metaswitch

WE ARE NOT THINGS!

IoT: The Internet of Talk
The Future of Voice Communications

INTRODUCTION

Telephony is the lifeblood of the telecom industry. But that blood has been getting thinner in recent years, as alternative modes of communication, as well as Internet-based providers of basic calling-type services, have grown at the expense of ordinary phone minutes.

And while mobile telephony has extended communications to billions around the world, it has recently faced an additional challenge. The arrival of 4G and all-IP connectivity has prompted a replacement of legacy circuit-telephony infrastructure with VoIP, and specifically VoLTE. This has been a complex transition and is still ongoing.

But simply replacing old phone networks with newer IP-based ones is not enough.

The challenge for the telecom industry is to bolster and reinvigorate its voice services, not just switch over the infrastructure to IP – although if this can lower costs or release spectrum it will be welcome. The update will involve many approaches. In part, it involves improving and extending normal "calls," including using core telephony systems as platforms for developers and telcos' internal innovators. It will also involve non-traditional voice applications for both consumers and enterprise, such as better push-to-talk systems for field workers, or new forms of "social voice" (and often integrated video) for teenagers. New pricing models have tended to mean bundling of "unlimited" minutes or large bundles – but there may be new approaches for added-value capabilities in future too.

But perhaps above all, it will mean blending voice with contextual and analytical data, to improve the usability and utility of phone calls, helping them compete better against the various substitutes. There are many sources of value in helping telephony meet the underlying needs of the people on the call – in scheduling, recording, blending in external sources of information, analyzing content and tone and so on.

Voice telephony isn't going away. But it needs to be enhanced and opened up, if it is to remain as relevant as in the past. It needs to go "beyond plain-vanilla calls."

THE HISTORY OF VOICE COMMUNICATION

"Mr. Watson – come here, I want to see you!". This was the first "distant" voice interaction, uttered by Alexander Graham Bell in March 1876 to summon his assistant when he spilled acid in his laboratory and needed help.

Yet despite the history, this landmark occasion was not really a "phone call" as we now understand it. It was an order, not a conversation. It was one-way (Watson's terminal did not have a microphone, only a speaker).

There was no service provider involved, just a wire owned by Bell. There was no numbering, no ringing and no polite etiquette ("Hi, this is Alex calling"). It was more like a push-to-talk system, to use a modern equivalent.

Other early examples of voice communication emerged – including the Theatrophone, invented in France five years after Bell's innovation, in 1881. It allowed people to listen to the opera remotely, from another building some distance away. Live audio streaming is therefore nothing new, either.

Over the subsequent decades, these interactions mostly matured into the familiar "call" format, becoming commonplace and helping transform the world's business and society. A huge telephony industry sprang up and integrated with the already-existing telegraphy business. (AT&T stands for American Telephone & Telegraph, and was founded in 1885.) The business model evolved mostly to a subscription model for "line rental," plus some fee-for-call charges, usually based on a per-minute cost, varying depending on distance or time of day. These largely reflected the cost structure of human switchboard operators (paid per hour) and the length of cabling involved for different calls. Over time, this model got further entrenched by regulators, licensing and market-monitoring approaches.

Subsequent evolutions of telephony drove the construction of new, dedicated networks. At first, they used humans to connect wires on switchboards, then automated analog switches, then digital, and more recently changing the signalling and transmission of voice from direct circuits to packetized IP data. The latter reflects the growing trend for networks to be used primarily for data connectivity – for enterprises, dedicated telecom/TV services and of course Internet access. Telephony is then layered on top.

Alongside the evolution of fixed telephony from service providers ("telcos" and the "PSTN"), the last few decades have seen the parallel evolution of business telecom systems (PBXs – private branch exchanges) and then unified communications, conferencing, hosted voice and a broad range of other enterprise collaboration options.

And of course, alongside telephony linking fixed end-points, we have also seen an immense explosion in mobile communications over the past 20 years or so. Until recently (see VoLTE, below), this all used circuit-switched systems for phone calls, albeit often using soft-switches that used IP signalling alongside non-IP media streams. Other mobile voice services such as push-to-talk have also grown, albeit only in small niches.

Outside of telco or enterprise telephony, some other voice platforms have also taken minor roles. Privately owned mobile communications systems (eg, walkie-talkies, secure trunking, CB-radio, etc.) have been used by groups such as emergency services, taxi dispatchers and enthusiasts – but have not really threatened the vast bulk of telco mobile telephony services and the main form of "voice" communications.

PRESENT-DAY TRENDS IN VOICE

Fast-forward to 2017 and six major new trends in and around voice have emerged. Of these, the first is being driven by the traditional telecom operators, vendors and standards bodies. The other five are mostly "adjacent" or "orthogonal" trends that sometimes intersect with the telephony market, although some telecom operators have adopted roles in those fields as well.

- **Growing adoption of carrier-grade VoIP** by telcos for mainstream standardized telephony, replacing circuit-based technology both in fixed networks with SIP-based solutions, and now -- gradually, but accelerating -- in 4G mobile networks with VoLTE, or over Wi-Fi with 3GPP VoWiFi. This transition has been the main focus area for many telcos' voice businesses in recent years, and its implications are examined in greater depth below.

- **Fragmentation and disruption of voice-based telephony** by new, mostly consumer-oriented VoIP-based applications and services – Skype, Viber and WhatsApp are some of the best known. Some are closely related to traditional "calls," but others use different formats – such as blending in social features, as well as video. Many are free to use for basic communications, with monetization through advertising or interconnection with the PSTN. They are increasingly treated as "first-class citizens" on a par with telcos' telephony in smartphone OSes (eg, with Apple CallKit).

- **Enterprise collaboration:** Focused efforts by enterprises and their suppliers to integrate telephony into broader business-process and collaborative toolkits. Developers embed call logic directly into contact centers, conferencing and a variety of workgroup platforms – as well as various other business solutions such as Microsoft Office 365. A broad array of UC (unified communications) suppliers offer both premise- and cloud-based systems (UCaaS), either sold direct, or provided via telcos as channels/integrators.

- **cPaaS:** The rise of communications platforms-as-a-service (cPaaS) for developers. Using cloud-based anchor points and flexible APIs, voice calls (and messages or video) can be added to web processes, mobile apps and other use cases easily. Twilio is probably the best-known example, but there are numerous others, including some telco-run PaaS efforts.

- **Non-telephony comms:** Broad adoption of substitutive non-telephony/non-voice forms of communication, notably including email, Internet- or app-based messaging, social networks, push

notifications and video-based calling. This has occurred in both the fixed world and, more recently, mobile. For instance, whereas in the past many people might have phoned for a taxi pick-up, many now use an app like Uber instead of a call.

- **Voice recognition and response**, either integrated into businesses' IVR systems, or as a "front end" for broader cloud and AI platforms such as Apple's Siri and Amazon's Alexa, built into iPhones and Echo devices respectively.

The net result has been a much greater overall use of communications and voice. But the corollary has been that "ordinary phone calls" (whether circuit or VoIP/VoLTE) have started to lose out, or just be viewed as a lowest common-denominator, where no better alternative is available.

VoLTE NEEDS TO BE MORE THAN "JUST A PHONE CALL"

The move from circuit-switched infrastructure to VoIP has been a huge move for most telecom operators. It has required a change in technology, and also often culture and organization, bringing together voice and data teams. While the process is well under way in the fixed/cable market and has been for some time, the mobile sector is still in the early phases of transition.

But if the fixed telecom world has had challenges, they have paled into insignificance compared with VoLTE, which some mobile operators have described as the hardest project they have ever undertaken. There are some eventual upsides -- potentially lower operating costs and physical space requirements compared to older systems, and the ability to "re-farm" spectrum, when 2G/3G networks can finally be switched off. Faster connection times are welcome, while for some networks, this will be the first time users have experienced HD-quality voice.

But rolling out VoLTE has involved not just implementing a SIP/IMS core, but also adaptations to the packet core, policy infrastructure, OSS/BSS – and dealing with the

radio network's ability to support VoIP.

This has meant considering coverage and capacity at different frequency bands, prioritizing specific radio bearers for latency-sensitive traffic, and dealing with handover to/from circuit telephony where 4G coverage ends and handsets drop back to 3G or 2G. (Most existing 4G networks have historically used CSFB – circuit switched fallback – for telephony. Blending it with VoLTE has needed yet another new technology called SR-VCC; single radio voice call continuity). Other work has ensured compliance with regulatory regimes (eg, emergency calls and lawful interception), roaming, interconnect and replicating old IN (intelligent network) services and other features in the new VoLTE infrastructure.

All this has been necessary as "table stakes" for 4G basic telephony. But it still doesn't address the underlying issue: voice telephony needs to be more than just phone calls, if it is going to be more than just a low-margin, lowest-common-denominator service. Prices for voice telephony are falling, but call volumes are flat or falling, despite that. Many parents complain that their children only communicate via text or messaging apps, not calls. Even among older groups or in business, there is a declining tolerance for receiving unexpected, interruptive, anonymous, or worthless spam calls.

What is really needed is for the transition to VoLTE, or the new behind-the-scenes architecture changes, to enable the operator to address some of the other five trends highlighted above. Otherwise, VoLTE faces "peak telephony" just as it is launched.

Total Outgoing Minutes of Fixed and Mobile Telephony *(billion)*
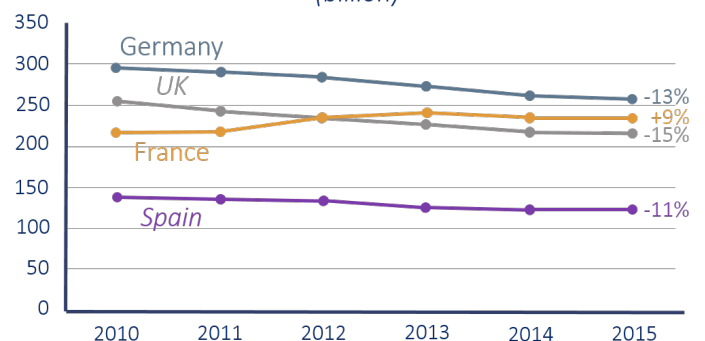
Peak Telephony: The crumbling telephony 101 establishment. Source: Disruptive analysis

EXPLOITING NEW SOURCES OF VALUE IN VOICE

In order for telco-offered voice services to maintain relevance, and hopefully grow revenues and loyalty, they need to find new "sources of value" in communications, such as helping users to better schedule, manage, filter, analyze, record or log their voice events. Context, purpose and user experience matter for voice, not just transmission quality.

Service providers should be thinking about the underlying reason that people want to make or receive a call, and the mechanism by which it is arranged, and help make the process and outcome easier or better.

To better understand the type of idea, think about the following scenarios:

- "The person you are calling is in a different time zone – it's dark and their phone is on charge. Press 1 to go straight to voicemail, or 2 to ring them anyway"
- "I'm on a beautiful beach. I wish you were here. Listen to the waves behind me"
- "I'm in a noisy bar, sorry. You were saying we need to reschedule tomorrow's meeting?"
- "Let's dial in John to the conference. He's deaf – are the rest of you all OK with real-time text captioning and text-to-speech for the rest of the call?"
- "We are recording the call for quality purposes." "So am I"
- "So, Erica, it sounds like this is a really good proposal, but maybe I should think overnight." [One-way automated whisper to Erica: "Close the deal now!"]
- "So, to move to my next slide, I think you can all see this year's trend --" <Wearable buzzes urgently – you remember that means "slow down & speak more clearly">
- Voice-analytics app screen display: "Last year, your boyfriend responded to voicemails in an average 17 mins. This year, it's 45 mins."
- "Am I speaking to a real customer service agent, or an AI? Please transmit your human/robot identity details for confirmation"
- "You want to book an appointment for the doctor to examine your throat? Could you please inhale deeply and then say the following sentence clearly, so we can perform initial diagnosis on your vocal projection immediately"

These are merely a handful of the hundreds or thousands of ways that the voice experience could be improved beyond the basic "call." Some will have mass appeal, while others will be minor niches. Some can be monetized directly, while others will be features that improve experience and brand loyalty. Some require all participants to have access to the same capabilities, while others can be deployed unilaterally. Some will require privacy opt-ins.

Certain possibilities can be ignored for the purposes of this paper: a karaoke app is based on "voice," but few would suggest that the phone network is the best platform on which to build one, nor telcos the best-placed providers to offer it. But many others fit better with telcos' customer bases and ambitions, at least in theory.

But at the moment, all these types of capability are either impossible to deliver, or are provided mostly by external developers and applications. All require an understanding of the network technology, plus also software development, and an appreciation of user behavior, as well as the capabilities of devices outside the telco's "comfort zone." They need design skills as well as engineering. They cannot all be neatly standardized, or packaged by vendors in a simple shrink-wrapped product. They need some hard thinking and value-adding work.

Three types of context can be blended into voice communication. Source: Disruptive Analysis

There are four broad sub-categories here:

- Application-embedded communications, either on the web, inside a mobile app, or integrated with another software or cloud application. In some cases, voice is fundamentally integral to a given app or workflow (eg, a sales call), while in others it is a secondary feature (eg, customer support in a travel-booking app).
- Novel formats for interaction, such as a one-way "whisper mode" overlay, or walkie-talkie capabilities
- Contextual and analysis-driven communications. This is likely to grow in importance and sophistication, as machine analytics and IoT sensors help make voice interactions more effective. For example, a smart contact-center system could highlight those callers who prefer brevity and "getting to the point," while others might appreciate pleasantries and humor. A conferencing application might be able to use IoT sensors to determine if you are in a room, or in the street – and apply a "mute" function automatically. Video communication may be unsuitable if you seem to be in a vehicle.
- Managing pre-, post- and mid-call processes, such as scheduling and logging. A collaboration tool with

"learning" abilities might highlight people outside of a known team, who could bring a particular skill, contact or experience set – and send them an invitation to join a conference call. We may be able to get auto-generated subtitles for people with poor network connections or impenetrable accents.

Strictly speaking, not all of these need VoIP or VoLTE as the underlying transport for the voice traffic – but the benefits of leveraging a standard-based infrastructure maximizes the potential of these techologies.

CONCLUSIONS AND RECOMMENDATIONS

The nature of voice applications and services is changing fast – as is the behavior of users who are looking for more than just a "vanilla phone call." If telcos want to defend their historic primacy, they need to swim with the flow, rather than hoping that regulators try to hold up the tide for them.

While telephony still has the advantage of ubiquity, it is increasingly seen as just a lowest common denominator when nothing better (or cheaper) is available. And as networks, phones, apps, and users become more sophisticated, this situation is only likely to worsen over time.

Operators should view carrier VoIP and VoLTE implementation as a timely catalyst for an overall rethink of "voice strategy." What new capabilities and services do they want to deliver, and how should it be done?

Beyond the technology, a lot of the problems are human – service providers need to start thinking of voice as a "raw ingredient" for a hugely long menu of possibilities. This needs "product managers" for voice that go beyond just implementation of new 3GPP standards or GSMA specifications. They need imagination, insight into real customers' problems, a willingness to seek out partners, and the software skills to implement and deploy on their own if necessary.

There is a need to think of communications services and features as going well outside the core network and application servers as well. Much of the value is added "at the edge" – for example, by integrating with local capabilities and APIs on phones (motion, screen orientation and so

on). Just relying on the "native" telephony stack on the smartphone is not enough. The cPaaS model uses "normal phone calls," but the real value lies in the application platform and integration, rather than the minutes of voice transported.

In many cases, there will also be a need to blend in various forms of messaging capability as well – plus, perhaps, video in either one- or two-way variants. Again, this will need to be appropriate for the given use case, customer group, devices – and application requirements and limitations. It could involve custom development, or use of existing services like SMS, Apple/Android push notifications, email, MQTT, WebRTC, XMPP and so on. (Disruptive Analysis is not a believer in RCS, ViLTE or the GSMA "universal profile," it should be noted.)

There is no "one size fits all" for telcos wanting to remain relevant – and generate new revenue – from voice communications. While there is a broad drift toward IP communications, contextual data around voice, and exploitation of developer platforms, the details and priorities will vary hugely for each operator.

Most obviously, fixed and mobile operators have very different drivers – as do those with large existing bases of enterprise customers, or which have well-developed plans for adopting NFV. Those telcos with a strong partnership and innovation ethos -- especially with external software providers – will have different opportunities from those with less internal expertise.

In summary: there is no single "silver bullet" for voice communications. But the idea that voice has no future relevance or value for telcos is wrong: there are plenty of options, but they need a more wide-reaching strategy to identify and exploit. The implementation of VoLTE or fixed-VoIP should be seen as an opportunity to start with a clean sheet of paper.
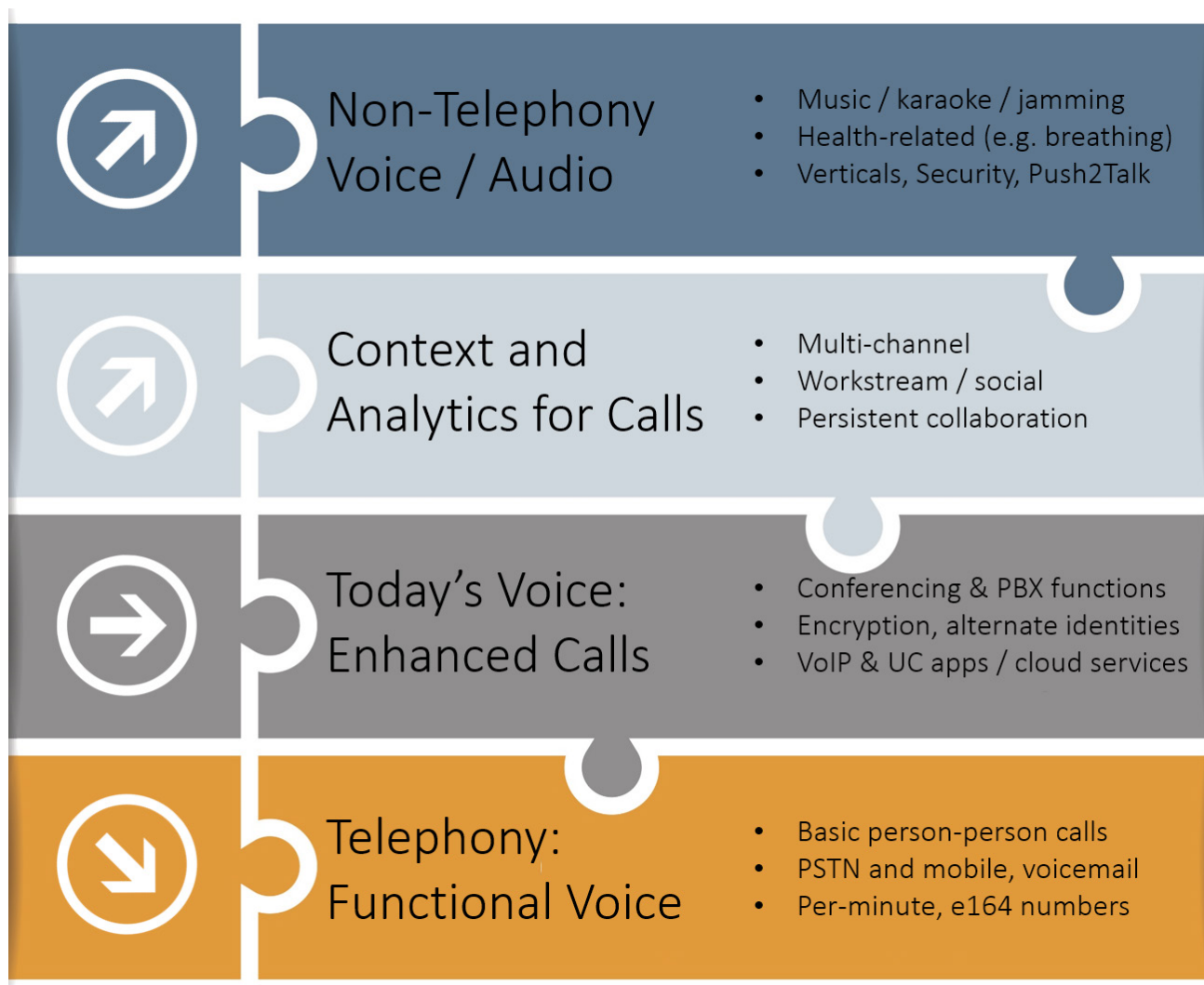
Multiple layers for voice, beyond basic functional telephony. Source: Disruptive Analysis