TRENDS IN CELL PHONE VOICE PROCESSING

Chris Forrester
Research in Motion, 295 Phillip St., Waterloo Ontario, Canada, N2L 3W8 cforrester@rim.com

1. INTRODUCTION

Cell phone voice signal processing technology has come a long way since the first digital cellular phone call in 1991 on a second generation (2G) wireless network. For those early digital handsets, DSP processors lacked sufficient horse power to implement much voice processing beyond the speech codec itself. Since then, the speed of DSP's has increased by more than an order of magnitude and modern cell phones have added multiple processor cores to support multi-media in addition to voice. So given this increase in capability, what trends are emerging in cell phone audio design and what impact is does this have on voice processing?

2. DISCUSSION

In 1989, Texas Instruments introduced the highest performing fixed point DSP in the industry, operating at 28 MIPS [1]. This DSP became very popular for use in cell phones. The main priority of the DSP is always to implement the radio signal processing needed to support mobile communications. By necessity, this means that voice signal processing is allotted only a small fraction of the MIPS budget. So for early 2G cell phones, there was very little time budgeted for voice processing.

This lack of available processing power made life difficult for the audio pioneers of digital cellular. The acoustic design of early 2G phones had to meet GSM voice-band phone specifications without the benefit of DSP voice equalization. This forced the use of analog circuitry to provide any filtering needed to meet voice-band specifications. As well, there was little DSP processing power available to perform echo or noise control functions that are common today. As a result, it was not uncommon to hear echo of your voice when talking to a person on a digital cell phone. Received speech volume levels needed to be kept low to control echo which led to 'can't hear' complaints by cell phone users. Without background noise reduction, users had to make calls in moderate background noise conditions or suffer complaints from the called party. But of course it did not take long for this to change.

By the mid to late 1990s, DSP processing power had increased sufficiently to provide more enhanced voice signal processing, targeted at addressing the early mobile voice quality issues. To address the echo problems, it became the standard to provide a basic time-domain echo canceller based on Normalized Least-Mean-Squares (NMLS) with some form of basic gating or gain switching function to control residual echo [2]. As much as 30dB of echo return loss enhancement is possible under typical conditions. The standard for noise cancellation became some form of sub-band or full-band gain reduction based on SNR (speech-signal-to-noise) calculations [3] - as much as 15-20dB of stationary noise reduction is achievable with this technology without harming speech quality too much. Other more basic features such as transducer equalization and dynamic range signal processing to limit high-level signals and boost low-level signals became common as well [4].

Of course, DSP processing speeds continue to increase and current platforms are more than 50 times more powerful than the early 2G platforms. The increase in processing power has enabled DSP audio engineers to use more sophisticated voice processing to provide features and solve problems that were only dreamed of as recent as five year ago.

Since the release of the first wireless phone there has been a continued evolution towards smaller, thinner and lighter phones. This trend to shrink the cell phone has had a huge impact on the acoustic design and has driven a need for ever greater voice processing features and performance.

As devices shrink, microphones move farther away from the mouth which reduces the signal-to-noise (SNR) at the microphone. The SNR may drop by as much as 10dB or more depending on the design of the phone. So the background noise reduction must be improved by this amount to maintain performance. At the same time, cellular service providers are demanding improvements in non-stationary noise reduction.

A trend that has emerged recently is to provide multiple microphones to address the need for improved background noise reduction. There has been a lot of interest lately in the area of blind-source separation (BSS) and beam forming signal processing techniques for application in cell phones [5]. With these techniques it is possible to provide improved noise reduction, especially to address non-stationary noise where one microphone is insufficient. It is expected that future devices will add more microphones with more complex signal processing to achieve even better voice quality performance in non-stationary noise environments.

With multiple microphones it is possible to achieve significant noise reduction to the point that the far end user finds it unnaturally quite. Additionally, it is not uncommon for noise reduction algorithms to severely change the quality of the noise, so that it becomes very unnatural to the far end listener. The ideal is to achieve a reduction in the transmission level of noise to some optimal level, but leave the character of the noise fully intact for a more natural experience. So some attention now is being given to these problems and it is expected that future noise reduction systems will provide good suppression without harming the quality of the noise.

The first cellular speakerphone devices used loudspeakers as large as 34mm diameter. To save space in modern cell phones, it is common to use speakers that are rectangular with dimensions on the order of 14x20mm – and the trend is towards even smaller speakers. Due to physical limitations, the impact on audio performance is a reduction in maximum acoustic output and an increase in distortion. As well, in some speakerphone designs, the loudspeaker is placed literately next to the microphone due to space limitations, which leads to a very high echo signal. So signal processing engineers are being challenged to provide improved dynamic range compression to increase output and minimize distortion, and better echo control algorithms to handle the increased distortion at high volume levels.

A modern approach for providing residual echo suppression is to use sub-band analysis or spectral matching techniques to suppress the bands only with high echo levels – this can provide a more virtual full-duplex experience [6]. As well, researchers continue to work on non-linear filtering to address echo components from distortion added by the loudspeaker and plastics vibrations. A common approach is to use a Volterra filter to model the loudspeaker distortion [7]. Often it is the third harmonic of the input signal which is the most significant distortion component which leads many to use third order filters. It is expected that these techniques will enhance the echo return loss at high volume levels by 6dB or more.

One of the biggest cellular voice quality enhancements expected to take place in the near future is the rollout of wideband telephony. Wideband doubles the voice bandwidth from about 3.5 kHz to 7 kHz and promises to provide a more natural voice experience with increased user satisfaction. Existing narrowband voice processing will need to be upgraded to support the extended bandwidth. So this will require modifications to echo cancellation, noise suppression and all other blocks. Generally it will require as much as twice the processing power to support wideband, but this will not be a problem for modern DSP's.

While wideband voice will provide a significant leap forward for voice quality, the Third Generation Partnership Project (3GPP) Technical Specification Groups are already working on enhanced mobile audio specifications for future networks. This includes such features as 'superwideband' which doubles wideband to 14 kHz, stereo telephony and further extensions to multi-channel audio such as 5.1. These capabilities will allow other audio program material from presentations or teleconferences to be more faithfully transmitted with voice.

Finally, a recent area of research focus is non-intrusive, single-ended voice quality monitoring and assessment. These techniques analyze the incoming speech using various algorithms which attempt to relate the quality of the speech to some subjective equivalent. This capability may be used to dynamically adapt the phone or network performance to improve voice quality, or to simply report the voice quality experience to the cellular operator for diagnostic purposes. Different approaches exist in the literature based on the ITU G.114 E-Model [8] and P.563 [9] methods. More work is needed to find simple but reliable solutions which are suitable for mobile devices which have limited resources if this capability is to be implemented however.

3. CONCLUSIONS

With the advancements in digital signal processing hardware, it is clear that voice processing engineers have more opportunity than ever to have a positive impact on voice quality. The ever shrinking cell phone is forcing voice processing engineers to use more sophisticated techniques to deliver more loudness yet provide better echo, noise and dynamic range control. However, the rollout of wideband and the continued evolution in voice processing will no doubt mean that one day soon wireless voice quality will exceed landline performance. Voice signal processing engineers will play a very important role in achieving this milestone in wireless voice.

REFERENCES

[1] Texas Instruments, www.ti.com.
[2] Hänsler, E. and Schmidt, G. Acoustic Echo and Noise Control: A Practical Approach. John Wiley & Sons, 2004.
[3] Loizou, P., Speech enhancement: Theory and Practice, CRC Press, Boca Raton, FL, 2007.
[4] Zölzer, U., Digital Audio Signal Processing, Wiley & Sons, 1997
[5] Makino, S. et al, Blind Speech Separation. Springer, 2007.
[6] Hoshuyama, O., Nonlinear Echo Suppression Technology Enabling Quality Handsfree Talk for Compact Equipment. NEC Journal VOl. 2 No.2/2007.
[7] Guerin et al, Nonlinear Acoustic Echo Cancellation Based on Volterra Filters. IEEE Transactions on Speech and Audio Processing, Vol 11, No. 6, 2003.
[8] ITU-T, 2005, G.107, 2005, E-Model, a computational model for use in transmission planning.
[9] ITU-T, 2004, P.563, 2004, Single-ended method for objective speech quality assessment in narrow-band telephony applications.