University of Würzburg
Institute of Computer Science
Research Report Series

Source Models for Speech Traffic Revisited

Michael Menth, Andreas Binzenhöfer, and Stefan Mühleck

Report No. 426 May 2007

University of Würzburg, Institute of Computer Science, Germany
{menth,binzenhoefer,muehleck}@informatik.uni-wuerzburg.de

Source Models for Speech Traffic Revisited

Michael Menth, Andreas Binzenhöfer,
and Stefan Mühleck
University of Würzburg, Institute of Computer Science,
Germany
{menth,binzenhoefer,muehleck}@informatik.
uni-wuerzburg.de

Abstract

In this paper, we analyze output traces of different voice codecs and present analytical models to describe them by stochastic processes. Both the G.711 and the G.729.1 codec yield constant bit rate traffic, the G.723.1 as well as the iLBC codec use silence detection leading to an on/off process, and the GSM AMR and the iSAC codec produce variable bit rate (VBR) traffic. We apply all codecs to a large set of typical speech samples and provide quantitative models that are based on standard and modified on/off processes as well as memory Markov chains. Our models are in good accordance with the original traces as they capture the complementary cumulative distribution function (CCDF) of the on/off phase durations and the packet sizes, the autocorrelation function (ACF) of consecutive packet sizes, and the queuing properties of the original traces. In addition, they are rather simple which makes them especially useful for application in analytical and simulative studies. The durations of the on/off phases in our model are an order of magnitude larger than those found by previous work and to the best of our knowledge this is the first paper presenting a model for VBR voice.

1 Introduction

Speech is usually sampled at a frequency of 8 kHz and each probe is coded by one byte resulting in a bit rate of 64 kbit/s. This information used to be transmitted continuously over circuits in the public switched telephone network. However, in packet-switched networks several probes are collected from intervals of fixed length T, put into a packet equipped with header information, and transmitted. This saves transmission overhead for individual probes. However, the packetization delay contributes to the end-to-end delay seen by the application and, therefore, it cannot be chosen arbitrarily large. Typical values for T are 20 or 30 ms depending on the voice codec.

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 the 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.

This paper presents simple stochastic models for different types of coded speech. For on/off processes we study several models of different complexity and accuracy, and for VBR traffic we take advantage of memory Markov chains [1]. We analyze a large set of sample traces to

parameterize the different models. We validate the models by showing that they well capture the complementary cumulative distribution function (CCDF) of the on/off phase durations of the original traces and the packet sizes, the autocorrelation function (ACF) of consecutive packet sizes, and the waiting time when several synthetic processes are fed simultaneously to a single server queue.

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 [2, 3, 4, 5, 6, 7, 8, 9, 10] and simulation tools like OPNET use by default an on/off model with exponentially distributed on/off phases with a duration of 352 and 650 ms, respectively. Most of them refer to [11] which cites "private work" [12]. We tried to track it, but without success. The work of Brady [13, 14, 15] 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.

Thus, the majority of research papers on this topic still relies on source models which were accurate in the 60ies and 70ies. However, our studies of recent voice codecs clearly show that those models are outdated and do no longer capture the characteristics of packetized voice traces on the packet level. Therefore, we present revised source traffic models which accurately describe the output of the currently most popular voice codecs. Since synthetically generated voice streams are often used for simulative or analytical performance studies in the telecommunication area, our findings are highly relevant and up-to-date.

The paper is structured as follows. Section 2 briefly reviews related work. In Section 3 we present measurement results of coded speech and derive quantitative stochastic models for typical vocoder output. We validate them by comparing the statistical properties and the queuing behavior of synthetic traces to those of the original traces. Section 4 summarizes our work.

2 Related Work

A general introduction to traffic models can be found in [16]. The paper describes the basic ability of different models to reproduce characteristics of the original process like a non-exponentially decaying autocorrelation. However, model parameters are not given.

Deriving source models for CBR speech traffic is a rather trivial task as those sources send packets of fixed size in regular intervals. Codecs using silence detection, in contrast, generate typical on/off packet processes which have often been characterized in literature. Silence or voice activity detectors (SD, VAD) may use a "hangover" to avoid "end-clipping" [17], i.e., they switch from the on-state to the off-state with delay and, thus, prolong the duration of the on-phase. The fill-in technique bridges a short gap between two intervals of voice activities and produces a longer on-phase. Thus, the output of vocoders depends on their parametrization. Most papers characterize the duration of uninterrupted activity or silence. Older papers measure analog voice while newer papers measure the generation or suppression of speech packets. Most of them study the duration of the on/off phases depending on the VAD sensitivity, the hangover, and the fill-in. They use an exponential or geometric approximation of the duration of the on/off phases, but point out that this simple model is not a good fit.

Early work [18] introduces the notion of talkspurts which is the duration of the speech of one party that may contain pauses. In later work, a talkspurt describes a contiguous interval of

recognizable speech, i.e., several talkspurts of a single party may follow each other. The work of Brady [13, 14] has reported different mean values for the duration of on/off phases depending on the sensitivity of the VAD: 1.31 s and 1.70 s for -45 dBm, 1.3 s and 1.72 s for -40 dBm, and 0.9 s and 1.66 s for -35 dBm. Altogether 137.4 min of two-way conversations were investigated, i.e., 274.8 min of speech. Parameters for a discrete-time Markov chain with two states are given in the paper to model the resulting output, but Brady also states that this is not a good fit. [15] presents an exponential model for generating on/off speech patterns in two-way conversations and reports a duration for the on/off phases of 1.2 s and 1.8 s. These parameters are used, e.g., in [19].

Most simulative and analytic studies use the values 352 ms and 650 ms for the duration of the on/off phases. They are reported in a paper of Sriram and Whitt [11] who cite the "private work" of May and Zebo [12]. Interestingly, many papers [2, 3, 4, 5, 6, 7, 8, 9, 10] use these values and some of them wrongly refer to some of Brady's works instead of citing [11] or [12]. In this paper, we refer to this traditional model by *Geom-352/650*.

The ITU P.59 [20] recommendation specifies an artificial on/off model for generating human speech. The durations of the talkspurts and silence intervals are 227 ms and 596 ms without hangover and 1.004 s and 1.587 s with hangover. Jiang and Schulzrinne investigated the G.729 Annex B VAD and the NeVoT SD [17] which use dynamic sensitivity thresholds to detect talkspurts and silence intervals. In addition, they discuss the impact of hangovers. They reported mean spurts and gaps of 293 ms and 306 ms for G.729B and 267 ms and 272 ms for NeVoT SD. They compared the queueing behavior of the empirical data with the one of an exponential model and showed that this is a bad approximation. The dependence of the talkspurt duration on the hangover interval has also been studied in [21, 22].

In [23] Deng et al. observe that the assumption of exponentially distributed talkspurts and silence intervals is not a good approximation. They tested packet voice from early VoIP tools like vat, NeVoT, Maven and recognized silence phases only if they are larger than 3 packets. As a consequence, Deng reports mean on/off phase durations for conversational speech of 7.24 s and 5.69 s which is already by an order of magnitude larger than the most widely used traditional *Geom-352/650* model. In [24] the distribution of the on/off phases of the codec traces is approximated by a Weibull distribution. Only on/off phases larger than 100 ms were recognized. The work reports mean talkspurt and silence durations of 1.58 s and 0.87 s. In [25], the codecs G.723.1, G.729B, and GSM FR were investigated. Their call level analysis provides a mean holding time of 114 s. Their packet level analysis reports mean durations of 2.28 s and 1.48 s, 2.37 s and 1.56 s, and 2.50 s and 1.55 s for the duration of the on/off phases for the three codecs. They propose to model their duration by a generalized Pareto distribution and found long range dependency in the rate of the superposition of several voice calls.

None of the above models considers the autocorrelations of the output of the codecs even though they are known to have an influence on the queuing behavior [26]. Li and Mark study the queue length distribution of multiplexed heterogeneous sources in [27]. Each source is modelled as a discrete-time on/off process with geometrically distributed on/off phases. The large impact of positive autocorrelations on the waiting time in queuing systems is mentioned but not expressed in terms of a quantitative measure.

In our work, we use a different interpretation of on/off phases which is similar to the one of [18]. On/off phases are recognized as such only if they are sufficiently long, otherwise we inter-

pret them just as short breaks or noise within on- or off-phases (cf. Figure 1). As a consequence, we report mean durations of the on/off phases in the order of 11 s which is an order of magnitude larger than those reported in the papers above.

In contrast to the above results, we show that based on our on/off phase durations a geometric model leads to a good approximation of the queuing properties of voice traffic. More elaborated models provide a good fit of the distribution function of the phase durations and the autocorrelation function of consecutive packet sizes. Thus, simple exponential or geometric models can be further applied, but analytical or simulative studies should use appropriate mean values for the duration of the on/off phases.

Although many papers model VBR video traffic [28, 29, 30], we are not aware of any source models for VBR voice codecs in the literature.

3 Source Models for Speech Traffic

In this section we consider two representatives of each of the three different vocoder types: constant bit rate (CBR) codecs, codecs with silence detection, and variable bit rate (VBR) codecs. We apply each codec to a large set of typical telephone conversations (3.5 h = 7 h speech) from [31], 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.

3.1 Voice Codecs with Constant Bit Rate

CBR codecs send a bit stream of constant rate which is independent of the voice input. The ITU G.711 [32] codec is mainly applied in digital telephony and uses pulse code modulation (PCM) sampled at a rate of 8 kHz and 8 bits per sample which results in a 64 kbit/s stream. The algorithmic complexity is very low and due to the relatively high bandwidth usage, the voice quality is very good and often used as a reference. The ITU G.729.1 standard [33] was also designed for voice communication and adds wideband functionality to the G.729 standard by offering different bit rates from 14 to 32 kbit/s in steps of 2 kbit/s. To analyze the behavior of the codecs in practice, we measured the output stream of the G.711 and the G.729.1 codec. Voice packets are usually transmitted using UDP over IPv4 entailing a header overhead of 8 and 20 bytes, respectively. However, it is also possible to use additional or alternative headers like RTP (at least 12 bytes) or IPv6 (40 bytes). To be independent of the network layer, we concentrate on the plain output of the codecs disregarding any headers.

We measured the G.711 codec using CounterPath's X-Lite [34], a freely available SIP based softphone which produces a main stream of 68.8 kbit/s. The implementation of the codec sends its control information separately as well as piggybacked on regular data packets. The trace of the G.729.1 codec was obtained using SkypeOut [35] to call a regular landline user from Skype. The codec strictly differentiates between control information and actual speech data. Thus, both

Table 1: Packet types of the G.711 and the G.729.1 codec.

Codec	Type	Packet size	Period
G.711	Control	4 bytes	30s
	Speech	172 bytes	20ms
	Speech + control	176 bytes	3s
G.729.1	Control	5 bytes	1s
	Speech	38 bytes	20ms

codecs send periodic control information in addition to their main audio stream. Table 1 gives a detailed description of the packet sizes and the periods at which they are sent. Due to the simplicity of the codecs in this category and the fact that their output rates are independent of the input Table 1 suffices to easily generate synthetic streams for simulations or analytical studies.

3.2 Voice Codecs with Silence Detection

Voice codecs with silence detection are able to detect voice activity in terms of "speech on" or "speech off" and transmit packets of fixed size only while the user is talking. Thus, the output on the network layer consist of contiguous talkspurts and silence intervals, the so-called on- and off-phases. Two prominent examples for such codecs are the G.723.1 [36] and the iLBC

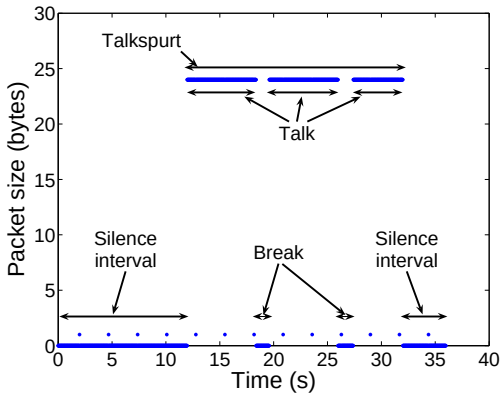


Figure 1: G.723.1 outputs an audio stream and control information. The audio stream consists of silence intervals and main talkspurts that are interrupted by short breaks and noise.

vocoder. The G.723.1 codec is an ITU-T standard since 1995 which was specially designed for voice coding 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 6.4 kbit/s with 24 bytes chunks or 5.3 kbit/s with 20 bytes chunks every 30 ms. We generate packet traces using the Picophone software [37] which relies on the reference implementations of Microsoft. G.723.1 produces a main audio streams of fixed packet sizes and sends additional control information of

1 byte every 3 s (cf. Figure 1).

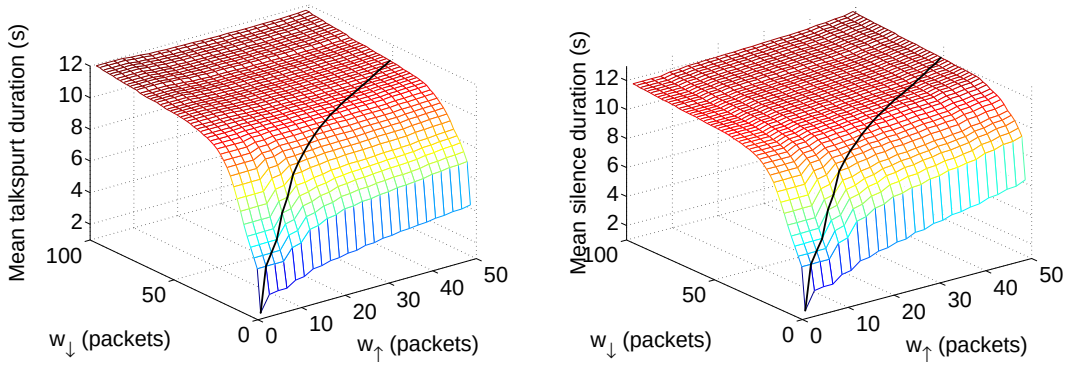
The Internet Low Bit Rate Codec (iLBC) [38] developed by Global IP Sound (GIPS) is suitable for robust voice communication over IP. It is designed for narrow band speech and results in a payload bit rate of 13.33 kbit/s for 30 ms frames and 15.20 kbit/s for 20 ms frames. The codec enables graceful speech quality degradation in the case of lost frames, which occurs in connection with lost or delayed IP packets. We used the implementation of XLite sending 62 bytes every 30 ms resulting in a bit rate of 16.53 kbit/s which is slightly larger than the one indicated in the standard. Thus, some control information seems to be piggybacked.

Figure 1 shows a typical packet trace of the G.723.1 codec. No audio packets are transmitted during a silence interval. The talkspurts, however, are interrupted by short breaks which arise from short pauses a speaker makes while talking. Obviously, the codec detects these pauses and temporarily stops the transmission of voice packets. Due to this noisy structure, the automatic detection of the beginning and end of major talkspurts is difficult. We discuss three different approaches for their recognition.

(0W) We take contiguous on- and off-phases as observed in the original trace such that major talk spurts are cut in pieces. This method has been applied by previous work.

(1W) We require that on- and off-phases start with at least w consecutive generated or suppressed packets, which can easily be controlled by a single moving window.

(2W) We require that on-phases start with at least $w\uparrow$ consecutive generated packets and that off-phases start with at least $w\downarrow$ consecutive suppressed packets, which can be controlled by two different moving windows.



(a) Talkspurts (b) Silence intervals

Figure 2: Impact of the window parameters $w\downarrow$ and $w\uparrow$ on the measured mean duration of the on-phases for the G.723.1 codec.

Figures 2(a) and 2(b) show the measured mean duration of the on-phases measured by the 2W-approach depending on the values of $w\downarrow$ and $w\uparrow$. The mean durations $E[D_{on}^{real}]$ and $E[D_{off}^{real}]$

increase with both window sizes, but we recognize a stable plateau for $w\downarrow \geq 50$ and $w\uparrow \geq 15$. This means that neglecting short breaks within on-phases of less than 1.5 s and short noise within off-phases of less than 450 ms leads to relatively stable measured mean values for the durations of on/off phases. We consider the window parameters $w\downarrow = 50$ and $w\uparrow = 15$ useful and use them in the following as standard for the W2-approach.

The measured mean values for $w\downarrow = 1$ and $w\uparrow = 1$ in both figures correspond to the W0-approach and the solid lines correspond to measured mean values for the W1-approach. The measured mean value for the W0-approach underestimates the length of "almost contiguous" on-phases by an order of magnitude. The W1-approach is unnecessarily insensitive to on-phases for large windows sizes.

A similar behavior is observed for the iLBC codec for the same values of $w\downarrow$ and $w\uparrow$. The statistical properties of the on/off phase durations are given for the G.723.1 and the iLBC codec in Tables 2 and 3 for the W0- and the W2-approach.

Table 2: Statistics about on/off phase durations *based on W0-measurements* including parameters for the corresponding *NBin* and the *Geom* approximation in packets.

Codec	G.723.1		iLBC	
Phase	on	off	on	off
$E[D^{real}]$	1.304 s	1.480 s	3.113 s	3.279 s
$c_{var}[D^{real}]$	1.7938	2.9858	0.7697	1.9152
r^{NBin}	0.31302	0.11243	1.71571	0.27330
p^{NBin}	$7.14891 \cdot 10^{-3}$	$2.27372 \cdot 10^{-3}$	$1.62657 \cdot 10^{-2}$	$2.49416 \cdot 10^{-3}$
p^{Geom}	$2.24859 \cdot 10^{-2}$	$1.98671 \cdot 10^{-2}$	$9.54523 \cdot 10^{-3}$	$9.06599 \cdot 10^{-3}$

Table 3: Statistics about on/off phase durations *based on W2-measurements* including parameters for the corresponding *NBin* and the *Geom* approximation in packets.

Codec	G.723.1		iLBC	
Phase	on	off	on	off
$E[D^{real}]$	11.54 s	11.98 s	11.23 s	11.31 s
$c_{var}[D^{real}]$	0.61003	0.60261	0.58344	0.61887
r^{NBin}	2.70609	2.77289	2.96094	2.62917
p^{NBin}	$6.98575 \cdot 10^{-3}$	$6.89591 \cdot 10^{-3}$	$7.84782 \cdot 10^{-3}$	$6.92564 \cdot 10^{-3}$
p^{Geom}	$2.59291 \cdot 10^{-3}$	$2.49792 \cdot 10^{-3}$	$2.66430 \cdot 10^{-3}$	$2.64550 \cdot 10^{-3}$

The voice activity factor (VAF) α is the fraction of the number of generated packets and the number of generated and suppressed packets. For the G.723.1 we get $\alpha = 0.44332$ from our measurements and for the iLBC we get $\alpha = 0.48835$. We approximate the distribution of the length of the talkspurts and the silence intervals in packets with the geometric distribution (*Geom*), i.e. $P(X^{Geom} = k) = p^{Geom} \cdot (1 - p^{Geom})^k$, and the negative binomial distribution

Table 4: Statistics about on/off phase durations *based on W2-APD-measurements* including parameters for the corresponding *NBin* and the *Geom* approximation in packets.

Codec	G.723.1		iLBC	
Phase	on	off	on	off
$E[D^{APD}]$	10.43 s	13.09 s	11.01 s	11.53 s
$c_{var}[D^{APD}]$	0.61003	0.60261	0.58344	0.61887
r^{NBin}	2.70812	2.77125	2.96141	2.62882
p^{NBin}	$7.73151 \cdot 10^{-3}$	$6.30966 \cdot 10^{-3}$	$8.00652 \cdot 10^{-3}$	$6.79197 \cdot 10^{-3}$
p^{Geom}	$2.86892 \cdot 10^{-3}$	$2.28604 \cdot 10^{-3}$	$2.71803 \cdot 10^{-3}$	$2.59457 \cdot 10^{-3}$

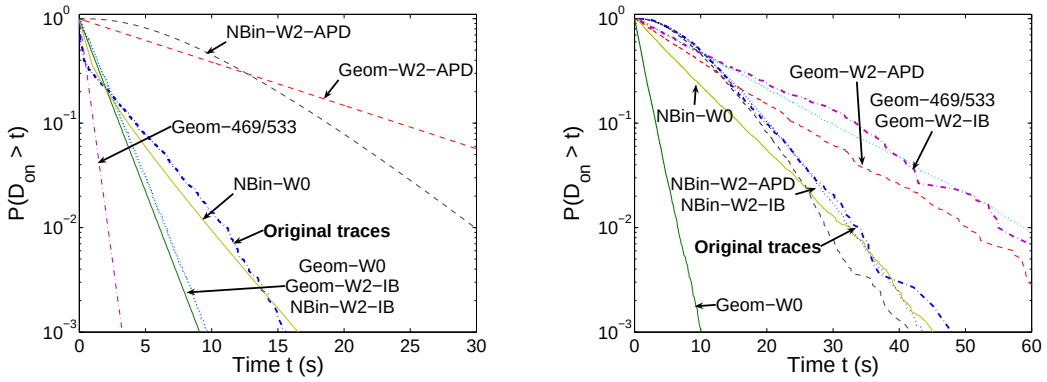
(*NBin*), i.e. $P(X^{NBin} = k) = \frac{\Gamma(r+k)}{k! \cdot \Gamma(r)} \cdot (p^{NBin})^r \cdot (1-p^{NBin})^k$, where Γ is the gamma function. Modelling truly contiguous on- and off-phase durations based on the measured mean values $E[D_{on}^{real}]$ and $E[D_{off}^{real}]$ given in Table 3 neglects the many breaks within a talkspurt resulting in an overestimated VAF. We propose two different approaches to tackle this problem:

(APD) Adapt phase durations: we adjust the mean duration of the on/off phases measured by the W2-approach in such a way that the original VAF is met, i.e., we use $E[D_{on}^{APD}] = \alpha \cdot (E[D_{on}^{real}] + E[D_{off}^{real}])$ and $E[D_{off}^{APD}] = (1-\alpha) \cdot (E[D_{on}^{real}] + E[D_{off}^{real}])$ to model the durations of contiguous on/off phases.

(IB) Introduce breaks: we use $E[D_{on}^{real}]$ and $E[D_{off}^{real}]$ of Table 3 to model the length of the major talkspurts and silence intervals and generate talk and break phases within the talkspurts as observed in Figure 1 by geometric distributions. To that end, we measure the average durations of the talk and break phases observed within talkspurts and obtain $E[D_{talk}^{real}] = 1.464$ s and $E[D_{break}^{real}] = 0.102$ s for G.723.1 and $E[D_{talk}^{real}] = 3.128$ s and $E[D_{break}^{real}] = 0.103$ s for iLBC.

We denote on/off phase durations generated by the geometric and negative-binomial distribution based on measurements from the W0-approach by {*Geom, NBin*}-*W0*. If they are based on the measurements from the W2-approach, we denote them by {*Geom, NBin*}-*W2*-{*APD, IB*} to indicate how the VAD is corrected. The parameters for the generation of the on- and off-phases (in packets) are derived as $r^{NBin} = \frac{E[D_{packets}^{real}]}{E[D_{packets}^{real}] \cdot c_{var}[D_{packets}^{real}]^2 - 1}$ and $p^{NBin} = \frac{1}{E[D_{packets}^{real}] \cdot c_{var}[D_{packets}^{real}]^2}$ and summarized in Table 2 for {*Geom, NBin*}-*W0* and in Table 3 for {*Geom, NBin*}-*W2*-{*APD*}. A fair comparison of the traditional model *Geom-352/650* and the G.723.1 output requires that both have the same VAD. Therefore, we adapt the average length of its on/off phases and get *Geom-469/533*.

Figures 3(a) and 3(b) show the complementary cumulative distribution functions (CCDF) of the on-phase duration for the original traces and different models for the G.723.1 codec. Figure 3(a) is obtained with W0-measurement while Figure 3(b) is obtained with W2-measurement. Looking at Figure 3(a) on the one hand, the traditional model *Geom-469/533* has significantly shorter on-phase durations and {*Geom, NBin*}-*W2-APD* have significantly longer on-phase durations compared to the original traces. The accordance of the curves for {*Geom, NBin*}-*W2-IB*,



(a) Measured by the W0-approach

(b) Measured by the W2-approach

Figure 3: CCDFs of the on-phase durations for the original traces and different traffic models.

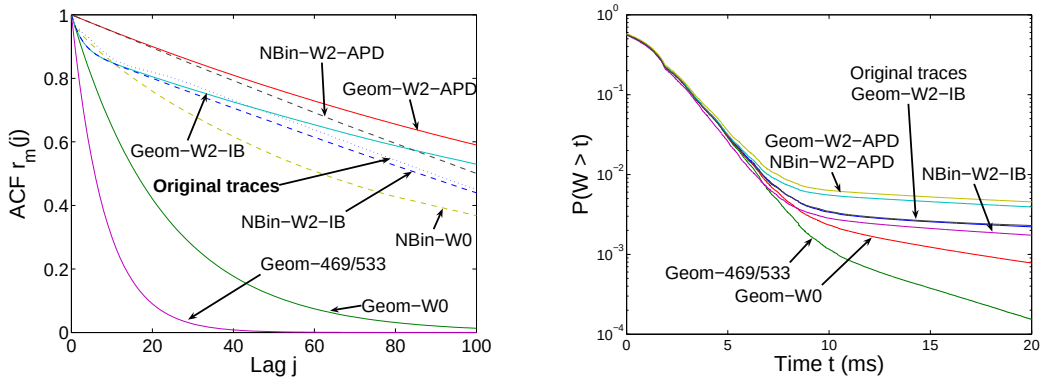
Geom-W0 with the curve for the original traces is acceptable, but not good. Only *NBin-W2-IB* seems to be a good fit.

Looking at Figure 3(b) on the other hand, the on-phase durations are clearly longer due to the W2-measurement method. The traditional model *Geom-469/533* surprisingly overestimates the on-phase durations ($E[D_{on}^{W2}] = 12.59$ and $E[D_{on}^{W2}] = 2.98$) because off-phases are recognized only if they are longer than 450 ms, i.e., the recognized on-phases contain many relatively large breaks. As a consequence, the VAF of the W2-measured trace is $\alpha = 0.8084$ instead of $\alpha = 0.44332$ for the W0-measured original traces. *Geom-W0* heavily underestimates the durations ($E[D_{on}^{W2}] = 1.39$ and $E[D_{on}^{W2}] = 1.58$) and so does *NBin-W0* ($E[D_{on}^{W2}] = 7.05$ and $E[D_{on}^{W2}] = 6.41$) although hardly visible in Figure 3(b). The CCDFs of the *Geom-W2-*{*APD, IB*} do not well approximate the CCDF of the original traces, but *NBin-W2-*{*APD, IB*} lead to a fairly good match. Combining the results of the W0- and W2-measurement, *NBin-W2-IB* provides the best fit for the original traces on different time scales.

The empirical autocorrelation function (ACF) for lag j can be calculated from m consecutive random variables (RV) X_i ($0 \leq i < m$) by $r_m(j) = \frac{\hat{C}_m(j)}{S_m^2}$ where S_m^2 is the empirical variance and $\hat{C}_m(j) = \frac{1}{m-j} \cdot \sum_{0 \leq i < m-j} (X_i - \overline{X}) \cdot (X_{i+j} - \overline{X})$ the empirical autocovariance of the m RVs. The values of $r_m(j)$ range between -1 and 1. If $r_m(j)$ is close to 1, RVs X_i and X_{i+j} have almost perfect correlation, if it is close to -1, they have almost perfect anti-correlation. If consecutive RVs are independent and identically distributed (iid), an ACF of $r_m(j) \approx 0$ can be expected for any lag $l > 0$.

To validate the different models, we consider the ACFs of consecutive packet sizes which are either zero or the standard packet size. The mean durations of the on/off phases have a significant impact on the ACFs. Figure 4(a) shows that the original traces reveal strong positive ACF values even for large lags. The ACF values for the traditional model *Geom-469/533* and *Geom-W0* are significantly lower than those of the original traces. The same hold for *NBin-W0* but to a minor degree. {*Geom, NBin*}*-W2-APD* clearly overestimate the ACF of the original traces, but

{*Geom, NBin*}*-W2-IB* match them fairly well.



(a) ACF of consecutive packet sizes

(b) CCDF of packet waiting times W of an $n \cdot D/G/1$–∞ queue

Figure 4: Validation of different traffic models for the G.723.1 codec with original traces.

Furthermore, we compare the queuing properties of the discussed models. To that end, we multiplex $n = 20$ periodic flows onto a single link, i.e., we consider an $n \cdot D/G/1-\infty$ queue. Packet sizes of different flows are independent of each other. They are generated either by one of our models or by randomly copying original traces. We choose the link capacity such that a link utilization of 60% is achieved and measure the resulting waiting time W of the packets. Due to the periodic nature of the traffic, this scenario is hard to simulate because the placement of the transmission instants of the flows within the periods impacts the waiting time significantly. Therefore, we repeat this experiment 500 times using random placements for different runs. We simulate 50000 periods for each run and cut off a warmup phase of 100 periods before collecting the statistics.

The CCDF of the packet waiting times are presented in Figure 4(b) for the G.723.1 codec. The CCDF values decrease rather quickly for increasing waiting times, but remain almost constant at a level of $2 \cdot 10^{-3}$. These long waiting times occur if sufficiently many flows are in the on-phase and if the bandwidth does not suffice to carry the traffic when all flows are in the on-phase. Thus, overload occurs which leads to significant queuing and potentially to packet loss due to buffer overflow. Reducing the utilization by increasing the virtual bandwidth in the experiment decreases the probability for very long waiting times. We have chosen a relatively large utilization of 60% to make the differences of the queuing properties of our models visible.

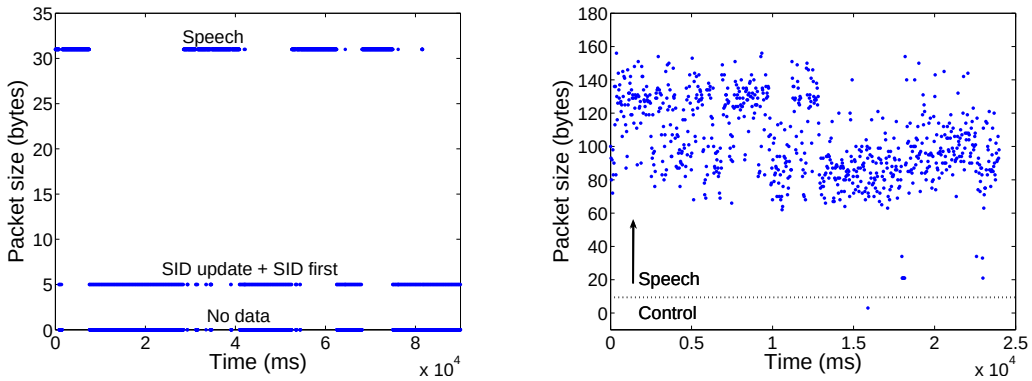
The adapted traditional model *Geom-469/533* heavily underestimates the waiting times of the original traces and so do {*Geom, NBin*}*-W0* to a minor degree. {*Geom, NBin*}*-W2-APD* overestimate them slightly, thus, they provide a conservative approximation for the queuing properties of the original traces. The packet waiting times of {*Geom, NBin*}*-W2-IB* are in good accordance with those of the original traces.

Summarizing, synthetic flows generated by the widely used source model for speech traffic *Geom-352/650* have too optimistic queuing properties as the duration of their on/off phases is

too short. In contrast, the queuing properties of the simple *Geom-W2-APD* model provide a conservative approximation of those of the original traces and *Geom-W2-IB* is a perfect match. However, the ACF of *Geom-W2-APD* overestimates the one of the original traces, so the more complex model *Geom-W2-IB* might be used. To meet the CCDF of the on/off phase durations in addition, *NBin-W2-IB* should be used.

3.3 Voice Codecs with Variable Bit Rate

Finally, we consider more sophisticated audio codecs that produce packets of different size depending on the speech input and lead to VBR traffic. The GSM Adaptive Multi-Rate (AMR) [39] codec is the default speech codec for third generation wireless systems and operates at a rate between 4.75 kbit/s and 12.2 kbit/s. GSM is the most widely used standard for mobile phones and the measurements were obtained using the 3GPP reference implementation of the GSM AMR. The iSAC [40] 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 [35]. Both codecs adapt their transmission rates to the quality of the communication channel. While GSM AMR decreases the size of its speech packets in times of bad transmission quality to save bandwidth, the Skype implementation of the iSAC codec increases its bit rate, possibly to counteract packet loss by increasing information redundancy. In this paper, however, we concentrate on the behavior of the codecs under perfect network conditions.



(a) GSM AMR (b) iSAC

Figure 5: Time series of packet sizes.

Figures 5(a) and 5(b) show typical traces for the GSM AMR and the iSAC codec. GSM AMR has a VAD, but in contrast to previous codecs, it sends empty packets for synchronization purposes instead of omitting them when there is no data to send. In addition, it produces silence descriptor (SID) packets which describe the recorded background noise to create adequate comfort noise at the receiver side in phases of silence. Therefore, the packet stream on the network layer does not result in an on/off process. The iSAC vocoder dynamically produces packets of many different sizes and yields a true VBR stream being significantly different from an on/off

process on the packet level. However, we still recognize clusters in Figure 5(b) with relatively large or small packets which correspond to on/off phases. Table 5 summarizes for both codecs information about individual packet sizes and periods at which they are sent.

Table 5: Statistical information about packet sizes and periods for original traces and the modelling MMC for both the GSM AMR and the iSAC codec.

Codec	GSM AMR
No data	0 bytes
SID update	5 bytes
SID first	5 bytes
Speech	31 bytes
Speech period	20 ms
$E[B^{\text{real}}]$	14.097 bytes
$c_{var}[B^{\text{real}}]$	1.0727
$E[B^{\text{MMC}}]$	14.096 bytes
$c_{var}[B^{\text{MMC}}]$	1.0728
n_s	3
n_a	10
W_a	15

Codec	iSAC
Control size	3 bytes
Control period	20 s
min(packet size)	21 bytes
max(packet size)	166 bytes
Speech period	30 ms
$E[B^{\text{real}}]$	71.319 bytes
$c_{var}[B^{\text{real}}]$	0.626
$E[B^{\text{MMC}}]$	71.321 bytes
$c_{var}[B^{\text{MMC}}]$	0.621
n_s	7
n_a	15
W_a	12

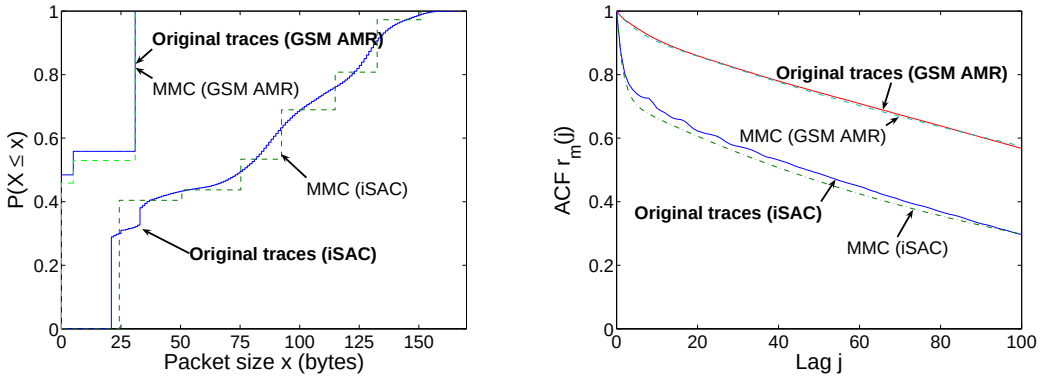
As on/off processes cannot model time series of different packet sizes, we use a memory Markov chain (MMC) [1] for that objective. An MMC is a Markov chain with a two-dimensional state (m_i^s, m_i^a). The values m_i^s and m_i^a can take n_s and n_a different values s_j and a_j, respectively. We use the following serialization of the two-dimensional state space:
$((s_0, a_0), ..., (s_{n_s-1}, a_0), ..., (s_0, a_{n_a-1}), ..., (s_{n_s-1}, a_{n_a-1}))$. This equivalent conventional one-dimensional Markov chain has a $(n_s \cdot n_a) \times (n_s \cdot n_a)$ transition matrix. In our context, the s_j are packet sizes and the a_j correspond to the average of the last W_a packets. Thus, the m_i^s-projection of the MMC-state yields a synthetic trace of packet sizes.

The MMC can model time series X_i with strong positive correlations and a recipe is given in [1]. The X_i are discretized into n_s different values s_j and the corresponding moving averages $\overline{X}_i = \frac{1}{W_a} \cdot \sum_{0<k\leq W_a} X_{i-k}$ are discretized into n_a different values a_j. Thus, the tuples $(X_i, \overline{X}_i)$ are discretized into tuples $(X_i^d, \overline{X}_i^d)$. The empirical transition probabilities of the discretized process $(X_i^d, \overline{X}_i^d)$ are taken as the entries in the transition matrix of the MMC. In the following, we characterize memory Markov chains MMC(n_s, n_a, W_a) by the values of their parameters n_s, n_a, and W_a.

We tested different parameter settings to model the vocoder output by an appropriate MMC. The search for optimal parameters was performed until the ACF of the original trace and the MMC matched sufficiently well. Removing iSAC's control traffic leads to better results. The parameters for both codecs are given in Table 5. Due to the lack of space we omit the presentation of the discretized packet sizes s_j and the transition matrices of the MMCs, but provide them for download from [41] or upon email request.

We validate the MMC models for the GSM AMR and the iSAC codec by comparing the statistical properties of their synthetic packet traces to those of the original traces. Table 5 shows that the corresponding mean values and coefficients of variation hardly differ. Figure 6(a)

compares the analytically derived CDFs of the packet sizes to those of the original traces. The CDF of the MMC$(3, 10, 15)$ model for the GSM AMR coincides with the one of the empirical data since the original codec also outputs only three different packet sizes. The trace of the iSAC codec has a more stepless distribution of the packet sizes, but the seven discretization levels of the MMC$(7, 15, 12)$ model reproduce the empirical distribution quite well. More discretization levels lead to a better approximation, but in this case the tradeoff was made towards a simpler and faster computable MMC.



(a) CDF of the packet size X

(b) ACFs of consecutive packet sizes

Figure 6: Validation of the MMC model with the original traces for the GSM AMR and the iSAC codec.

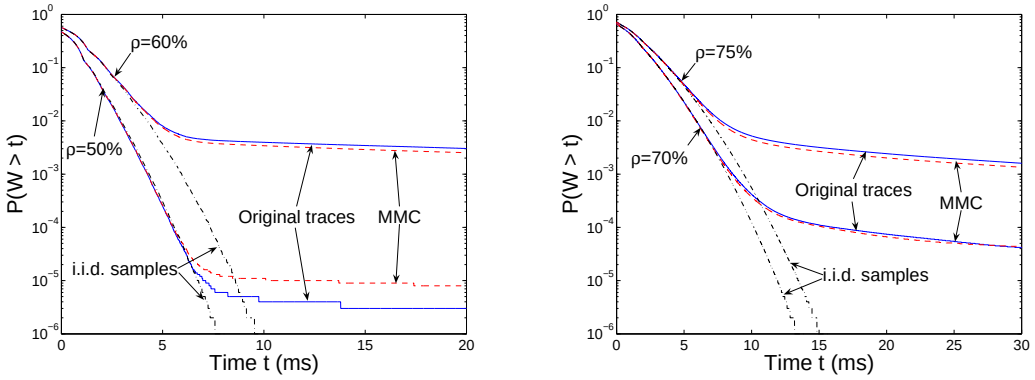
Figure 6(b) shows that the ACFs of the presented MMCs match those of the original sample traces very well for both codecs. We omitted the ACF for iid packet sizes that are generated based on the empirical distribution because they yield $r_m(j) = 0$ for all lags $j > 0$.

To compare the queuing properties of the analytical models to those of the original traces, we feed their output to an $n \cdot D/G/1 - \infty$ queue like in Section 3.2. We use $n = 20$ sources and choose the link bandwidth such that the system operates at different utilizations. Figures 7(a)–7(b) show the CCDF of the obtained packet waiting times for the GSM AMR and for the iSAC codec. The CCDFs for the original traces and the MMC match quite well for different load levels while the CCDFs of iid packet sizes sampled according to the empirical distribution function underestimate the waiting time of the sample traces significantly.

Summarizing, the MMC-based model approximates the CDF of the packet sizes, the ACFs, and the queuing properties of VBR voice sources quite well while periodically sampled iid packet sizes fail to do so. It is simple enough to be integrated in any simulation software and appropriate parameter sets are available at [41] for GSM AMR and iSAC.

4 Conclusion

In this paper, we studied the output of fundamentally different voice codecs. To that end, we sampled a large set of standard telephone conversations [31] and analyzed the vocoder output.



(a) GSM AMR codec

(b) iSAC codec

Figure 7: CCDFs of packet waiting times W of an $n \cdot D/G/1 - \infty$ queue fed by original traces and synthetic traffic.

We proposed stochastic models approximating the properties of original traces and validated them. These models are useful for analytic and simulative performance studies.

G.711 and G.729.1 are constant bit rate codecs sending packets of fixed size in regular intervals. They differ in the length of these intervals, in the transmitted packet size, and the associated control information.

G.723.1 and iLBC are codecs with silence detection producing fixed packet sizes but in an on/off manner. Individual on- and off-phases are difficult to determine as on-phases are interrupted by short breaks. However, they can easily be filtered by a two-window approach. The average durations of on- and off-phases in literature are in the order of 0.5 s, but we found them in the order of 1.4 s and 3.1 s depending on the codec without filtering and 11.5 s independently of the codec when the short breaks are filtered. The validation showed that synthetic traffic generated by geometrically distributed on/off phases with durations of 10.7 s and 12.1 s (*Geom-W2-APD* for G.723.1) have the same queuing properties as the original traces. The short breaks within the on-phases need to be modelled (*Geom-W2-IB*) to produce the same autocorrelation function (ACF) for consecutive packet sizes like in the original traces. To fit the cumulative distribution function (CDF) of the original traces, the duration of the on/off phases are better modelled by a negative-binomial distribution (*NBin-W2-IB*).

Variable bit rate (VBR) codecs such as the GSM AMR and iSAC also send data in regular intervals, but use variable packet sizes. We modelled the time series of consecutive packet sizes by a memory Markov chain (MMC). The synthetic output of the MMC matches the CDF of the packet sizes, the ACF of consecutive packet sizes, and the queuing properties of the original traces very well. IID packet sizes generated according to the empirical distribution function have a significantly different ACF and queueing properties. The full parametrization of the MMC can be downloaded from our website [41].

The most important result of this work is that the mostly used durations of on/off phases of 352 ms and 650 ms are too short such that their use in performances studies underestimates

packet loss and delay. Thus, future simulations or analyses using synthetic voice sources should better rely on our parameters. Furthermore, we provided an accurate traffic model for VBR codecs producing different packet sizes which has not been studied before.

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

[2] S. S. Wang and J. A. Silvester, "A Discrete-Time Performance Model for Integrated Service ATM Multiplexers," in *IEEE Globecom*, 1993.

[3] S. Sibal, K. Parthasarathy, and K. S. Vastola, "Sensing the State of Voice Sources to Improve Multiplexer Performance," in *IEEE Infocom*, 1995.

[4] S.-H. Jeong and J. A. Copeland, "Cell Loss Ratio and Multiplexing Gain of an ATM Multiplexer for VBR Voice Sources," in *IEEE Conference on Local Computer Networks (LCN)*, 1998.

[5] C. Liu, S. Munir, R. Jain, and S. Dixit, "Packing Density of Voice Trunking Using AAL2," in *IEEE Globecom*, 1999.

[6] A. Bengt, A. Anders, H. Olof, and M. Ian, "Dimensioning Links for IP Telephony," in *Internet Telephony Workshop*, (New York, USA), pp. 14–24, Apr. 2001.

[7] H. Sze, S. Liew, J. Lee, and D. Yip, " A Multiplexing Scheme for H.323 Voice-over-IP Applications," *IEEE Journal on Selected Areas in Communications*, vol. 20, 2002.

[8] M. K. Ranganathan and L. Kilmartin, "Performance Analysis of Secure Session Initiation Protocol Based VoIP Networks," *Computer Communications*, vol. 26, p. 552565, 2003.

[9] A. Koubaa and Y.-Q. Song, "Loss-Tolerant QoS using Firm Constraints in Guaranteed Rate Networks," in *IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, (Washington, DC, USA), 2004.

[10] S. Obeidat and S. Gupta, " Towards Voice over Ad Hoc Networks: an Adaptive Scheme for Packet Voice Communications over Wireless Links," in *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Aug. 2005.

[11] K. Sriram and W. Whitt, "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data," *IEEE Journal on Selected Areas in Communications*, vol. 4, pp. 833–846, Sept. 1986.

[12] C. J. May and T. J. Zebo, “private work,” 1981.

[13] P. T. Brady, “A Technique for Investigating On-Off Patterns of Speech,” *Bell Systems Technical Journal*, vol. 44, pp. 1–22, Jan. 1965.

[14] P. T. Brady, “A Statistical Analysis of On-Off Patterns in 16 Conversations,” *Bell Systems Technical Journal*, vol. 47, pp. 73–91, Jan. 1968.

[15] P. T. Brady, “A Model for Generating ON-OFF Speech Patterns in Two-Way Conversations,” *Bell Systems Technical Journal*, vol. 48, pp. 2445–2472, Sept. 1969.

[16] A. Adas, “Traffic Models in Broadband Networks,” *IEEE Communications Magazine*, pp. 82–89, July 1997.

[17] W. Jiang and H. Schulzrinne, “Analysis of On-Off Patterns in VoIP and their Effect on Voice Traffic Aggregation,” in *IEEE International Conference on Computer Communications and Networks (ICCCN)*, 2000.

[18] A. C. Norwine and O. J. Murphy, “Characteristic Time Intervals in Telephone Conversation,” *Bell Systems Technical Journal*, vol. 17, pp. 281–291, Apr. 1938.

[19] J. N. Daigle and J. D. Langford, “Models for Analysis of Packet Voice Communications Systems,” *IEEE Journal on Selected Areas in Communications*, vol. 4, pp. 847–855, Sept. 1986.

[20] ITU-T, “P.59: Telephone Transmission Quality Objective Measuring Apparatus: Artificial Conversational Speech,” Nov. 1996.

[21] J. G. Gruber, “A Comparison of Measured and Calculated Speech Temporal Parameters Relevant To Speech Activity Detection,” *IEEE Transactions on Communications*, vol. 30, pp. 728–738, Apr. 1982.

[22] H. H. Lee and C. K. Un, “A Study of On-Off Characteristics of Conversational Speech,” *IEEE Transactions on Communications*, vol. 34, pp. 630–637, June 1986.

[23] S. Deng, “Traffic Characteristics of Packet Voice,” *IEEE International Conference on Communications (ICC)*, June 1995.

[24] C.-N. Chuah and R. H. Katz, “Characterizing Packet Audio Streams from Internet Multimedia Applications,” in *IEEE International Conference on Communications (ICC)*, 2002.

[25] T. D. Dang, B. Sonkoly, and S. Molnár, “Fractal Analysis and Modeling of VoIP Traffic,” in *International Telecommunication Network Strategy and Planning Symposium (Networks)*, (Vienna, Austria), pp. 217 – 222, June 2004.

[26] M. Livny, B. Melamed, and A. K. Tsiolis, “The Impact of Autocorrelation on Queuing Systems,” *Management Science*, vol. 39, no. 3, pp. 322–339, 1993.

[27] S.-Q. Li and J. W. Mark, “Traffic Characterization for Integrated Services Network,” *IEEE Transactions on Communications*, vol. 38, pp. 1231–1243, Aug. 1990.

[28] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-Range Dependence in Variable-Bit-Rate Video Traffic," *IEEE Transactions on Communications*, vol. 43, no. 2–4, pp. 1566–1579, 1995.

[29] O. Rose, "Simple and Efficient Models for Variable Bit Rate MPEG Video Traffic," *Performance Evaluation*, vol. 30, no. 1–2, pp. 69–85, 1997.

[30] M. Krunz and A. M. Makowski, "A Source Model for VBR Video Traffic Based on $M/G/\infty$ Input Processes," in *IEEE Infocom*, 1998.

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

[32] ITU-T, "G.711: Pulse Code Modulation (PCM) of Voice Frequencies."

[33] ITU-T, "G.729: Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)."

[34] CounterPath Solutions, Inc., "X-Lite 3.0." `http://www.xten.com/`.

[35] "Skype." `http://www.skype.com`.

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

[37] M. Vitez, "Picophone." `http://www.vitez.it/picophone/`.

[38] S. Andersen, A. Duric, H. Astrom, R. Hagen, W. Kleijn, and J. Linden, "RFC3951: Internet Low Bit Rate Codec (iLBC)," Dec. 2004.

[39] 3rd Generation Partnership Project (3GPP), "GSM AMR Speech Codec." `http://www.3gpp.org/ftp/Specs/archive/26_series/26.073/`.

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

[41] Michael Menth, Andreas Binzenhöfer and Stefan Mühleck, "Transition Matrices and Parameters of the MMCs Modelling VBR Traffic." `http://www3.informatik.uni-wuerzburg.de/TR/mmc`.