An Alternative Approach for Investigating the Impact of Mobile Phone Technology on Speech

Esam A. S. Alzqhoul, Balamurali B. T. Nair, Bernard J. Guillemin

***Abstract*—Creating a database of mobile phone speech recordings is the first step to studying the impact of mobile phone networks on various speech applications. The conventional way of creating such a database involves transmitting speech between two mobile phone devices across an actual network. However, this can only encompasses a small subset of all possible scenarios that could occur in the network. To better represent these scenarios, the speech would need to be recorded at many different times of the day, at many different locations, and under both stationary and mobile conditions. In this paper we propose an alternative approach for comprehensively simulating mobile phone speech. This is based on a detailed understanding of the various aspects of a mobile phone network that directly impact the speech signal. The differences between the Global System for Mobile Communication (GSM) and Code Division Multiple Access (CDMA) networks in respect to their ways of handling speech are explained. The speech codec in these networks is identified as the only component that is directly responsible for the quality of the coded speech and any changes that might occur to it. These codecs have many modes of operation, but changing these modes is linked to global factors occurring in the network. These global factors have been incorporated into the design of two software platforms which can be used to create mobile phone speech under various realistic scenarios in the GSM and CDMA networks. The impact on the speech signal for some of these aspects is presented.**

***Index Terms*— mobile phone networks, mobile speech codecs, dynamic rate coding, frame loss, background noise.**

I. Introduction

THERE has been a rapid growth in mobile phone communications over the last few decades. The use of mobile phones goes beyond simple communication and includes a variety of demanding applications including user authentication, forensic voice comparison, speech-to-text and speaker recognition [1], [2]. In these applications it is crucially important to understand the impact of the mobile phone network on the speech signal and to what extent this might compromise the performance of these applications.

Clearly the first task for any study in this arena is collecting a database of mobile phone speech recordings representative of the very large number of scenarios that could occur during the transmission process.

Manuscript received July 5, 2014; revised July 30, 2014.

E. A. Alzqhoul is a PhD student at the Dept. of Electrical and Computer Engineering, Faculty of Engineering, The University of Auckland (e-mail: ealz002@aucklanduni.ac.nz).

B. B. T. Nair is a PhD student at the Dept. of Electrical and Computer Engineering, Faculty of Engineering, The University of Auckland (e-mail: bbah005@aucklanduni.ac.nz).

B. J. Guillemin is a Senior Lecturer at the Dept. of Electrical & Computer Engineering, Faculty of Engineering, The University of Auckland (phone: (09) 373 7599 Ext. 88190, Fax: (09) 373 7461, e-mail: bj.guillemin@auckland.ac.nz).

A common approach to achieving this involves transmitting speech across an actual mobile phone network. Though this might seem an obvious and appropriate strategy, it is in fact far from ideal because it is likely to encompass only a small subset of all possible transmission scenarios. The impact of a mobile phone network on the speech signal is dependent on many highly variable factors such as the wireless channel conditions present, capacity (i.e., number of users accessing the system), interference levels, etc. These factors are both time varying and location dependent. Hence, in order to collect a truly representative mobile phone speech database using this approach, it would be necessary to conduct a large number of experiments at different times of the day and in many different locations, for example, city centers (typically referred to as urban canyons), urban and rural areas. Even then, there would be no way of knowing whether all possible transmission scenarios had been represented because such information is not available in the received speech signal.

There are a number of mobile phone technologies in use today, the most widely used being Global System for Mobile communication (GSM) and Code Division Multiple Access (CDMA). According to recent surveys the total number of users utilizing GSM technology worldwide is approximately 3 billion, compared to 500 million using the CDMA network. The GSM network is no doubt the most popular worldwide with operators in 212 countries [3]. Nonetheless, the CDMA network is still very popular in North America, China and India with a presence in 118 countries worldwide [4].

The underlying design of these networks is fundamentally different, leading to subtle, but important, differences in the way they impact the speech signal. So experiments would need to be repeated using these different network technologies as well. Overall, this approach to producing a mobile phone speech database by transmitting speech over actual mobile phone networks is likely to be very time consuming and far from comprehensive.

In this paper we present an alternative, less time consuming, approach for creating mobile phone speech databases which has the potential to encompass a large set of all possible transmission scenarios. With this approach, speech data is passed through the speech codec of these networks under various possible modes of operation, taking into account the underlying rules under which these modes might be initiated. Using this approach it is possible, for example, to relatively easily transform existing speech databases into mobile phone speech databases. The approach is based on the fact that it is only the speech codecs in these networks that directly impact the speech signal, and thus any changes that might occur to it [5]. The codecs have many modes of operation, but changing which mode is in operation at any particular time is a process controlled and initiated by the network as a whole. Constraints in respect to how these modes can be changed have been built into these codecs, as well as into the design of the networks. So, though the number of possible codec operational modes is large, it is possible to group these into a much smaller number of representative scenarios.

The most widely used speech codecs in these networks are the Adaptive Multi Rate Codec (AMR) in the GSM network [6] and the Enhanced Variable Rate Codec (EVRC) in the CDMA network [7]. The differences between these two codecs in respect to their ways of handling the speech signal, and thus their impact upon it, as well as their different operational modes, are discussed in Section 2. This section also provides an overview of the most representative transmission scenarios that could be present in these networks. Section 3 discusses two software platforms that have been developed in order to simulate typical scenarios in the GSM and CDMA mobile phone networks and their resulting impact on the speech signal.

II. AMR and EVRC Operational Modes

For both the AMR and EVRC codecs speech is segmented into 20 ms frames. There are three key aspects of the operation of these codecs which ultimately impact upon speech quality: (i) Dynamic Rate Coding (DRC), (ii) Frame Loss (FL) and (iii) Background Noise (BN). The design of the software platforms presented in this paper is based upon a detailed understanding of these aspects and how they are linked to what is happening in the network as a whole.

A. Dynamic Rate Coding (DRC)

As the name implies, this aspect concerns dynamically changing the bit rate at which speech is coded in response to certain events in the network. Bit rate is one aspect which impacts speech quality. In the GSM network, DRC is primarily a function of changing channel conditions, but also to a lesser extent on the number of users present in a cell site. The GSM network instructs the AMR codec to increase or decrease the speech coding bit rate in response to good or bad channel conditions, respectively [5], [8]. Though the AMR codec has the capability of coding sequential speech frames at different bit rates (i.e., changing bit rate every 20 ms), a constraint has been built into the GSM network limiting bit rate changes to a maximum of every 2nd frame (i.e., every 40 ms). The result is that for any specific call there is in fact a large number of possibilities in terms of resulting speech quality from very poor to very high. However, our approach is to broadly classify the resulting speech quality into one of three categories corresponding to poor, average and good channel condition, respectively. The choice of these categories is motivated by constraints imposed on the DRC process, as discussed below.

The AMR codec can operate at one of the following eight bit rates: 4.75, 5.15, 5.90, 6.70, 7.40, 7.95, 10.2 and 12.2 kbps [8]. Switching between bit rates is performed in a highly controlled and systematic manner, a process best understood through the concept of Active Codec Set (ACS). For any particular call, the choice of bit rates is limited to a subset of a maximum of four bit rates, called the ACS. For example, the ACS might contain 1, 2, 3 or 4 bit rates and these are assigned according to the channel conditions present when a call is first established [8]. (Note: the ACS can change if the user moves from one cell to another, a process called handover.)

There are additional constraints imposed upon the ACS. It can be selected either from the lowest five bit rates, in the case of Half Rate (HR) channel mode, or from all the eight bit rates, in the case of Full Rate (FR) channel mode. The FR or HR modes are also selected at the commencement of a call, but the choice of these is dependent upon the number of users present in a cell site, FR being chosen when the number is small, HR when it is large [9]. The use of low bit rates in the HR mode increases the robustness of transmission, but the resulting speech quality is relatively poor. Under good channel conditions, the ACS is normally formed from the highest five bit rates of the FR mode. With this scenario, the resultant speech is generally of high quality. When channel conditions are average, the ACS can be formed from either the FR or HR modes, resulting in medium speech quality. To give an example of the number of ACS that could occur, if channel conditions are poor, then the ACS contains bit rates selected from the HR mode (the lowest five bit rates). On the assumption that the ACS assigned is three bit rates long, then the number of possible ACSs which can be formed using five bit rates would be: $\frac{5!}{3!(5-3)!} = 10$.

Finally, an additional constraint is imposed upon DRC in respect to the first bit rate used in a call, referred to as the initial codec mode (ICM). If the ACS contains four bit rates, the ICM must be the second lowest in the set; if the ACS contains less than four, then the ICM will be the lowest bit rate in the set [8].

It should be clear from the above that the DRC process associated with the GSM network is quite complicated. Nonetheless, we believe that its impact on the speech signal can be simplified into three broadly representative speech qualities, namely low, medium and high quality. This is one of the simplifications that have been built into our software platform.

The process of DRC is somewhat simpler in the CDMA network. The EVRC codec produces speech frames at one of only four bit rates: 0.8, 2, 4, and 8.55 kbps. These bit rates are primarily chosen in accordance with the speech characteristics present in each input speech frame (i.e., voiced, unvoiced, transient and silence), but the network capacity condition present at the time (i.e., number of users in the cell site) also plays a part. The result is an Average Data Rate (ADR), which in turn determines speech quality. As with the GSM network, there is a then a continuum of resulting speech quality from very poor to very high. We are of the view that this also can be broadly classified into one of three categories: low, medium and high speech qualities corresponding to high, medium and low capacities, respectively. Again this simplification is motivated by codec operational aspects to be discussed next.

The EVRC codec uses three modes of operation, known as Anchor Operating Points (AOPs): OP0, OP1 and OP2.

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The AOPs control the general behavior of the codec and each one of them uses a different set of algorithms for coding the speech signal. The EVRC codec can be instructed by the CDMA network to achieve a certain ADR every 20 ms. The ADR value is then translated by the codec into one of the three AOPs, as shown in Fig. 1 The ADR can take on any value between 4.8 kbps to 9.6 kbps. Under low capacity conditions, ADR values in the upper-half of this range results, and this is achieved by switching between OP0 and OP1 in a predetermined manner. Such a scenario will produce a relatively high quality speech. When the network is congested with users (i.e., high capacity), the ADR values are selected from the lower-half of this range, which is achieved by switching between OP1 and OP2. This in turn produces a low speech quality. Under medium capacity conditions (i.e., the cellular site is neither congested nor having a small number of users), ADR values can be selected from the entire range of possibilities. This in turn is achieved by switching between OP0, OP1 and OP2 in a pre-determined manner, resulting in medium quality speech.

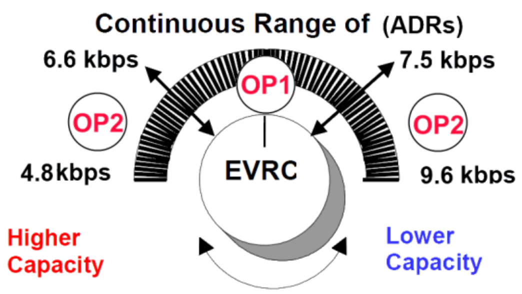

Fig 1. Mapping of ADR values into AOPs in the EVRC codec.

B. *Frame Loss (FL)*

The wireless channel in mobile phone networks is typically very poor and this often results in frames being lost or corrupted during transmission. Highly sophisticated mechanisms have been developed in both the GSM and CDMA networks to mitigate the impact of this on speech quality [10]. Specific strategies for achieving this in the GSM and CDMA networks differ, but there are some common features. If a frame is received in error, error correcting processes such as convolutional coding try to recover the missing data [11]. If this process fails, or indeed if a frame is lost completely, a new frame will be created to replace it. The data in this frame could be either a repeat of a previous good speech frame, or an extrapolation from past good frames received. Artificial segments of up to 16 frames (i.e., 320 ms duration) could be generated as a result [12], but the amplitude of these frames is systematically reduced in a nontrivial manner in order to minimize any unpleasant perceptual artifacts that might occur. In both networks the amount of frame loss is continuously monitored and if this exceeds 10-15% over short time frames (a value known to result in severely compromised speech quality), the call will be dropped [13].

At the outset it needs to be emphasized that unlike the frame-loss strategy for the CDMA network (to be explained in the next section), the frame-loss strategy for the GSM network leaves the original bit rate for individual frames unchanged. The spectral data for a lost frame in the GSM network is derived from the last 'Good' frame received. On the other hand, the amplitude adjustment process mentioned works on a sub-frame basis (i.e., each 20 ms speech frame is segmented into four 5 ms sub-frames). When a frame is lost, the amplitude of its subframe is adjusted using a history of the last five sub-frames. Further, amplitude discontinuities resulting from this process, that could lead to unpleasant perceptual artifacts, are minimized by the use of sophisticated strategies [10].

In the CDMA network, if a frame that was preceded by an active speech frame (i.e., not a silence frame) is lost or irrecoverably corrupted, that frame is replaced by an artificially generated frame at 8.55 kbps, irrespective of the bit rate associated with the lost or corrupted frame. If the preceding frame was coded at 8.55 kbps, then the spectral data for this new frame will be the same as that of the last good frame received, otherwise a sophisticated bandwidth expansion is applied to match the higher bit rate. The amplitude of this new frame is made the same as that of the preceding frame. However, if successive frames are lost, their spectral data will be derived from the last 'Good' frame with an associated reduction in amplitude by a factor $(0.75)^{N-1}$, where N is the consecutive lost frame number. Further, to minimize discontinuities in the speech signal, the pitch information of a 'Good' frame is adjusted if it was preceded by a lost frame [7].

C. *Background Noise (BN)*

Background noise at the transmitting end is frequently present in mobile phone communications. This situation becomes even worse when there is a reasonable separation between the mobile phone microphone and the speaker, such as in the case of hands-free terminals. In the GSM network, no strategies have been put in place to reduce the impact of background noise on the coding of the speech signal. Therefore, high levels of BN may well compromise the coding process, and consequently the overall quality of the coded speech. On the other hand, the CDMA network incorporates a unique strategy, namely Noise Suppression (NS), which subtracts the background noise present in every input speech frame [14], [15]. This process is primarily used to aid in the more accurate classification of speech frames into voiced, unvoiced or transient prior to the coding stage. The NS is repeated twice for every frame (i.e., every 10 ms) and uses a set of energy estimators and voice metrics to determine characteristics of the noise signal, and thus assist in its subsequent removal. Despite the beneficial effects of NS in removing background noise, it can add additional distortion to the coded speech signal when background noise levels are high.

III. Software Platforms for Simulating Mobile Phone Speech

Two platforms have been developed at the University of Auckland based upon this alternative approach for simulating mobile phone speech. They make use of publicly available software routines of the AMR and EVRC codecs

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[16], [17]. The layouts of the GSM and CDMA codec platforms are shown in Fig. 2 and Fig. 3, respectively. These platforms incorporate functionality which can be used to operate the mobile speech codec under a wide range of possible scenarios. They also have the capability to process, view or listen to multiple or single speech files (.wav format). Using these platforms, any existing speech database can be transformed into mobile quality speech by adjusting the platform settings in accordance with the desired scenario. The input speech samples must be sampled at 8 kHz and 16 bit digitized.

A. *Simulating the Impact of Dynamic Rate Coding (DRC)*

The GSM platform has the capability to create all the possible active codec sets (ACSs) which conform to a certain channel condition. Based upon the standards and constraints mentioned in Section 2, the ACSs will be generated automatically upon specifying a speech quality: low, medium and high (see Fig. 2). Each ACS generated is a combination of 1, 2, 3 or 4 bit rates. The bit rates used and the total number of possible ACSs for each speech quality are listed below:

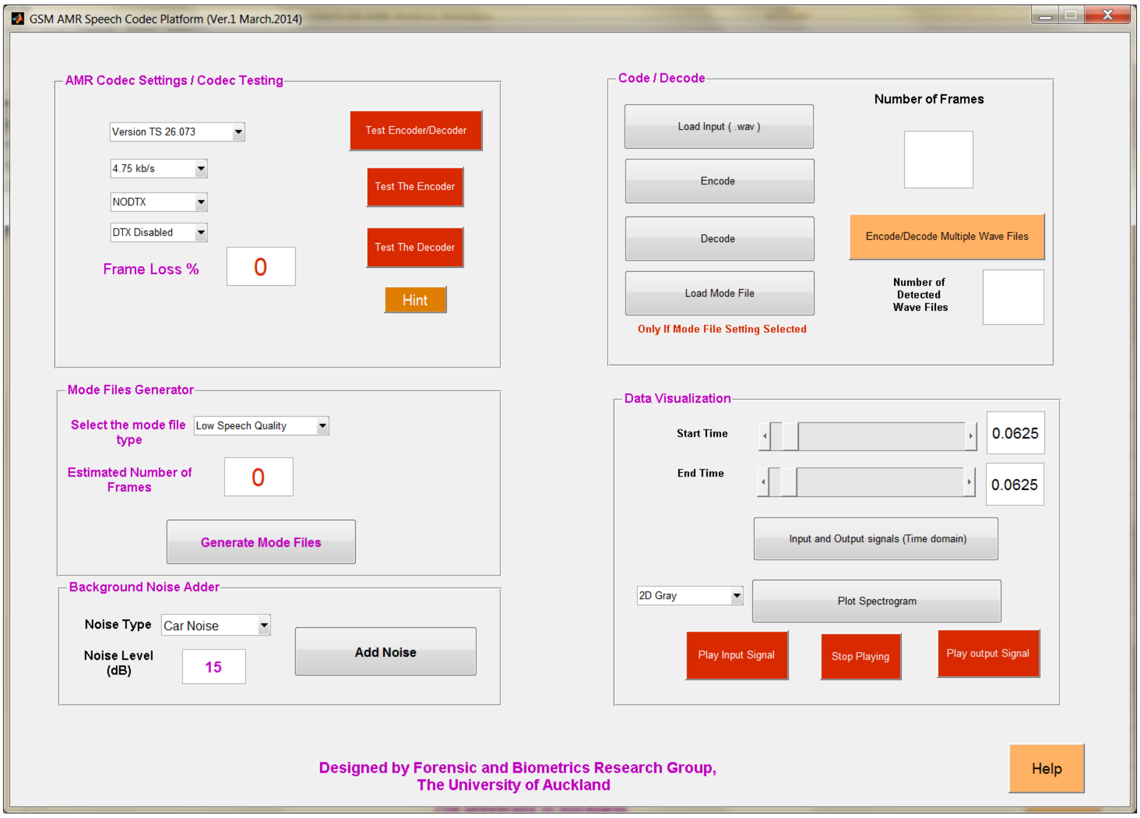

Fig 2. The AMR codec platform layout and settings.

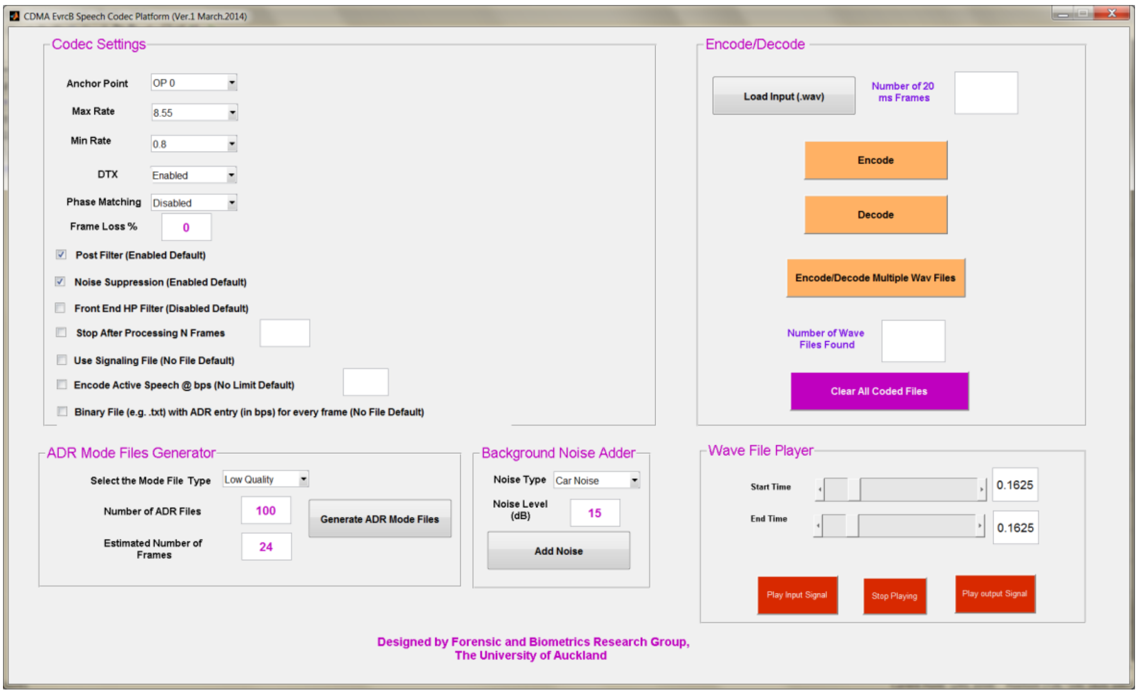

Fig 3. The EVRC codec platform layout and settings.

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

--Low speech quality: (4.75, 5.15, 5.90, 6.70, 7.40) kbps. This selection results in 30 ACSs.

--Medium speech quality: (4.75, 5.15, 5.90, 6.70, 7.40, 7.95, 10.20, 12.20) kbps and results in 162 ACSs.

--High speech quality: (7.40, 7.95, 10.20, 12.20) kbps, which results in 15 ACSs.

The ACSs will be selected randomly, according to a uniform distribution, from the sets listed above. The choice of a uniform distribution is arbitrary at this stage, and there is ongoing research into what a more appropriate choice might be.

The ICM constraint has also been taken into account, where this will be the second bit rate in the ACS if it contains a combination of four bit rates; else it is the lowest bit rate in the set.

The CDMA platform also provides practitioners with a facility to create ADR files in accordance with different qualities of speech (or capacity conditions) as shown in Fig. 3. The ADR values are automatically generated upon specifying the speech quality desired. For each speech quality, the ADR values are sampled at different points of the continuous range according to the following: (i) lower half (low speech quality), (ii) entire range (medium speech quality) and (iii) upper half (high speech quality). The ADR values chosen for each of the ranges are as follows:

--Low speech quality: selected from the lower-half ADR range (4.8, 5.8, 6.2, 6.6) kbps.

--Medium speech quality: selected from the entire ADR range (4.8, 5.8, 6.2, 6.6, 7.0, 7.5, 8.5, 9.6) kbps.

--High speech quality: selected from the upper-half ADR range (7.0, 7.5, 8.5, 9.6) kbps.

The selection of discrete values ensures a better coverage of these ranges and thus enforces switching between anchor points in order to reflect more realistic scenarios in the network. We know for instance that ADR values for medium speech quality can take on any value between 4.8 and 9.6 kbps [7], but random selection from such a continuous range can result in these being more concentrated at one part of the range than another. Using discrete values increases the likelihood of switching between AOPs and ensures a better coverage of the range.

C. Simulating the Impact of Frame Loss (FL)

We have had to modify the original software code of both the AMR and EVRC codecs in order to allow the user to specify a percentage of FL. The lost frames are selected randomly according to a uniform distribution and the total number of lost frames is made equal to the FL percentage specified, multiplied by the number of 20 ms speech frames found in the target speech file. Again, the choice of a uniform random distribution is arbitrary at this stage and we are currently investigating whether there is a more appropriate distribution that should be used. In terms of the modifications we have had to make to the original software codec simulations in order to implement this feature, we have done this by noting that the frame loss mechanism is internally activated in the AMR codec upon changing the "RX_TYPE" flag value to "SPEECH_BAD" [10]. Similarly, this process is initiated in the EVRC codec by changing the "data_packet.PACKET_RATE" flag value to "0xE" [7]. The network operators will drop a call if the FL exceeds 10-15%, so we have constrained accordingly the possible values for FL in our platforms.

One of the interesting aspects which can be easily investigated using these platforms is the impact of the temporal locations of lost frames on the speech signal. Examples of this are shown in the spectrograms of Fig. 4 and Fig. 5 for the AMR and EVRC codecs, respectively. These relate to a nine-frame segment of the diphthong /aI/. Fig. 4(a) and Fig. 5(a) relate to the original speech segment. Fig. 4(b) and Fig. 5(b) relate to the recovered speech signal with no frame loss. Here the speech has been coded at 12.2 kbps and OP0, respectively. Fig. 4(c) and Fig. 5(c) show the impact of a single lost frame, namely Frame 3, on the recovered speech signal, while Fig. 4(d) and Fig. 5(d) show what happens when the single lost frame is Frame 5 instead. It is clear from these figures that a single lost frame can negatively impact on the recovered frames that follow it. Further, the extent of this impact varies depending on the temporal location of the lost frame. For both codecs the loss of the 5th frame (around the transition region of the diphthong) has had a greater impact on the recovered speech waveform than loss of the 3rd frame. It is also interesting to note, by comparing Fig. 4 and Fig. 5, that there are noticeable differences between the recovered speech signals for each of the codecs for the same input speech segment, even when there has been no frame loss.

D. Simulating the Impact of Background Noise (BN)

Both platforms also provide facility to add different kinds of BN at any specified SNR prior to processing the speech files. The most common types of noise in mobile phone communications are babble, car and street noises, and these noise types have therefore been included in our platforms. The background noise files used have been acquired from the soundjay database [18]. However, if the user prefers to use a different set of background noise files, they can be simply replaced. It is known that in mobile phone communications typical SNR levels at the transmitting end vary from 9 to 20 dB [19].

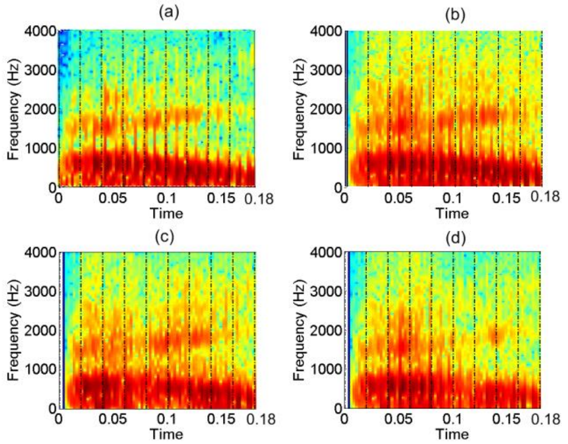

Fig 4. Spectrograms showing the impact of frame loss with the AMR codec on the diphthong /aI/. (a) uncoded speech, (b) recovered speech with no frame loss, (c) recovered speech when Frame 3 is lost, (d) recovered speech when Frame 5 is lost. (Note: dashed lines show the frame boundaries).

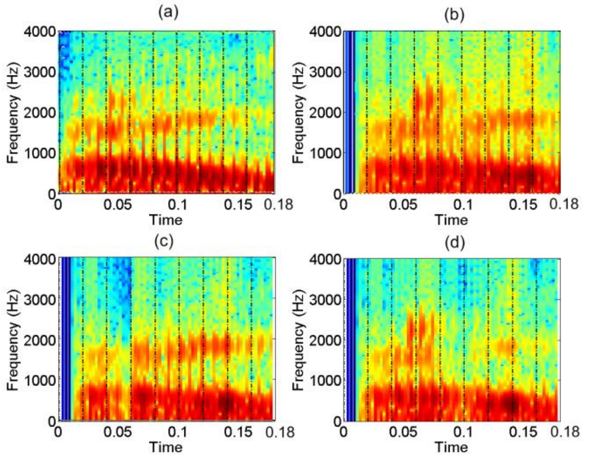

Fig 5. Spectrograms showing the impact of frame loss with the EVRC codec on the diphthong /aI/. (a) uncoded speech, (b) recovered speech with no frame loss, (c) recovered speech when Frame 3 is lost, (d) recovered speech when Frame 5 is lost. (Note: dashed lines show the frame boundaries).

Though, the process of NS is enabled by default in the EVRC codec, our CDMA platform allows the user to enable or disable this feature. This can be quite useful in order to understand whether the distortion of the speech signal has been caused by the coding process itself, or is as a result of NS.

To illustrate differences between both codecs in respect to handling of BN, a set of time waveforms and their corresponding spectrograms have been produced for the diphthong /aI/. These are shown in Fig. 6 and Fig. 7 for the AMR and EVRC codecs, respectively. Fig. 6(a) shows the time waveform of the original speech segment. Fig. 6(b) shows the same speech segment with babble noise added at SNR = 9 dB. The noisy speech has then been coded at 12.2 kbps in the case of the AMR codec and OP0 for the EVRC codec. Their resulting recovered speech waveforms are shown in Fig. 6(c) and Fig. 6(d) for the AMR and EVRC codecs, respectively. Examination of the resulting spectrograms of Fig. 7 shows that the EVRC codec has done a marginally better job of reducing the BN, as evidenced by a reduction in the energy of the high frequency noise components.

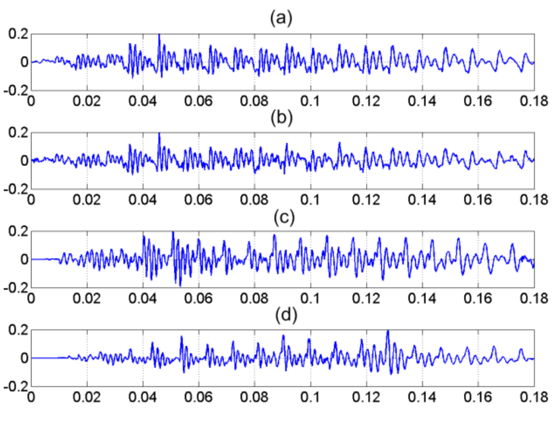

Fig 6. Time waveforms showing the impact of BN on the coding of the diphthong /aI/. (a) uncoded speech, (b) uncoded speech with added babble noise at SNR = 9 dB, (c) AMR recovered speech when BN is present, (d) EVRC recovered speech when BN is present. (Dashed lines show the frame boundaries).

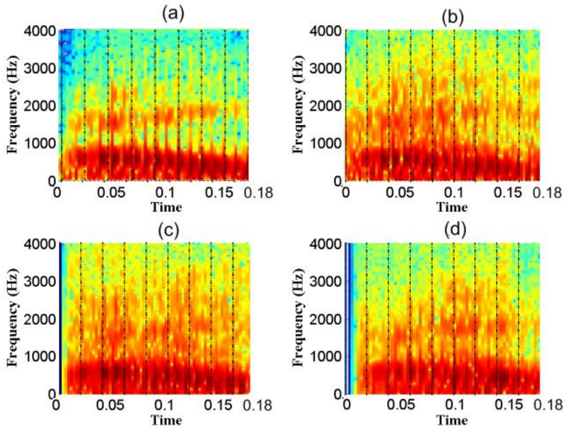

Fig 7. Spectrograms showing the impact of BN on the diphthong /aI/. (a) uncoded speech, (b) uncoded speech with added babble noise at SNR = 9 dB, (c) AMR recovered speech when BN is present, (d) EVRC recovered speech when BN is present. (Dashed lines show the frame boundaries).

However, the characteristics of the speech signal have also slightly changed as a result of this noise subtraction process, which will likely impact upon the measurement of any speech parameters of interest.

IV. Conclusions

This paper has presented an alternative approach to investigating the impact of mobile phone technology on speech. It makes use of the speech codec software routines of these networks to simulate the impact of various aspects of the mobile phone network. The rationale behind this approach is that it is the codec in these networks which alone directly impacts the speech signal. We have focused on the GSM and CDMA mobile phone networks and noted that they incorporate somewhat different mechanisms in respect to how they handle the speech signal.

The three key aspects in mobile phone communications that can negatively impact the speech signal are dynamic rate coding, frame loss and background noise. A thorough understanding of these can be used to drive the speech codecs in a manner similar to an actual mobile phone network. Two software platforms have been developed to aid researchers in examining and understanding the impact of each of these on the speech signal, either separately, or in various combinations. The integration of all three aspects is essential in order to accurately reflect realistic scenarios in these networks.

(Note: These platforms will be made available online soon, but can also be obtained by contacting the authors.)

References

[1] H. Lu, A. B. Brush, B. Priyantha, A. K. Karlson, and J. Liu, "SpeakerSense: energy efficient unobtrusive speaker identification on mobile phones," In Pervasive Computing, Springer Berlin Heidelberg, pp. 188-205, 2011.

[2] L. Besacier, S. Grassi, A. Dufaux, M. Ansorge, and F. Pellandini, "GSM speech coding and speaker recognition," in proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP' 00), 2000, vol. 2, pp. II1085-II1088.

[3] GSM, Available: http://www.gsmworld.com/.

[4] CDG, Available: http://www.cdg.org/index.asp.

[5] B. J. Guillemin, and C. Watson. "Impact of the GSM Mobile Phone Network on the Speech Signal–Some Preliminary Findings," International Journal of Speech Language and the Law, vol. 15, no. 2, 2008, pp. 193-218.

[6] 3GPP, "TS 26.071 V11.0 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects;

Mandatory speech CODEC speech processing functions; AMR speech CODEC; General description,". Available: http://www.3gpp.org/.

[7] 3GPP2, "S0018-D, S0018-D, Minimum Performance Specification for the Enhanced Variable Rate Codec, Speech Service Options 3, 68, 70, and 73 for Wideband Spread Spectrum Digital Systems," Available: http://www.3gpp2.org/.

[8] 3GPP, "TS 45.009 3rd Generation Partnership Project; Technical Specification Group GSM/EDGE Radio Access Network; Link adaptation," Available: http://www.3gpp.org/.

[9] Nokia, "Guidelines for practical implementation of AMR in Nokia's Network element. General description," Available: http://www.scribd.com/doc-/104330368/14/Initial-codec-mode-selection.

[10] 3GPP, "TS 26.091 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Error concealment of lost frames," Available: http://www.3gpp.org/.

[11] 3GPP, "TS 45.003; Technical Specification Group GSM/EDGE; Radio Access Network; Channel coding," Available: http://www-.3gpp.org/.

[12] E. A. Alzqhoul, B. B. T. Nair, and B. J. Guillemin, "Speech Handling Mechanisms of Mobile Phone Networks and Their Potential Impact on Forensic Voice Analysis," Speech Science and Technology (SST), Sydney, Australia, 2012.

[13] D. Networks, "Voice Quality Solutions for Wireless Networks," Available: http://www.ditechnetworks.com/.

[14] S. H. Isabelle, "Noise suppression for low bitrate speech coder," Google Patents, U.S. Patent No. 6,122,610, issued September 19, 2000.

[15] S.-m Lee, and Y.-J. Kim, "Method for eliminating annoying noises of enhanced variable rate codec (EVRC) during error packet processing," Google Patents, U.S. Patent No. 6,385,578, issued May 7, 2002.

[16] 3GPP, "TS 26.073 ANSI-C code for the Adaptive Multi Rate (AMR) speech codec," Available: http://www.3gpp.org/.

[17] 3GPP, "Software Distribution for Enhanced Variable Rate Codec (EVRC), Speech Service Options 3 and 68, Minimum Performance Specification (MPS)," Available: http://www.3gpp2.org/.

[18] Soundjay, "Sound Effects website," Available: http://www.soundjay.com/.

[19] 3GPP, "TS 26.077, Minimum performance requirements for noise suppresser; Application to the Adaptive Multi-Rate (AMR) speech encoder," Available: http://www.3gpp.org/.

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)