A Note on Source Models for Speech Traffic

Michael Menth[1], Andreas Binzenhöfer[2], Stefan Mühleck[3]

Abstract

Speech traffic is often used in simulations to evaluate the performance of control mechanisms in communication networks. Therefore, trustworthy models are required that capture the fundamental statistical properties of typical voice sources. The G.723.1 codec produces on/off traffic streams with fixed size packets. The iSAC codec strongly periodic packet streams with variable packet sizes. We propose new models for the traffic output of both codecs and show that their queuing properties are in good accordance with those of original traffic traces, while existing traffic models, that are frequently used in literature, lead to significant discrepancies.

1 Introduction

Due to the high redundancy in human speech, voice data can be well compressed, but different voice codecs exploit this fact to a different degree. The G.711 and the G.729.1 codec simply encode speech into packets of fixed size. The G.723.1 and the iLBC codec detect silence phases during which they suppress the generation of data leading to an on/off process on packet level. Finally, the GSM AMR and the iSAC codec take additional advantage of the characteristics of speech and compress it into packets of different size leading to variable bit rate (VBR) streams.

Source models for speech traffic seem to be an old and well-studied topic. However, a look into the literature shows that a large number of simulative or analytical studies and simulation tools use an on/off model with exponentially distributed on/off phases with a duration of 352 and 650 ms, respectively. Most of them refer to [10] which cites "private work" [7]. We tried to track it, but without success. The work of Brady [2–4] seems to be the next popular source which reports mean durations for on/off phases of about 1.3 s and 1.7 s, respectively. These source models were accurate in the 60ies and 70ies, but our studies of recent voice codecs clearly show that they are outdated and do no longer capture the characteristics of packetized voice traces on packet level.

We apply each of the vocoders mentioned above to a large set of typical telephone conversations (3.5 h = 7 h speech) from [1], a publicly available database of English speech sources which were specifically designed to be used in research and speech technology. We then analyze the original packet traces and provide quantitative models describing the codec output. To validate the accordance of the stochastic models and the original traces, we compare the cumulative distribution function (CDF) of the packet sizes, the complementary CDF (CCDF) of the on/off phase durations, the autocorrelation function (ACF) of consecutive packet sizes, and the CCDF of the packet waiting time when several voice streams are fed to a single server queue.

In this note, we review the most important measurement results and appropriate traffic models of our study in [8]. Section 2 covers G.723.1 as a representative for on/off voice codecs and Section 3 covers iSAC as a representative for VBR codecs. Finally, we give a short conlcusion in Section 4.

2 On/Off Voice Codecs

The G.723.1 codec [6] is an ITU-T standard since 1995. It was specially designed for voice encoding at low bandwidth and is mostly used in VoIP applications, e.g., in Netmeeting or Picophone. G.723.1 can operate in two different modes generating up to 6.4 kbit/s with 24 bytes chunks or up to 5.3 kbit/s with 20 bytes chunks every 30 ms. Voice activity detection is used to suppress the generation of packets during silence phases which leads to an on/off output. We generate packet traces using the Picophone software [11] which relies on the reference implementations of Microsoft. G.723.1 produces a main audio stream of packets with fixed size and sends additional control information of 1 byte every 3 s. Figure 1 shows a typical time series of consecutive packet sizes. Talkspurts consist of several contiguous talk phases that are interrupted by short breaks. In contrast, the silence intervals are mostly contiguous, but they are sometimes also interrupted by very short noise. Hence, measuring only contiguous talk and silence phases (**W0**) underestimates the length of the talkspurts and, thereby, the autocorrelation function of consecutive packet sizes.

We solved this problem by assuming that real talkspurts start with $w\uparrow$ generated packets and real silence intervals with $w\downarrow$ suppressed packets, respectively (**W2**). Otherwise, we interpret generated or suppressed packets as noise or breaks. It is clear that the average duration of on/off phases increases with increasing $w\uparrow$ and $w\downarrow$. Figure 2 shows the impact of both parameters on the mean duration $E[D_{on}^{real}]$ of the on-phases for the G.723.1 codec. The parameter combination $w\uparrow = 15$ and $w\downarrow = 50$ leads to stable results and we get $E[D_{on}^{real}] = 11.54$ s and $E[D_{off}^{real}] = 11.98$ s.

[1]University of Würzburg, Institute of Computer Science, Germany
[2]University of Würzburg, Institute of Computer Science, Germany
[3]University of Würzburg, Institute of Computer Science, Germany

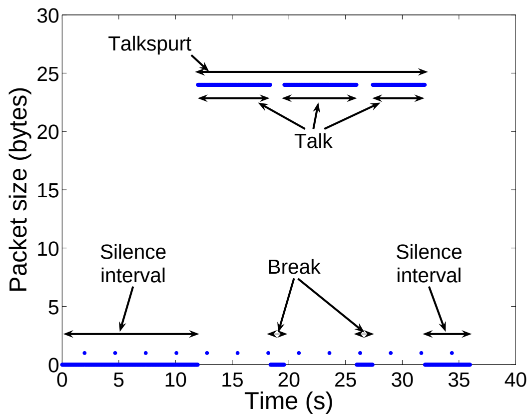

Figure 1: G.723.1 outputs an audio stream and control information. The audio stream consists of silence intervals and main talkspurts that are interrupted by short breaks and noise.

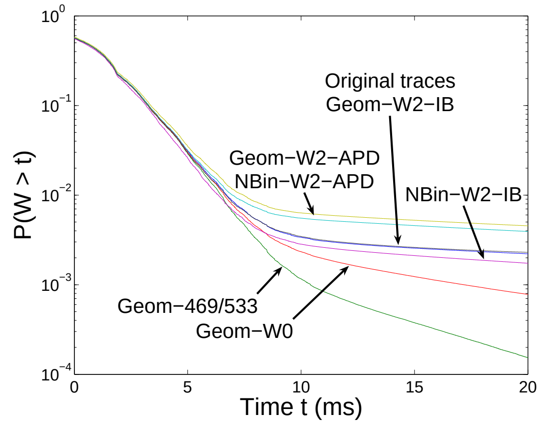

Figure 3: Validation of different traffic models for the G.723.1 codec with original traces using the waiting time of multiplexed sources.

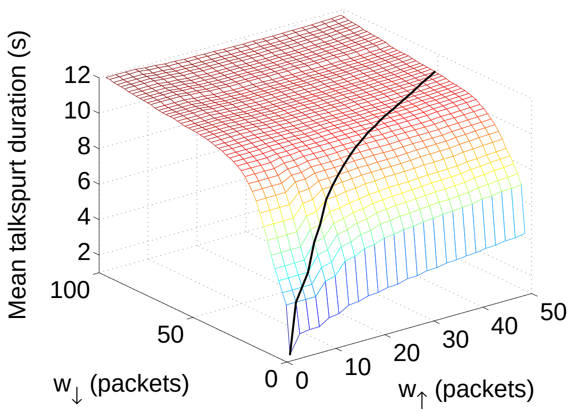

Figure 2: Impact of the window parameters $w\downarrow$ and $w\uparrow$ on the measured mean duration $E[D_{on}^{real}]$ of the on-phases for the G.723.1 codec.

As this procedure disregards the short breaks within a talkspurt, it overestimates the voice activity factor (VAF) α which is the fraction of generated packet vs. generated and suppressed packets. However, the VAF impacts the net load and needs to be modelled accurately. We propose to adapt the phase durations (**APD**) of the measured on/off phases in such a way that the original VAF is met and that the sum of the average durations of the on/off phases is the same as the one obtained from measurement. Alternatively, we introduce breaks (**IB**) within the on-phase to model the original trace structure. The distribution of the talkspurts and silence phases shows a coefficient of variation of about $c_{var}[D] = 0.6$. While the geometric distribution (Geom) cannot fit this parameter when the mean value $E[D]$ is given, a fit with the negative binomial distribution (NBin) takes both parameters into account.

We calculated the autocorrelation function (ACF) of consecutive packet sizes assuming a packet size of 0 bytes for suppressed packets. The ACF of existing traffic models and those obtained from W0 measurements heavily underestimate the ACF of the original traces while our proposed models capture it quite well. To validate the queuing properties of the models, we multiplex 20 synthetically generated flows onto a single link at a utilization of 60% and compare the obtained waiting times with those gained from original traces in the same experiment. Figure 3 shows that the widely used model from [10] yields clearly shorter waiting times than the original traces when we adapt the mean duration of its on/off phases to match the VAF of the original traces (Geom-469/533). Geometrically sampled on/off durations with a mean value according to W0 measurements also lead to shorter waiting times. APD-based models cause overestimation, but IB-based models capture the queuing properties of the original traces very well. The exact parameters of the traffic models are presented in [8].

3 VBR Voice Codecs

The iSAC [5] codec is a proprietary codec by Global IP Sound (GIPS) which produces a bit rate between 10 kbit/s and 32 kbit/s. It is one of several codecs being used by the VoIP client Skype. It adapts its transmission rate to the quality of the communication channel, i.e., the Skype implementation of the iSAC codec increases its bit rate, possibly to counteract packet loss by increasing information redundancy. However, we concentrate on the behavior of the codec under perfect network conditions. Figure 4 shows a typical time series of consecutive packets from the iSAC codec which every 30 ms produces packets between 21 and 166 bytes. In contrast to Figure 1 no talkspurt and silence phases can be recognized.

The simplest idea to generate such streams is to sample consecutive packet sizes randomly according to their empirical distribution gained from the original packet traces. However, this leads to uncorrelated packet sizes while the analysis of the empirical data shows a strong positive correlation even for packet sizes with a distance of 100, i.e. a duration of 3 s. To

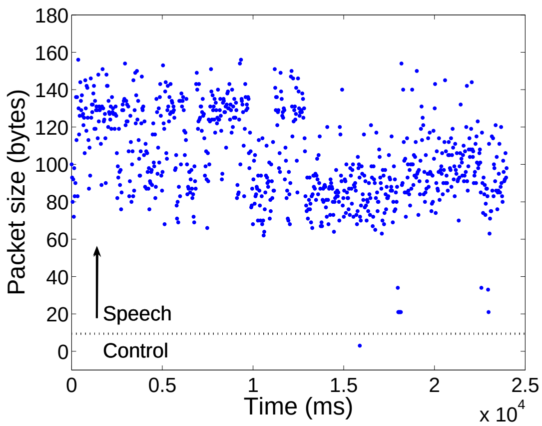

Figure 4: Time series of consecutive packet sizes for the iSAC codec.

capture this fact, we propose to apply a memory Markov chain (MMC) [9] and the procedure to fit its parameters according to a given data set. The MMC is a two-dimensional Markov chain with a two-dimensional state space (m_i^s, m_i^a) where m_i^s corresponds to different packet sizes and m_i^a to the average size of the last packets. It is able to produce time series with positive correlations. We chose a small number of different packet sizes for our model to keep it small, thus, it approximates the empirical distribution using a few discrete packet sizes. We observe that its ACF is in good accordance with the one of the original traces.

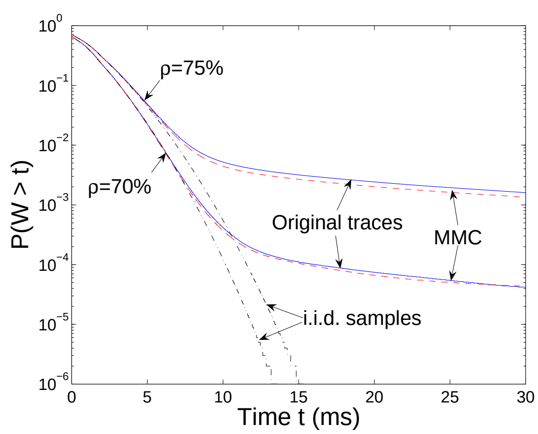

Figure 5: Validation of different traffic models for the iSAC codec with original traces using the waiting time of multiplexed sources.

Like in Section 2, we validate the MMC models by comparing its queuing property with the one of the original traces. Figure 5 shows the waiting time of 20 multiplexed streams at a utilization of 70 and 75%. The waiting time of the flows generated by the MMC model matches the one of the original traces very well. In contrast, the flows consisting of packets sampled i.i.d. from the empirical distribution heavily underestimate the waiting times.

4 Conclusion

In this work we have shown that frequently used source models for speech traffic underestimate the waiting time of multiplexed sources compared to real traffic because they do not capture the autocorrelation of consecutive packet sizes sufficiently well. Choosing the duration of the on/off phases for on/off traffic long enough leads to a good accordance regarding the queuing properties of synthetical flows and original traces. The same is achieved for VBR traffic by using a memory Markov chain with appropriate parameters. The proposed models are simple and can be used for simulative or analytical performance studies in the telecommunication area.

Acknowledgements

The authors would like to thank Paul Kühn, Danielle Liu, Daniel Minder, Oliver Rose, Kotikalapudi Sriram, Phuoc Tran-Gia, and Ward Whitt for valuable pointers and fruitful discussions.

References

[1] Bavarian Archive for Speech Signals (BAS). Verbmobil 6.1. `http://www.phonetik.uni-muenchen.de/Bas/BasHomedeu.html`, 1996.

[2] P. T. Brady. A Technique for Investigating On-Off Patterns of Speech. *Bell Systems Technical Journal*, 44(1):1–22, Jan. 1965.

[3] P. T. Brady. A Statistical Analysis of On-Off Patterns in 16 Conversations. *Bell Systems Technical Journal*, 47(1):73–91, Jan. 1968.

[4] P. T. Brady. A Model for Generating ON-OFF Speech Patterns in Two-Way Conversations. *Bell Systems Technical Journal*, 48(9):2445–2472, Sept. 1969.

[5] Global IP Sound. iSAC. `http://www.globalipsound.com/datasheets/iSAC.pdf`.

[6] ITU-T. G.723.1: Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 And 6.3 kbit/s.

[7] C. J. May and T. J. Zebo. private work, 1981.

[8] M. Menth, A. Binzenhöfer, and S. Mühleck. Source Models for Speech Traffic Revisited. Technical Report, No. 426, University of Würzburg, Institute of Computer Science, May 2007.

[9] O. Rose. A Memory Markov Chain Model for VBR Traffic with Strong Positive Correlations. In 16^{th} *International Teletraffic Congress (ITC)*, pages 827–836, Edingburgh, United Kingdom, June 1999.

[10] K. Sriram and W. Whitt. Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data. *IEEE Journal on Selected Areas in Communications*, 4(6):833–846, Sept. 1986.

[11] M. Vitez. Picophone. `http://www.vitez.it/picophone/`.