Introduction

In survey after survey potential and actual users of wireless communications indicated that voice quality topped their reasons for selecting a specific service provider. While providers have been well aware of this key component powering their offering, they have not always been certain as to the specific methodology, resolution elements, equipment type, architecture, trade-offs, and rate of return on their particular investment that elevate the perceived voice-quality performance in their network.

It is only natural that voice quality in wireless networks has become a key differentiator among the competing service vendors. Network operators, network infrastructure planners, sales representatives of equipment vendors, their technical and sales support staff, and students of telecommunications seek information and knowledge continually that may help them understand the components of high-fidelity communicated sound.

Throughout the 1990s applications involving voice-quality enhancements, and specifically echo cancelation, have induced fresh inventions, new technology, and startling innovations in the area of enhanced voice performance. The initial echo canceler (EC) product implementations existed for about a decade before a diverse array of voice-quality enhancement realizations emerged to meet the evolving needs of digital wireless communications applications.

Early EC implementations were limited to very long distance (e.g., international) circuit-switched voice and fax applications where echo was perceived (in voice conversations) due to delays associated with signal propagation. The EC application soon expanded beyond strictly very-long-distance applications as further signal processing and dynamic routing along the communications path added delay to end-to-end voice transport. Consequently, EC equipment became a necessity for all long-distance calls (rather than just very-long-distance).

In the late 1980s, AT&T promoted a voice-transmission quality plan called the "zero-mile policy."[1] AT&T installed EC equipment next to each one of their toll switches. As a result, every trunk in the AT&T backbone network was equipped with EC coverage regardless of distance. The "zero-mile policy" made sense because of the dynamic routing architecture that AT&T put into practice, a treatment that essentially removed the correlation between geographic distance and physical wire span. Furthermore, innovations such

[1] AT&T introduced "dynamic routing" in the late 1980s. The procedure called for the use of lightly utilized links regardless of distance when routing calls along busy corridors. For example, morning calls between New York and Boston were routed via the West Coast, where traffic was still on its third shift. Late-night calls along the West Coast were routed through the East, taking advantage of the Late Night Show.

as call forwarding and call mobility have rendered geographic distance a weak predictor of signal delay. Consequently, there was a wider deployment of ECs by service providers, who have effectively duplicated the AT&T's experience.

Throughout the 1980s and until the mid nineties, AT&T with its leading and prominent Bell Laboratories, a body packed with Nobel-Prize scientists, continued to be the most innovative leader in the voice-quality arena. In 1991, AT&T revolutionized the notion of what voice-quality enhancement was about when it introduced TrueVoice. This new and innovative application was first to incorporate into their echo-canceler equipment a graphic-equalizer-type manipulation of speech levels by varying the extent of signal amplification across the various frequency bands of the speech impulse. Although the new technology improved voice quality, it did not interact well with certain modes of voice-band data. The new technology produced operational headaches, a fact that led to a slow withdrawal and an eventual suspension of TrueVoice in the mid to late nineties.

In the mid nineties AT&T introduced a subscription-based service – TrueVoice2 – a noise-reduction feature designed to enhance communications over a small cluster of international circuits. TrueVoice2 represented a first, a pioneering effort, intended to enhance voice quality by way of reducing circuit noise.

In the 1990s, mobile telephony introduced a new variety of impairments and challenges affecting the quality of voice communications. The impairments comprised acoustic echo, noisy environments, unstable and unequal signal levels, voice-signal clipping, and issues related to echo cancelation of vocoder compressed voice. The technological focus moved to developing remedies that would handle these new parameters properly. Furthermore, the challenges brought about by the wireless era imposed more stringent performance requirements on electrical-echo cancelation.

The 1990s experienced a vast growth in digital wireless communications. During the first half of the decade, voice-quality considerations other than echo cancelation were confined to low-bit-rate codec performance while hybrid-echo cancelation was delegated to a select group of stand-alone systems providing 8 E1/T1 per shelf with four shelves per bay (or 32 E1/T1) at best. During the second half of the decade, it was gradually acknowledged that mobile handsets failed to follow standards, and acoustic echo was not taken care of at the source. Echo-canceler equipment vendors such as Lucent, Tellabs, and Ditech Communications later on seized the opportunity for added revenues. They started incorporating acoustic-echo control while enhancing the offering with noise reduction and level optimization with their stand-alone echo-canceler systems. The market reacted favorably and voice-quality systems (VQS) embarked on an expandable path for replacing plain echo cancelers.

The commencement of the new millennium witnessed a continuing, yet substantial, advancement in microprocessor technology and in signal-processing software and algorithmic design. These developments spawned a fresh trend in the implementation and delivery of VQS. They allowed for high-density, lower cost per line, innovative products, and systems that could easily be integrated inside larger systems like a mobile switching center or a base-station controller. More and more VQS began selling as switch or base-station features and the stand-alone-system market was reduced to a secondary segment. Ericsson came up with their own integrated echo-canceler version. Nokia purchased echo-canceler technology from Tellabs, only to integrate it in its MSC,[2] Siemens created their version of application-specific integrated circuit (ASIC) echo canceler, and Alcatel acquired echo-canceler software technology from Ditech Communications. Lucent created a voice-quality (VQ) software version and integrated it with their codec software on their mobile-switching center, and Nortel purchased echo-canceler modules from Tellabs, which were inserted and integrated in their switch.

Stand-alone systems continued to hold a quality edge over their integrated cousins, which focused on hybrid-echo cancelation exclusive of acoustic-echo control, noise reduction, and level optimization. These voice-quality applications and a higher-performance echo cancelation were still part of a classy club of customers (or others whose mobile-switching center did not contain any voice-quality features) who preferred the higher voice quality and the operations inertia that history had bestowed upon them. Verizon Wireless, Nextel, T-Mobile USA, and T-Mobile Germany (as of 2004), Japan NTT DoCoMo and KDDI, most Korean and Chinese service providers, as well as numerous others continued to insist on stand-alone voice-quality systems rather than switch integrated voice-quality features.

Voice-quality systems and, more specifically, echo cancelation, have also become a crucial component of VoIP and ATM voice platforms. Still, the implementation is persistently integrated within the packet gateway, and stand-alone voice-quality systems have not been able to take a significant hold in this market segment.

This book regards a voice-quality system as a functional system. Most of the analysis and the descriptions are independent of whether the system is implemented inside a codec, a switch, a base-station controller, or as a stand-alone. The generic nature of the applications may be delivered via any of these, and most of the analysis except for a few sections, which declare themselves as particular to specific implementation, is not implementation specific.

Plan of the book

The book is divided into six major parts.

Part I – *Voice-quality foundations* – opens the discussion in Chapter 1 with an overview of voice-coding architectures in digital wireless networks. It provides an overview of the GSM, TDMA, and CDMA codecs from the early 1990s through the first half of the next decade. It provides a high level analysis of the architecture principles that make low bit-rate codecs effective in dictating the transformation of natural speech to an electronic signal and back into speech. And it articulates in plain language how this transformation alone produces changes in the quality of transmitted voice.

Chapter 1 reviews the major voice codecs, their history, and their relative perceived quality. Voice-coding architectures are the building blocks of transmitted voice. They are the core that shapes the characteristics and quality of transmitted speech. Nevertheless, they are treated in the book only as background to the main subject, which deals with

[2] Nokia added a home-made voice-quality suite on their base-station controller as an optional module later on.

impairments due to transmission architecture and environment, and corresponding remedies that immunize and repair any potential or actual spoil. Since the effectiveness of the various remedies depends on that underlying coding, it is essential that these designs be understood so that remedies can be fine tuned and customized to suit the particular characteristics of the underlying voice architecture.

Quantitative assessment of the perceived voice quality as it relates to a particular codec is postponed to the next chapter. Instead, the presentation conveys a sense of relative performance standing, and how voice quality has been improving over time.

Chapter 2 – *Quantitative assessment of voice quality* – kicks off the presentation with an overview of the standard metrics and methodologies followed by a description of specialized tools employed for obtaining subjective voice quality scores through genuine opinion surveys and via computer modeling emulating people's perceptive evaluation of speech quality. It then relates voice-quality scores obtained from surveys or computer evaluations to the perception of worth. It elaborates on the relationships between opinion scores and the potential return on investment in voice-quality technology. It examines the results of voice-quality studies with reference to the three popular GSM codecs – full rate (FR), enhanced full rate (EFR) and half rate (HR). The presentation includes a discussion of the effect of noise and transmission errors on the relative performance of these codecs.

Part II – *Applications* – opens the discussion in Chapter 3 with an overview of echo in telecommunications networks, its root causes, and its parameters. It follows the presentation with the methods used for controlling electrical echo, including network loss, echo suppression, linear convolution, non-linear processing, and comfort noise injection. The chapter covers the application of echo cancelation in wireless communications. And in view of the fact that today's wireless networks include long-distance circuit-switched VoIP and VoATM infrastructures (specifically as part of third-generation architectures), the chapter covers echo cancelation in long distance and voice-over-packet applications as well.

Chapter 4 – *Acoustic echo and its control* – examines the sources and the reasons for the existence of acoustic echo in wireless networks. It explains how acoustic echo is different from hybrid or electrical echo, and how it can be diagnosed away from its hybrid relative. The chapter follows the description of the impairment by examining the present methods for properly controlling acoustic echo in wireless communications. It also gives details of how background noise makes it more difficult to control acoustic echo properly. It describes those particular impairments that may be set off by some acoustic-echo control algorithms, specifically those built into mobile handsets, and it describes how they can be remedied by proper treatment brought about by means of voice-quality systems (VQS) in the network.

Both electrical echo cancelation and acoustic-echo control require a comfort noise injection feature. Discussion of acoustic-echo control must include methods and algorithms designed to generate suitable comfort noise. Although comfort noise injection is not a stand-alone application resembling the treatment of electrical or acoustic echo, it is, nonetheless, an essential component supporting these applications, and it can make an enormous difference in the perception of how good the voice quality is. It may employ an algorithm chosen from a range of uncomplicated to very sophisticated and, hence, it is essential that the book allocates space to this critical feature.

Chapter 5 is devoted to the subject of noise reduction. Noise reduction is the most complicated feature among the voice-quality assurance class of applications. It also requires a higher-level understanding of mathematics. The discussion, however, substitutes numerous mathematical expressions for intuition, ordinary analogies, and logical reasoning, supplemented by graphical and audio illustrations.

The analysis gets underway with the definition of noise, a definition consistent with the principles and characterization employed by a typical noise-reduction algorithm. It then introduces and explains the mathematical concept of time and frequency domains and the transformation process between the two. Once the reader is armed with the understanding of time and frequency domain representations, the analysis proceeds to a discussion of the noise-estimation process. The presentation then moves ahead to examine the suppression algorithm, which employs the noise-estimation results in its frequency-band attenuation procedures. The next segment contains a presentation covering the final algorithmic steps, which involve scaling and inverse transformation from frequency to time domains.

The next section in Chapter 5 reflects on key potential side effects associated with noise-reduction algorithms including treatment of non-voice signals. It points to key trade-offs and adverse-feature interactions that may occur in various GSM and CDMA networks – a subject that is covered much more thoroughly in Chapter 12 – *Trouble shooting and case studies*. The final section offers an examination of the network topology and placement of the noise-reduction application within it.

Chapter 6 is dedicated to the subject of level-control optimization. The presentation is divided into three parts and an introduction. The presentation starts the ball rolling in the introduction by defining standard methodologies for measuring and quantifying signal levels. The first part deals with automatic level control (ALC), how it works, and its placement within the network. The second part describes the adaptive level control, a.k.a. noise compensation (NC), how it works under different codecs, and where it is placed in the network. The third part describes the high-level compensation procedure along the same outline.

Part III – *Wireless architectures* – is essential to the understanding of the VQS contributions, its relevance to the delivery of high-performance mobile voice-communications service, its compatibility with data services, and its place within the 3G network architecture.

Part III commences with Chapter 7: Mobile-to-mobile stand-alone voice-quality system architectures and their impact on data communications. The chapter reviews the optional placements of the voice-quality system functions relative to the mobile-switching center and the base-station controller, since placement impacts voice performance, applications, deployment cost, and data-detection algorithms. The second part of the chapter presents an analysis of the techniques employed by a voice-quality system when coping with data communications without interfering or blocking its error-free transmission. The analysis includes descriptions of data-detection algorithms based on bit-pattern recognitions. The scope encompasses circuit-switched and high-speed circuit-switched data (CSD and HSCSD respectively) services as well as tandem-free operations (TFO).

Chapter 8 – *The VQS evolution to 3G* – portrays the 2G- and 3G-network topologies and their impact on VQA feasibility and architecture. It provides an evolutionary examination of the process leading to 3G from the popular 2G wireless architecture. It presents a parallel progression of placement and applicability of the voice-quality system that supports the evolving infrastructure.

Part IV offers a practical guide for the service provider. It guides product managers who are faced with the dilemma of whether or not they should invest in voice-quality systems and, if so, what performance tests they ought to implement, and what system capabilities they must require from their supplier.

Chapter 9 – *A network operator guide to testing and appraising voice-quality systems* – describes test and evaluation procedures of performance associated with the various voice-quality applications. The telecommunications-equipment marketplace is filled with a variety of echo-canceler (EC) and voice-quality systems (VQS) promoted by different vendors. Noticeably, the characteristics and performance of these products are not identical. In addition, the non-uniformity and arbitrary judgment that is often introduced into the EC and VQS product-selection process makes the network operator's final decision both risky and error prone. Chapter 9 describes the criteria and standards that are available to facilitate methods for objectively analyzing the benefits of EC and VQA technology when confronted with multiple EC and VQS choices. The scope includes procedures for evaluating performance of electrical (hybrid), acoustic-echo control, noise reduction, and level optimization via objective, subjective, laboratory, and field-testing.

Chapter 10 – *Service-provider's system, management, and delivery requirements* – presents a basic template that may be used by service providers as part of their request for information from vendors. The chapter elaborates on the various elements beyond voice performance that make the voice-quality system easy to manage, and easy to integrate within the operation of the network. The information is rather dry, but highly useful as a reference. Readers of this book who are not interested in the system-engineering requirements may skip this chapter in their pursuit for understanding of the magic that makes voice-quality systems enhance speech communications.

Chapter 11 – *Making economically sound investment decisions concerning voice-quality systems* – discusses key differences between stand-alone and integrated systems. It points to the pros and cons of investing in a full VQS suite versus a minimal set containing a hybrid-echo canceler only, and it closes the chapter with a simple model providing guidelines for assessing return on investment.

Part V – *Managing the network* – presents an extensive account of conditions that must be accommodated for healthy voice communications to come about. The presentation is carried out as if the reader is in charge of running a mobile-switching center where all equipment, including voice-quality assurance gear, has been operating satisfactorily up to that moment in time when the specific trouble at issue has been reported. The specific troubles are analyzed for root cause and remedial actions.

Part VI – *Afterthoughts and some fresh ideas* – concludes the book with a discussion of application ideas and inventions that may be incorporated into forthcoming voice-quality systems.

Chapter 13 presents an application concept referred to as *Tracer probe*. This application may be highly useful in promptly detecting and isolating network troubles without the tedious and laborious effort of the current methods and procedures.

Chapter 14 presents an application concept, *My sound*, that assigns voice-processing control to subscribers of a preferred class, where they may be able to play games with sound effects.

Chapter 15 presents procedures for evaluating voice quality of new codecs and the impact of voice-quality systems on their overall voice quality. The main reason for including this chapter in this part is the fact that even as of today, after many studies and standard works that have been published, there are still skeptics. And they want to verify new codec introduction in their own network, and justify the notion that noise reduction and optimal level control do in fact make a difference in the perception of voice quality. For those who do not take published studies as truth, we outline a way to come up with their own studies.

Chapter 16 presents a concept I named the *Theory of sleep*. The discussion challenges the concept of present system default settings where voice applications are turned on and only disabled when detecting non-voice transmission. It presents a complementary concept, whereas certain voice applications are permanently disabled unless the system detects voice transmission. The writing does not recommend this approach. It simply challenges existing paradigms.

Part VII – *Recordings* – provides a summary of the recordings on the accompanying website, www.cambridge.org/9780521855952. These recordings are intended to highlight and clarify particular aspects of sound and voice communications.

A bibliography provides a list of references relevant to the material discussed throughout the book.

The glossary contains a compendium applied to echo cancelation and voice-quality enhancement technology. Uncertainty, confusion, and misinterpretation often result when acronyms or terminology that are specific to the field of echo cancelation and voice-quality enhancement are used freely. The glossary is a collection of commonly used acronyms and other terms, accompanied with brief descriptions. This material is intended to provide clarity and insight into the unique language of echo cancelation and voice-quality enhancement.

Part I

Voice-quality foundations

Part I reviews the major voice codecs, their history, and their relative perceived quality. Voice-coding architectures are the building blocks of transmitted voice. They are the core that shapes the characteristics and quality of transmitted speech. Nevertheless, they are treated in this book only as background to the main subject, which deals with impairments due to transmission architecture and environment, and their corresponding remedies that immunize and repair any potential or actual spoil. Since the effectiveness of the various remedies depends on that underlying coding, it is essential that these designs be understood so that remedies can be fine tuned and customized to suit the particular characteristics of the underlying voice architecture.

Cambridge University Press
978-0-521-85595-2 - Fundamentals of Voice-Quality Engineering in Wireless Networks
Avi Perry
Excerpt
More information

1 An overview of voice-coding architectures in wireless communications

1.1 Introduction

It must have happened to most of us. At some point through our lives we came across someone whom we deemed an "audio nut." (Some of us may have even impersonated that one special character.) That singular person would not listen to music unless it was played back on an exceedingly pricey hi-fi system. He or she actually did hear a titanic disparity in sound quality between what we would be taking pleasure in on a regular basis and what he or she regarded as a minimum acceptable threshold.

In all probability, we might have adopted the same mind-set had we been accustomed to the same high-quality sound system. It is a familiar human condition – once a person lives through luxury, it grows to be incredibly hard getting used to less. How quickly have we forgotten the pleasure we took in watching black-and-white TV, listening to the Beatles on vinyl records, Frank Sinatra on AM radio, etc. But hey, we have experienced better sound quality and, thus, we refuse to look back.

Wireless telecommunications is entering its third generation (3G). Infancy started with analog communications. It developed through childhood in the form of GSM and CDMA, and has crossed the threshold to puberty with cdma2000 and W-CDMA – its third generation. Voice quality in wireless telecommunications is still young and looking up to adolescence, but technology has advanced appreciably, and most of us have been content with its voice performance. We got mobility, connectivity, and voice performance that had never been better, except of course, on most wireline connections. But we were forgiving. We did not notice the relative degradation in voice quality. The little price paid was well worth the mobility.

Why then, why is voice quality over wireline networks – the older, almost boring, plain old telephone service (POTS) – so much better than the newer, cooler, wireless mobile-phone service? Why is there more distortion, more delay, more occasional echo, different noises that come and go, plenty of speech clipping, noise clipping, and voice fading, that we have been putting up with just to gain mobility? Why does it have to be this way? Or does it?

This chapter provides a partial rejoinder to the questions above. It provides an overview of the transmission plans and coding schemes that command the transformation of natural speech to an electronic signal and back into speech in wireless networks. It explains in simple terms how this transformation alone produces changes in the quality of transmitted voice. It focuses on the coding and decoding methods that transform sound into a binary code and the network infrastructure that carries this digital signal to its final destination where it is transformed back into a close relative (rather than a perfect replica) of the original sound. It also addresses the transmission media that adds a variety of distortions to the original signal before it arrives at its final destination.

1.2 Pulse-code modulation

To date, pulse-code modulation (PCM)[1] and the PSTN G.711 ITU standard[2] have been the most common digital-coding scheme employed in wireline telecom networks worldwide. Voice quality associated with PCM is regarded as "toll quality" and is judged to be unsurpassed in narrowband telecommunications. Although PCM exhibits a degraded voice quality relative to the natural voice, it is perceived to be rather fitting for the application of speech transport over the wires.

Pulse-code modulation occupies a band of 0 to 4 kHz. Considering that the young human ear is sensitive to frequencies ranging from 16 Hz to 20 kHz, the bandwidth allocated to a telephone channel seems to be spectrally depriving. Nonetheless, since the principal body of speech signals (energy plus emotion) dwell in the ∼100 Hz to 4 kHz band of frequencies, the 4 kHz channel provides a dependable nominal speech channel. (Musicians may not welcome listening to a Beethoven symphony over that channel. Some of the musical instruments may sound heavily synthesized and somewhat distorted. Speech, on the other hand, is not as spectrally rich as a symphonic performance; neither does it have the benefit of a comparable dynamic range.)

When the band-limited nominal 4 kHz channel is sampled at a rate equal to twice the highest frequency, i.e., ($4000 \times 2 =$) 8000 samples per second, then the sample takes account of the entire information contained in the sampled signal. Since each sample comprises 8 bits of information, the total bandwidth required for representing an analog signal in its PCM form is (8000 samples × 8 bits =) 64 kbps.[3]

Pulse-code modulation is a speech-coding method classified as waveform coding. This class of codecs employs extensive sampling and linear quantization representing the original analog signal in a digital form. The delay associated with PCM coding is negligible and is equal to its sampling rate, i.e., 0.125 milliseconds.[4] A round trip between two wireline (PSTN) phones involves one code/decode operation. Accordingly, the delay attributed to the PCM coding/decoding process is 0.25 ms. The significant portion of the delay experienced in wireline communications is, by and large, the result attributed to signal propagation over long distances.

1 K. W. Cattermole, *Principles of Pulse Code Modulation*, Elsevier, (1969).
2 ITU-T Recommendation G.711, *Pulse Code Modulation (PCM) of Voice Frequencies*, (1988).
3 As a reference point, it should be noted that a 22 kHz, 16-bit mono music sample requires a data rate of 44 100 samples per second.
4 8000 samples per second produce a single frame. The sampling frequency is 1/8000 seconds = 0.125 ms = 125 μs.