-talk)

b) leaking any echo through under any condition

That is, an in-vehicle AEC it must perform under the harsh acoustic conditions of the automobile cabin and the equally harsh transmission conditions of the mobile

[1] See Shreyas Paranjpe *et al.*, "Acoustic Echo Cancellation for Wideband Audio". <http://www.qnx.com/download/feature.html?programid=21750>

cellular network. It must perform with stability during quantized noise, network dropouts, and large gain variation across phones.

Noise reduction and speech reconstruction

Noise reduction must not impact the quality of speech, but improve the signal-to-noise ratio and attenuate the effects of mobile network gating.

Automobiles are subject to large and rapid fluctuations in noise levels from HVAC systems, windows, the engine, the tires and the road surface. Most engine and road noise is low frequency, but it cannot be removed by simply filtering out these frequencies, because this would also significantly thin the near-end person's voice.

Network gating

Mobile phone networks (CDMA and GSM) were not designed with the nature and levels of noise heard in automotive hands-free calls in mind. To make optimal use of available bandwidth, mobile networks "gate" transmission. In a CDMA2000 network, for instance, when the near-end person pauses or speaks very softly, to save bandwidth the network closes the gate and stops transmitting, re-opening the gate only when that person resumes speaking. Because conversations from a vehicle often carry a lot of background noise, the person on the far-end of the conversation hears an unpleasant oscillation between conversation with background noise and sudden silences.

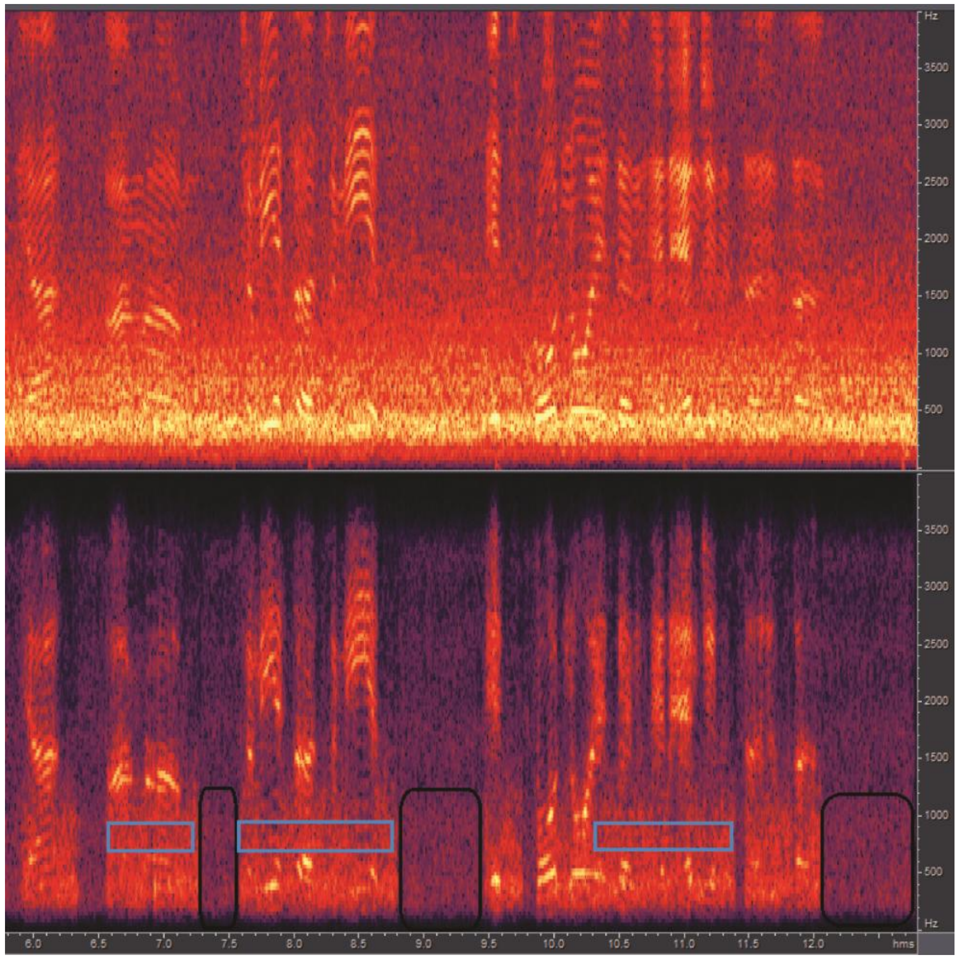


Figure 3. The top image shows the raw unprocessed microphone signal. The bottom image shows a post-CDMA2000 signal. Instances of noise gating between speech utterances are framed in black; missing speech harmonics are framed in blue..

Recent advances

Reducing the unpleasant effect of mobile network gating demands significant noise reduction and improvement of the signal-to-noise ratio (SNR), especially in the lower frequencies. Noise reduction that is effective in an automobile's acoustic environment requires more than 16 dB of noise attenuation. Unfortunately, before 2011 no solution was available that could provide more than about 12-14 dB of attenuation without degrading speech.

Fortunately, recent advances have made it possible to achieve the required levels of noise reduction without damaging the speech quality using the following techniques:

- Dynamic noise removal—increasing the level of noise removal where there is more noise (e.g., low frequencies or tones)
- Speech reconstruction—reconstructing speech that has been completely masked by noise

- Advanced algorithms—mitigating issues such as CDMA gating on downstream cell phone codecs

A good hands free system must do more than basic noise removal, and should not use algorithms that fake greater noise attenuation by gating the speech themselves, as this will only make the problem worse.

Multi-channel support

Multiple input channels must be available so that the vehicle manufacturer can offer different microphone arrays and configurations, as well as block some inputs during speech recognition activities.

In-vehicle hands-free systems are often required to support either only one talker at the near-end or many talkers, who may need to be blocked during activities that use voice recognition. These sorts of spatially selective tasks are often at odds with another important requirement of the microphone system: to reduce diffuse background noise.

To meet these conflicting requirements (variable number of participants, blocked microphones, and background noise reduction) the hands-free system can use two or more microphones with well considered inter-microphone spacing, and specialized algorithms for handling multi-channel input. A system with multi-channel support can:

- Reduce noise—well-placed microphones can pick up clearer speech and less background noise
- Include multiple near-end talkers—with separate microphones for different talkers, the conversation can include (and exclude) different people in the vehicle, as needed
- Improve voice recognition activities—blocked microphones exclude everyone except the person speaking the instructions

Microphone array independence

The same manufacturer may use a variety of microphone arrays and configurations. Even within the same project, different vehicle trims may have different arrays, and these arrays may even change during production. A good hands-free system is not dependent on a particular microphone array, and can be adapted to different arrays and configurations without changes to the software.

Algorithms

Many textbook algorithms that work well in an anechoic chamber do not work in a car due to the multiplicity of hard glass and plastic surfaces near the microphones. Further, many solutions fail robustness tests when faced with a sensitivity mismatch or loss of one of the microphones. To avoid such pitfalls, look for a system with a proven record of multi-channel support in automotive environments.

Automatic gain control

Gain control should accurately differentiate between voice and other sounds, such as wind buffeting, and it should use a soft look-ahead limiter to maintain a consistent voice level both when sending and receiving.

A hands-free system must maintain a consistent voice level for talking on the near end without increasing gain in response to wind buffeting, GSM buzz from the mobile phone, or any other non-speech sounds. If a person in the back seat speaks, the automatic gain controller (AGC) should amplify this person's voice automatically; the person on the far end of the conversation should not need to ask anyone to speak up. Similarly, if someone suddenly speaks more loudly, the AGC should adjust to this change smoothly.

The AGC must function equally well for the receiving (downlink) side of a conversation. Far side terminals, cell phones, networks, and the people speaking contribute to huge variations in voice level coming in. Variations can reach 30 dB very rapidly. The AGC should compensate for these variations, and maintain stable voice levels without sudden, annoying or otherwise noticeable increases or decreases in receive noise.

Look-ahead limiter

The AGC should compensate for quantization noise (or quantization error: the difference between the analog and digital signal values) and network dropouts. A soft look-ahead limiter in the AGC is an effective technique for handling sudden changes in loudness, and ensuring that speech is neither clipped not distorted.

Equalization

Equalization should be simple and handle both high- and low-noise conditions.

An in-vehicle hands-free system must accommodate significant differences in send and receive frequency response in order to accommodate differences in microphones, amplifiers, speakers, phones, far-side terminals and cabin acoustics. The system should, therefore, include a simple means of implementing an equalizer (EQ), such as a parametric equalizer, that allows specification of node centers, widths and gain values.

The best EQ for low noise conditions is often not the best choice for high noise conditions. Therefore, many systems use an EQ with both a low noise curve and a high noise curve; the EQ performs automated dynamic blending between curves, based on either noise level or signal-to-noise ratio (SNR).

Wind buffet suppression

An efficient algorithm for handling wind buffeting is a key component of any hands-free system.

Automobile interiors are filled with constantly changing winds and turbulence: HVAC systems, windows, sun roofs and so on. The acoustic effect of wind on a microphone varies with the force of the wind and the sensitivity of the microphone. Effected frequencies range from about 100 Hz to about 4 kHz:

- 100-200 Hz—minimal wind

- 400-800 Hz— (normal)
- 2-4 kHz—extreme wind, causing clipping

Figure 4 shows a raw microphone signal affected by wind buffeting, and the same signal with wind buffeting reduced. Note that the quality of the raw speech signal is not compromised.

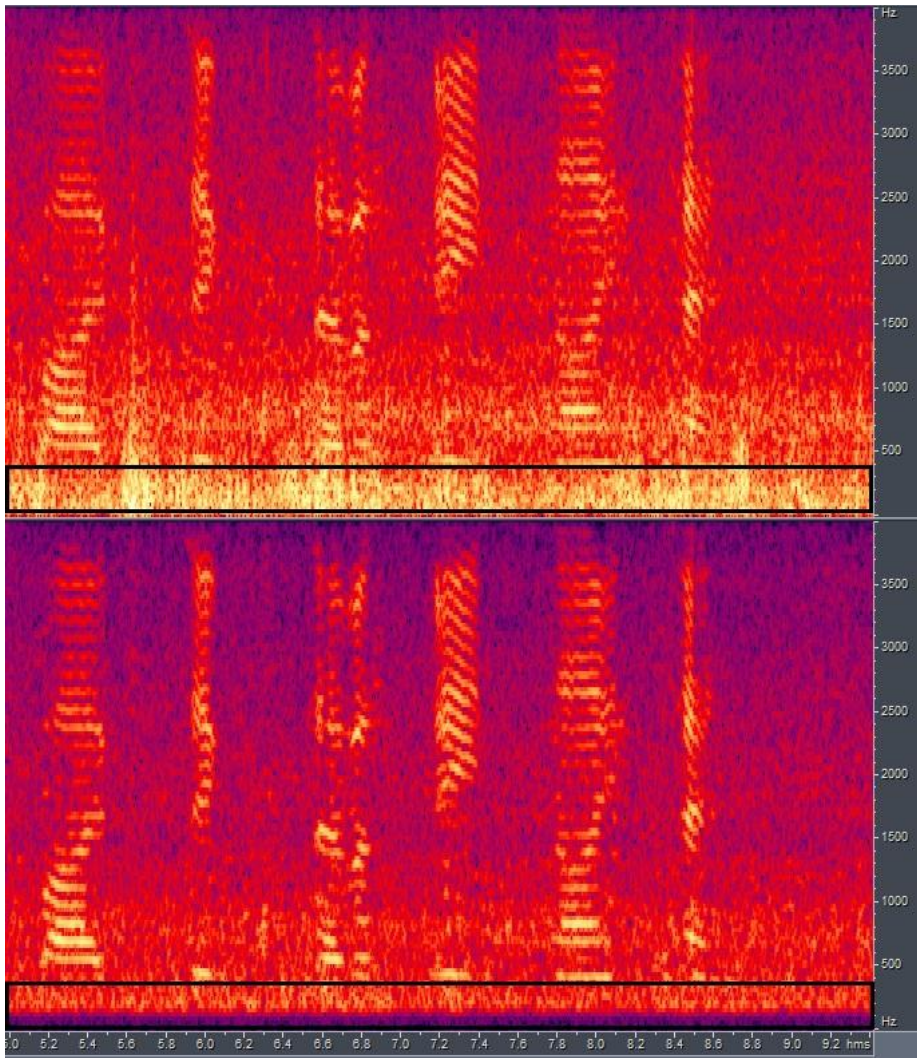


Figure 4. The top image shows a raw microphone signal affected by turbulent wind flow (buffeting). The bottom image shows a novel way of cleaning up the buffeting without compromising the quality of the original speech signal.

Unfortunately, directional microphones, which are the microphones of choice for in-vehicle systems, are more susceptible to the acoustic effects of wind buffeting than are other microphone designs. To counter these effects, a hands-free system needs wind buffet suppression capabilities. The algorithm it uses should not be limited to a passive high-pass filtering of the microphone signal, but should identify and selectively remove the acoustic effects of wind and turbulence.

Intelligibility enhancement

Intelligibility enhancement techniques must complement noise reduction.

In hands-free communication, lower frequency consonants (e.g. /p/, /t/ /k/, etc.) are often masked by the noise in the vehicle. Fortunately, because noise is predominantly low frequency (below 3 kHz), higher frequency consonants (e.g. /s/, /f/. etc.), which many people find difficult to distinguish even in the best acoustic environments, are often not masked, making them good candidates for effective intelligibility enhancement.

Noise removal does not directly improve the intelligibility of a call, because it only removes information. However, it can be used with intelligibility enhancement techniques. These techniques can complement noise removal by reconstructing masked low energy consonants and boosting high-energy consonants to make them easier to distinguish.

Noise dependent receive gain

Noise dependent receive gain should be smooth, with parametric control of change rates, and it should use algorithms that make minimal assumptions about the acoustic environment.

To help compensate for loudness masking, the hands-free system should automatically adjust the receive level supplied to the head unit, based on the noise in the vehicle, as estimated by the hands-free microphone.

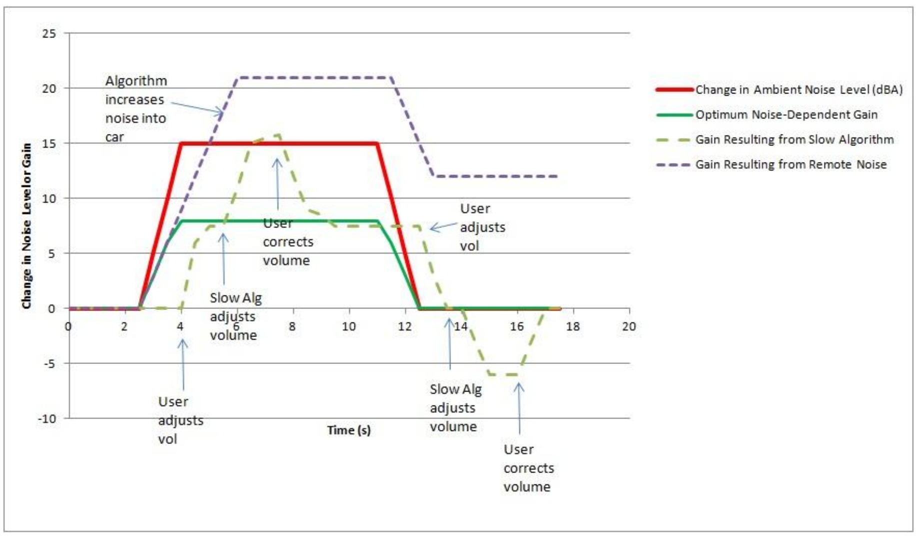


Figure 5.Receive gain that is too slow can precipitate "gain-chasing", a situation where the driver increases the volume before the gain has taken effect, requiring more gain, and so on.

Gain adjustments use ambient noise compensation or dynamic level control algorithms. These algorithms should provide smooth responses to changes in noise levels:

- If the adjustment is too slow, the driver may reach for the volume, which can provoke "gain chasing". See figure below.

- If the adjustment is too quick, the sudden increase or decrease may seem unnatural, or even startle the driver.

Look for a solution that supports parametric control of the gain change rates and has a proven record in production. Look also for algorithms that make minimal assumptions. In fact, the best algorithms make only a single assumption: that the noise measured at the microphone is the same as at the ear.

Bandwidth extension

Improvements to speech through bandwidth extension is difficult to accomplish, but is important for today's dominant narrowband communications.

Improving the intelligibility of far-end speech is especially valuable in a noisy automobile. Bandwidth extension (BWE) can improve the quality of far-end speech heard by the driver. Bandwidth extension uses filtering and wave shaping to reconstruct high and low frequency portions of the incoming signal that were dropped by the narrowband mobile network.

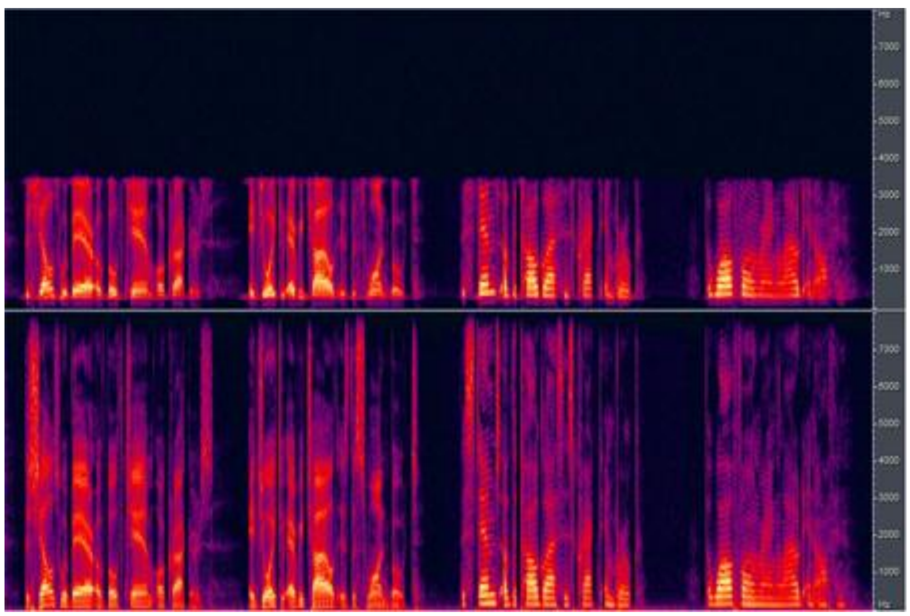


Figure 6. Bandwidth extension can improve intelligibility of incoming speech, but is difficult to implement effectively. In this image the plot on the left is the original signal without BWE (narrowband), while the plot on the right is the same signal with BWE (wideband). Note that with BWE both the low and high frequencies are extended.

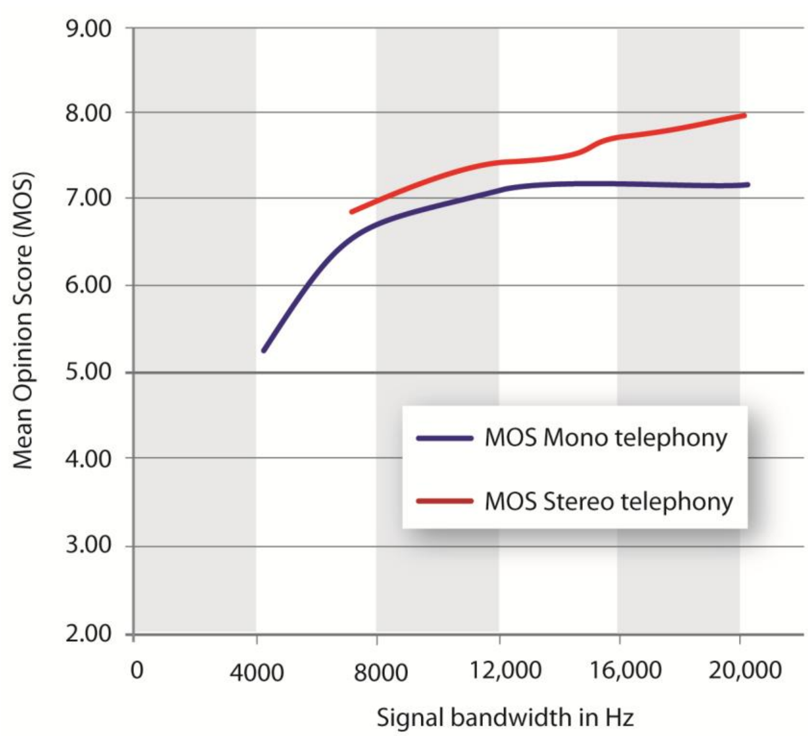
Bandwidth extension is difficult to implement, because there is often little speech information to work with and it requires a judicious balance between too little reconstruction and too much. Too little reconstruction brings little or no benefit, while too much can produce unwelcome exaggerated bass and treble speech components. The best measure of a bandwidth extension solution, then, is its record of acceptance by customers and users in the field.

Wideband support

Mobile networks are beginning to adopt wideband communications, which will soon render obsolete any hands-free system that does not support this technology.

Cellular networks in Europe and North America are beginning to support bandwidths greater than 3.5 kHz. With the advent and growing adoption of 4G and LTE networks and devices, mobile VoIP calls are just around the corner. Wideband, with frequencies of up to about 7 kHz, or super-wideband, with frequencies of up to about 15 kHz, are coming to automotive. In fact, they make most sense in automotive, as they improve intelligibility, which can help reduce driver cognitive load during a phone call.[2]

Given that an automotive hands-free system should remain current from start of production (SOP) through the entire life of the vehicle, that is to say, for 10 or more years, any system that cannot run at a sample rate of 16 kHz is already dated. Hence, look for evidence that the supplier is working to support a 32 kHz sample rate and already has experience with VoIP codecs.



[2] See Scott Pennock, *et al.*, “Wideband Speech Communications for Automotive” <http://www.qnx.com/download/feature.html?programid=21293> and Scott Pennock and Andy Gryc, “Situation Awareness: A Holistic Approach to the Driver Distraction Problem” <http://www.qnx.com/download/feature.html?programid=22439 >

Figure 7. We see a significant improvement in Mean Opinion Score (MOS) by moving from narrowband (4kHz bandwidth) to wideband (8kHz bandwidth)[3].

Note that while the frequency range of calls is likely to double or even quadruple in the next few years, it is unlikely that the available computing resources will easily accommodate a similar increase in processing requirements. Hence, the best solutions will not scale linearly with sample rate, but will offer super-wideband processing with little more resources than they currently use for narrowband.

[3] From Kari Järvinen *et al.*, "Media coding for the next generation mobile system LTE", *Computer Communications* 33 (2010): 1916-1927.<www.elsevier.com/ locate/comcom>

Appendix: Decision checklists

The checklist below follows the same general organization as the whitepaper, plus other items that will affect the success of a project. It may be useful when evaluating a hands-free system and its vendor.

Performance

- ☐ **Acoustic echo cancellation**
  - ☐ Full duplex Type 1 Send
  - ☐ Full duplex Type 1 Receive
  - ☐ Full Duplex in high noise
  - ☐ No echo during movement of driver in car, especially near loud-speakers or microphone
  - ☐ Robust handling of high acoustic coupling
- ☐ **Noise Reduction**
  - ☐ Up to 24 dB of noise attenuation without speech degradation
  - ☐ Dynamic noise reduction; greater attenuation where noise is higher
  - ☐ Speech reconstruction for smoother, more natural speech in high road and wind noise
- ☐ **Multichannel Support**
  - ☐ Self calibrating to handle microphones with unknown spacing
  - ☐ Works with parallel or splayed microphone arrays
  - ☐ Supports running production change from single to dual microphones without retuning
  - ☐ Robust to phase inverted microphone inputs
  - ☐ Passenger support for dual microphones
  - ☐ Seamless mixing of driver and passenger speech
  - ☐ Robust response catastrophic failure of a microphone (e.g. crash)
  - ☐ Rejects or suppresses transient noises coming from a direction other than the driver.
- ☐ **Multichannel Support (con’td)**
  - ☐ Robust response catastrophic failure of a microphone (e.g. crash)
  - ☐ Rejects or suppresses transient noises coming from a direction other than the driver.
- ☐ **Automatic gain control**
  - ☐ Uplink direction
    - ☐ Adapts only in response to voice, not to wind buffets
    - ☐ Does not increase gain in the presence of GSM buzz (edge networks)
    - ☐ Combined with look-ahead (soft) limiter
  - ☐ Downlink direction
    - ☐ Robust to quantization noise
    - ☐ Robust to network dropouts
    - ☐ Combined with look-ahead (soft) limiter
- ☐ **Equalization control**
  - ☐ Uplink direction
  - ☐ Downlink direction
  - ☐ Parametric control
  - ☐ Dynamically modifies uplink equalization based on SNR at the microphone
- ☐ **Wind buffeting suppression for single directional microphone**
- ☐ **Intelligibility enhancement**
- ☐ **Noise-dependent receive gain**
- ☐ **Bandwidth extension**
- ☐ **Wideband support**

Documentation

- ☐ **Fully documented API**
  - ☐ Sufficient overview of each library feature and switch
  - ☐ Recommended use cases
  - ☐ Explanation of default parameter values and ranges
  - ☐ Recommended tuning parameter values
  - ☐ Example code for calling API
- ☐ **Documentation of the Bluetooth in-vehicle demonstration system**
- ☐ **Fully documented operation of the tuning tool**
- ☐ **Recommendation and best practices documents**

Tools

- ☐ **Tuning tool**
  - ☐ Available for customers to use
  - ☐ Fully-documented user guide
  - ☐ Easy-to-use and intuitive GUI
  - ☐ Real-time visual control interface for parameters, especially setting and observing the send and receive EQ curves
  - ☐ Stream and inject all channels simultaneously
  - ☐ Inject wav files into any or all audio tap points from host to target for repeatable issue diagnosis
  - ☐ Cue markers in wav files to mark real-time switch and parameter changes
  - ☐ Supports a serial-port interface for embedded targets with no Ethernet
  - ☐ Capture parameter state of the library at tuning time, save to a binary configuration file, and push to target or other targets for repeatability
  - ☐ Remote configuration control
  - ☐ Facilitates diagnostics, coherence, distortion tests
- ☐ **Average vehicle tuning time < 2 hours**

Feasibility

- ☐ **Reduced parameter set**
  - ☐ Only include parameters that can be readily described and understood
- ☐ **Support multiple H/W platforms:**
  - ☐ Renesas SH 2/3/4
  - ☐ Freescale PPC 5200, i.MX 31/35
  - ☐ TI Jacinto, C64x
  - ☐ ARM 9/11, Cortex A8, A9
  - ☐ Intel Atom/x86
- ☐ **Support multiple code formats**
  - ☐ Fixed-point
  - ☐ Floating-point
  - ☐ DSP
- ☐ **Support multiple OSs**
  - ☐ QNX
  - ☐ μ-ITron
  - ☐ WinCE
  - ☐ Linux
  - ☐ Android

Vendor

- ☐ **Generally available releases**
- ☐ **Mature library design**
  - ☐ Both fixed-point and floating-point formats
    - ☐ Identical APIs for both solutions
    - ☐ Near identical performance for both
  - ☐ C (not C++) interface
  - ☐ Simple API, designed for both forward and backward compatibility
- ☐ **Company**
  - ☐ Extensive worldwide automotive production experience
  - ☐ Dedication to OEM and Tier 1 space
  - ☐ Active roadmap with proven ability to support narrowband to fullband speech systems for both mono and stereo modes
  - ☐ Innovative group, with on-going research and development
  - ☐ ITU-T participation & leadership
- ☐ **Company (con'td)**
  - ☐ Mature coding standards and strong QA
  - ☐ ISO 9001:2008 certified management system
  - ☐ Worldwide technical support structure
  - ☐ Diverse team of researchers, developers, QA, field application engineers
- ☐ **Measurement and testing**
  - ☐ Objective measurement equipment and facility for ITU-T P.1100/P.1110 & VDA 1.6 tests
  - ☐ Objective measurement experience based on ITU-T P.1100/P.1110 & VDA 1.6
  - ☐ Subjective Test Experience
  - ☐ Demonstrate a working BT hands-free system in a vehicle
  - ☐ Demonstrate the tuning tool

About QNX Software Systems

QNX Software Systems Limited, a subsidiary of BlackBerry, is a leading vendor of operating systems, development tools, and professional services for connected embedded systems. Global leaders such as Audi, Cisco, General Electric, Lockheed Martin, and Siemens depend on QNX technology for vehicle infotainment units, network routers, medical devices, industrial automation systems, security and defense systems, and other mission- or life-critical applications. Founded in 1980, QNX Software Systems Limited is headquartered in Ottawa, Canada; its products are distributed in more than 100 countries worldwide. Visit www.qnx.com and facebook.com/QNXSoftwareSystems, and follow @QNX_News on Twitter. For more information on the company's automotive work, visit qnxauto.blogspot.com and follow @QNX_Auto.

www.qnx.com