



HAL
open science

Joint effect of channel coding and AMR compression on speech quality

Minh Quang Nguyen, Hang Nguyen

► To cite this version:

Minh Quang Nguyen, Hang Nguyen. Joint effect of channel coding and AMR compression on speech quality. WICOM 2015: 11th International Conference on Wireless Communications, Networking and Mobile Computing, Sep 2015, Shanghai, China. pp.291 - 295. hal-01272054

HAL Id: hal-01272054

<https://hal.science/hal-01272054v1>

Submitted on 10 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Joint effect of channel coding and AMR compression on speech quality

Minh-Quang Nguyen and Hang Nguyen
Department of Wireless Network and Multimedia Services
Institut Mines-Telecom - Telecom SudParis
Samovar Laboratory, UMR 5157, CNRS, Evry, France
{minh_quang.nguyen, hang.nguyen}@telecom-sudparis.eu

Abstract—In this paper, we estimate the effect of joint Adaptive Multirate (AMR) compression and channel coding on quality of speech data. In our simulation, speech data is transferred over AWGN channel after performing 8 AMR codec modes (corresponding to 8 different compression rates) and 12 convolution code rates. The joint effect is estimated in the receiver by analysing quality of received speech data by Mean Opinion Score. Our result shows that conventional scheme performing source coding and channel coding separately and sequentially does not guarantee obtaining the optimal solutions for speech data.

Keywords—AMR, source-channel coding theorem, convolution code, Shannon Theory, compression.

I. INTRODUCTION

Speech data must be compressed (source coding) and encoded by redundant way (channel coding) before transmission in order to transfer over erasure channel. Conventionally, these two steps are performed separately. During compression, we choose the most effective representation of data to remove all the redundancy and form the most compressed version possible. Conversely, during data transmission, we add redundant information to message prior to data transmission, and receiver can recover original data which contain no apparent errors. There is a need to evaluate the joint effect of source coding and channel coding on quality of data to know whether traditional method is optimal or not.

In Shannon's 1948 paper [1], the relationship between source/channel coding is first mentioned in joint source-channel coding (JSCC) separation theorem which is combination of two main branches of Shannon theory. It stated for discrete memoryless source over discrete memoryless channel and contained two parts: the direct part stated that the source can be transmitted over channel in reliable way if source coding rate is strictly below the channel capacity; and converse part stated that if the source coding rate is smaller than channel capacity, reliable transmission is impossible.

According to two parts of Shannon theory, to attain a good quality of data transmission, a conventional scheme is to do two works separately and sequentially: (1) compress data with highest rate which satisfies rate-distortion threshold and (2) protect compressed data by adding redundancy in channel encoder.

The theorem is proven in different classes of sources and channels. In general case, Da Wang et al. [2] proved a strong

converse to JSCC: for any joint source-channel scheme, the success probability approaches zero as the block length increases if source is compressed with distortion below threshold which is lower than the optimal distortion. However, Shannon source-channel coding separation theorem is only proven in case of large data blocks length. On the other hand, works of Verdú et al. [3] showed examples where the converse to the separation theorem fails to hold: a memoryless information stable source/channel pair in which minimum achievable source coding is strictly greater than channel capacity, but the source is transmissible through the channel with zero error. They claimed that, in general, the channel capacity and the minimum source coding rate are insufficient to determine whether the source can be transmitted reliably or not.

For speech data, dependence of speech quality on compression method or FEC was widely explored [4]–[8]. Different adaptation algorithms were proposed to improve speech quality in wireless IP networks and wireless Ad Hoc networks. Johnny Matta et al. [8] presented a dynamic joint source and channel coding adaptation algorithm for the AMR speech codec based on the ITU-T Emodel. They found the optimal choice of source and channel bit rate given QoS information about the wired and wireless IP network. Yicheng Huang et al. [4] proposed an algorithm to optimize speech quality with AMR along with a FEC scheme. The effect of Package Loss Rate and compression rate on speech quality is described. Work of Hongqi Zhang et al. [6] shown the performance of speech transmission in wireless Ad Hoc networks. They also proposed a mechanism to control source network rate to decrease packet loss. The main idea is to choose an appropriate AMR codec mode with respect to the available network bandwidth, to maximize speech quality.

Although results have been published for good strategy to speech data transmission, no results are available for the joint effect of both source coding and channel coding speech data's quality. The impact of source/channel sequence on speech data's quality needs to be evaluated to determine whether the traditional strategy can provide optimal solution for data transmission.

This paper will investigate the dependence of speech's quality on sequence of AMR compressor/convolution code. Our simulation is performed in AWGN channel with 8 AMR code rates and 12 convolution code rate. Results show how pairs of compression rate and channel rate affect on quality of speech. Moreover, we also give an example in which

conventional strategy does not attain the best quality of speech data.

II. FROM JSCC THEOREM TO SPEECH DATA TRANSMISSION

In [9], the two-stage scheme is proven to be as efficient as any other method which could be designed for transmitting information over a noisy channel. This result has significant implication in practical communication systems. A communication system can be considered as two separate parts: source coding and channel coding. Source codes can be chosen for the most efficient representation of data, i.e. representation with the least redundancy. Independently, channel coding can be designed to be appropriate for the channel condition. The combination of two designed parts would be as effective as any other system which is designed by considering source coding and channel coding together.

However, the process of two separate steps is not always effective. There are examples of multimedia data transmission which is perceived by human senses. A simple example is sending English text over an erasure channel. The most efficient binary representation of the text can be used and sent over the erasure channel. The error appearing in received messages makes it difficult to decode. On the other hand, when English text is sent directly over the channel, human is able to make sense out of the message although a significant part of message can be missed. Similarly, some special characteristics of ears allow human to recognize speech under very high distortion. In such cases, it may be appropriate to send the uncompressed speech over the noisy channel rather than the compressed representation. The redundancy in the source apparently must be suited to the channel.

For speech data, there is a wide range of source coding and channel coding methods for transmission. Modern compression methods can be divided into 2 classes: wideband speech coding (AMR-WB for WCDMA networks; VMR-WB for CDMA2000 networks; G.722, G.722.1, Speex, IP-MR for VoIP and video conferencing...) and narrowband speech coding (FNBDT for military applications; SMV for CDMA networks; full rate, half rate, EFR and AMR for GSM networks; G.723.1, G.726, G.728, G.729, iLBC for VoIP or video conferencing...). To control errors in transmission over noisy channel, forward error correction (FEC) is widely used. Examples of FEC include: convolution code [10]–[12], Turbo code [13], [14], low-density parity-check [15], [16]. Because of page limitation, this paper will investigate the effect of AMR and convolution code. Other pairs of source/channel coding for speech data are considered as future work.

III. SIMULATION CHAIN DESCRIPTION

Our simulation is described in Figure 1. The original speech data is transferred to receiver through six steps:

- 1) Compression: remove redundancy from original raw speech data to form compressed versions;
- 2) Channel encoding: add redundancy to compressed data to control errors when transferring over noisy channel;
- 3) BPSK modulation: convey data by modulating the phase of a reference signal (the carrier wave);

TABLE I: Compression rates corresponding to AMR codec modes

AMRmode	4.75	5.15	5.9	6.7	7.4	7.95	10.2	12.2
Compr.rate	24.60	22.86	19.98	17.78	15.99	15.23	11.85	10.00

- 4) Transmission: send modulated data to receiver over AWGN channel;
- 5) Channel decoding: translate received messages into AMR format;
- 6) Source decoding: decompress data to form raw speech data.

After six steps, speech data with errors is obtained at receiver. Quality estimation on speech is performed to compare received speech data with original data.

IV. SIMULATION PARAMETERS

In our simulation, the compressor for speech data is 3GPP Adaptive Multi-Rate Floating-point Speech Codec (3GPP TS 26.104 V11.0.0) [17]. After compressing each raw speech data file, 8 AMR files are obtained. They correspond to 8 codec modes: 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2 and 12.2 KB/s. Let R^{Comp} be the compression rate. It is calculated by formula:

$$R^{Comp} = \frac{Uncompressed\ Size}{Compressed\ Size}$$

Compression rates of each codec mode is shown in Table 1. Higher bit rate corresponds to lower compression rate and associates with higher speech quality.

Convolution coding is used in the second step. Convolution code rate is defined as the ratio of input data size to output data size in convolution encoder. By using different puncturing matrices, each AMR file is protected with 12 different convolution code rates: 5/6, 4/5, 3/4, 2/3, 4/7, 1/2, 4/9, 2/5, 4/11, 1/3, 1/4 and 1/5. Note that the convolution code rates are inversely proportional to redundancy added to input message in convolution encoder.

The channel in our simulation is AWGN, the energy per bit to noise power spectral density ratio is 0.63. With each raw speech file, 96 files (corresponding to 8 compression rates and 12 convolution code rates) is transferred over channel.

At receiver, Viterbi algorithm [18], [19] is used in channel decoder. AMR files after channel decoder are then converted to raw speech files by AMR decoder. We obtain 96 raw speech files from a single original speech file at transmitter.

Quality of each received file (corresponding to one codec mode and one convolution code rate) is estimated by Mean Opinion Score (MOS) which rates a processed speech signal in relation to the original signal. It provides a numerical indication of the perceived quality from the user's perspective of received media after compression and transmission. The MOS is expressed as a value in the range -0.5 to 4.5, where -0.5 is lowest perceived audio quality, and 4.5 is the highest perceived audio quality measurement. Original raw speech data includes 100 files, which are in PCM format. Each file is 108560 milliseconds length, with 5428 frames. Hence, 542.800 frames are transferred.

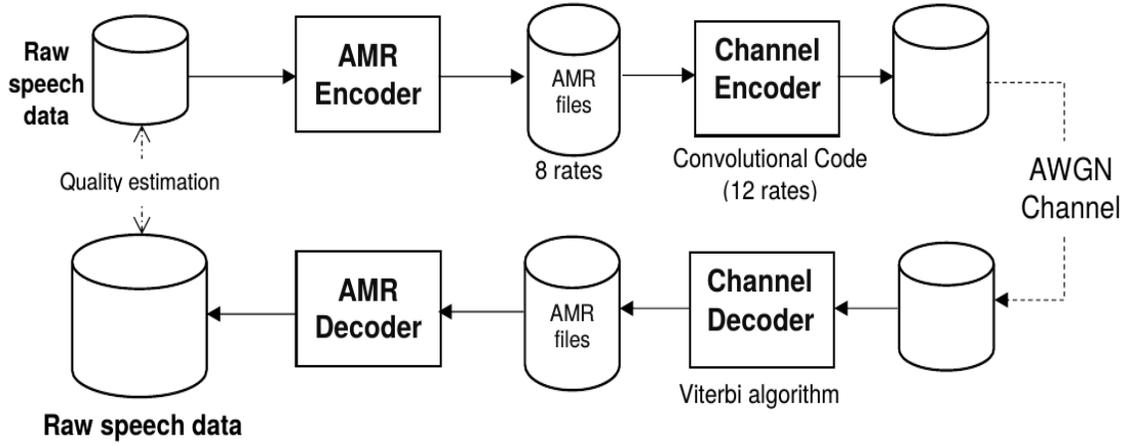


Fig. 1: Simulation chain

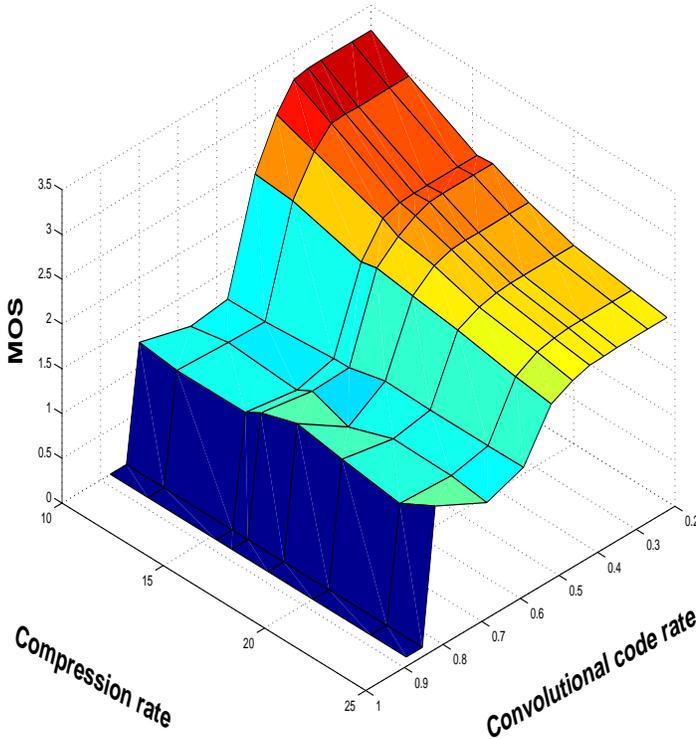


Fig. 2: Dependence of MOS on compression rate and convolution code rate

V. THE RESULT

The dependence of MOS on AMR compressor/convolution code is shown in Figure 2. Each point in 3D-curve corresponds to speech data quality attained after transferring over channel by specific pair of a compression rate and a convolution code rate.

With a specific range of quality, there are several corresponding pairs of compression rates and channel rate. Let $R_U(\text{bits/s})$ and R^{Conv} be bit rate of original speech data and convolution code rate, respectively. Rate of encoded data

which is transferred over channel is:

$$R_e = \frac{R_U}{R^{Comp} * R^{Conv}} (\text{bits/s})$$

Channel must transfer encoded data at R_e to ensure fluency of speech at the receiver. Figure 3, 4 and 5 show all the pairs of compression rates and convolution code rate for speech transmission to attain MOS with values ranging from 2.0 to 2.1, 2.1 to 2.2 and 2.9 to 3.0, respectively. R_e corresponds to each pair is given.

In addition, we shown the best point that obtain specific MOS with respect to number of bits transfer over channel in one second. These points, with their rate of encoded data is marked as triangle point in Figure 3, 4 and 5.

VI. RESULT ANALYSIS

As shown in Figure 2, overall, the general trend of MOS is high in data transferred by high compression rate and low convolution code rate, and is low in data transferred by low compression rate and high convolution code rate.

The two highest convolution code rates (5/6 and 4/5, corresponding to smallest amount of added redundancy to control errors over channel), AMR decoder failed to decode received data. Hence, MOS in this case is expressed as zero.

With three lower convolution code rates (3/4, 2/3 and 4/7), MOS followed a fairly similar pattern over all 8 compression rates, all remaining at between 1.2 and 1.6. But in 7 lowest convolution code rates (i.e. more redundant bits added), quality of speech is gradually increased with the deduction of compression rate. The quality reaches to peak at compression rate 24.60 and convolution code rate 0.2, which are the highest compression rate and lowest convolution code rate in our simulation.

Figure 3, 4 and 5 show that, if low codec modes are used (high compression rate which corresponds to large amount of redundancy removed) then more redundancy have to be added to compressed file to obtain specific MOS. In contrast, if higher codec modes are used, number of redundant bits is less.

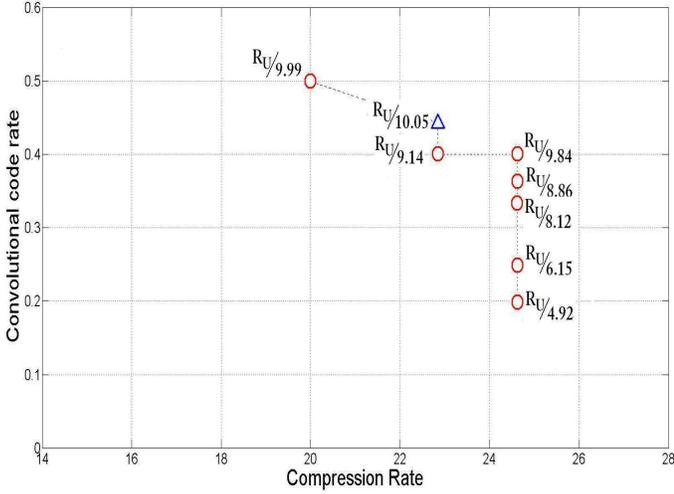


Fig. 3: Convolution code rates and compression rates for MOS between 2.0 and 2.1

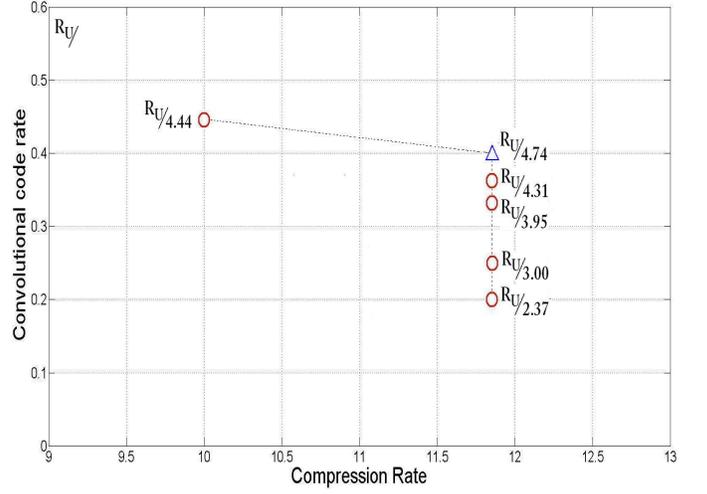


Fig. 5: Convolution code rates and compression rates for MOS between 2.9 and 3.0

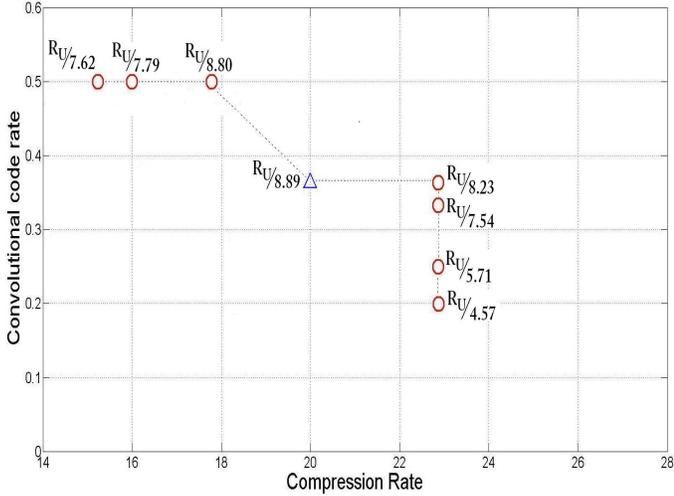


Fig. 4: Convolution code rates and compression rates for MOS between 2.1 and 2.2

The interesting result found in Figure 3 and 4 is that the best encoder-decoder sequence does not correspond to the highest compression rates of source encoder.

It is shown that the highest compression rate to reach MOS 2.0 - 2.1 is 24.60 while the optimal choice of source and channel coding corresponds to compression rate 22.85. Similarly, with MOS from 2.1 - 2.2, the optimal compression rate is 19.98 whether compression rate 24.60 (with different convolution code rates) can be used to reach this MOS.

In the view of these two cases, it is clear that performing separately AMR compression (source coding) and convolution code (channel coding) may not find optimal solution for reliable transmission of a source over channels. An example can be brought out from Figure 3. Suppose that the channel

can transfer data at rate

$$\mathcal{R}_{channel} = \frac{R_U}{9.9} (\text{bits/s})$$

During compression, in order to ensure MOS greater than 2.0, conventional strategy is to remove redundancy in the data to form most compressed version which is strictly smaller than channel capacity and AMR mode 4.75 (with compression rate $R_1^{comp} = 24.60$) can be used. To combat errors in channel, the biggest convolution code rate can be used is shown in Figure 3: $R_1^{conv} = 2/5$. Hence, data must be transferred at rate:

$$R_1 \geq \frac{R_U}{R_1^{conv} * R_1^{comp}} = \frac{R_U}{9.84} > \mathcal{R}_{channel}$$

Although data is compressed as much as possible with AMR mode 4.75, adding redundancy to guarantee MOS greater than 2.0 leads to data rate greater than $\mathcal{R}_{channel}$. Interestingly, AMR mode 5.15 (corresponding to compression rate 22.86) can be used. Namely, compressing speech data with AMR mode 5.15 ($R_2^{comp} = 22.86$) and add redundancy with convolution code rate $R_2^{conv} = 2/5$ leads to data rate

$$R_2 = \frac{R_U}{10.05} < \mathcal{R}_{channel}$$

Apparently, it can be reliably transferred over channel to get MOS greater than 2.0. We also can bring an similar example from Figure 4.

Our simulation found that, the conventional scheme in which source coding and channel coding are chosen separately based on channel conditions, does not guarantee the best speech quality.

VII. CONCLUSION

We simulate speech data transmission with AMR compression and convolution code. We estimate speech quality by MOS and describe the dependence of speech quality on source and channel coding. From simulation result, joint effect of channel coding and AMR compression on speech quality

is described. We find an example that separate strategy in transmission does not guarantee optimality in case of speech codec with AMR and convolution code. We suggest that beside channel capacity and the minimum source coding rate, more information is needed for the best performance of speech data transmission.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communications," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.
- [2] D. Wang, A. Ingber, and Y. Kochman, "A strong converse for joint source-channel coding," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 2117–2121.
- [3] S. Vembu, S. Verdu, and Y. Steinberg, "The source-channel separation theorem revisited," *Information Theory, IEEE Transactions on*, vol. 41, no. 1, pp. 44–54, 1995.
- [4] Y. Huang, J. Korhonen, and Y. Wang, "Optimization of source and channel coding for voice over ip," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 4–pp.
- [5] M. Karlsson, M. Almgren, S. Bruhn, K. Larsson, and M. Sundelin, "Joint capacity and quality evaluation for amr telephony speech in wcdma systems," in *Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall. 2002 IEEE 56th*, vol. 4. IEEE, 2002, pp. 2046–2050.
- [6] H. Zhang, J. Zhao, and O. Yang, "Adaptive rate control for voip in wireless ad hoc networks," in *Communications, 2008. ICC'08. IEEE International Conference on*. IEEE, 2008, pp. 3166–3170.
- [7] O. Corbun, M. Almgren, and K. Svanbro, "Capacity and speech quality aspects using adaptive multi-rate (amr)," in *Personal, Indoor and Mobile Radio Communications, 1998. The Ninth IEEE International Symposium on*, vol. 3. IEEE, 1998, pp. 1535–1539.
- [8] J. Matta, C. Pépin, K. Lashkari, and R. Jain, "A source and channel rate adaptation algorithm for amr in voip using the emodel," in *Proceedings of the 13th international workshop on Network and operating systems support for digital audio and video*. ACM, 2003, pp. 92–99.
- [9] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.
- [10] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *Information Theory, IEEE Transactions on*, vol. 13, no. 2, pp. 260–269, 1967.
- [11] Viterbi and A. J., "An intuitive justification and a simplified implementation of the map decoder for convolutional codes," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 2, pp. 260–264, 1998.
- [12] B. Sklar, *Digital communications*. Prentice Hall NJ, 2001, vol. 2.
- [13] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *Communications, IEEE Transactions on*, vol. 44, no. 10, pp. 1261–1271, 1996.
- [14] S. Benedetto and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *Information Theory, IEEE Transactions on*, vol. 42, no. 2, pp. 409–428, 1996.
- [15] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 619–637, 2001.
- [16] T. Richardson, "Error floors of ldpc codes," in *Proceedings of the annual Allerton conference on communication control and computing*, vol. 41, no. 3. The University; 1998, 2003, pp. 1426–1435.
- [17] "3gpp specification detail: Ansi-c code for the floating-point adaptive multi-rate (amr) speech codec," sep 2014. [Online]. Available: <http://www.3gpp.org/DynaReport/26104.htm>
- [18] A. J. Viterbi, "A personal history of the viterbi algorithm," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 120–142, 2006.
- [19] G. D. Forney Jr, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.