# Improvements of Hidden Chunk Models

*Harald Höge*[1] *and Panji Setiawan*[2]

[1]*Universität der Bundeswehr München, Munich, Germany*
[2]*Siemens Enterprise Communications GmbH & Co. KG, Munich, Germany*

**Abstract:** The statistical properties of segments [8] using a specific acoustic model called the hidden chunk model (HCM) is investigated. We call the sequence of feature vectors assigned to a segment a *chunk* of length $\ell$. The HCM still assumes that the feature vectors are statistically independent. In contrast to hidden Markov model (HMM) we introduce emission probabilities which depend on $\ell$. Segment error rates (SERs) are calculated on a database with over 33 million chunks aligned to 607 segments. The HCM achieves more than 10 % absolute improvement in SER compared to the HMM. Based on the estimated Shannon's entropy, the proposed HCM model paves the way to create acoustic models which are heading towards the lowest possible SER.

## 1 Introduction

Based on the pattern recognition theory [1], the minimum error rate is achieved by the principle of maximum likelihood

$$\widehat{W}^* = \arg\max_{W^*} P(W^*|\mathbf{x}^T) = \arg\max_{W^*} \frac{p(\mathbf{x}^T|W^*)P(W^*)}{p(\mathbf{x}^T)},$$

where $W^*$ denotes a sequence of words which is the symbolic representation of an utterance and $\mathbf{x}^T = \{\mathbf{x}(1), \cdots, \mathbf{x}(t), \cdots, \mathbf{x}(T)\}$ denotes a sequence of feature vectors which is the acoustic representation of an utterance. Advances in speech recognition depend to a great extent on the improvements in: the quality of the feature vectors $\mathbf{x}$, language modeling $P(W^*)$, and acoustic modeling $p(\mathbf{x}^T|W^*)$. This paper is focused on an acoustic model defined by the hidden chunk model (HCM)

$$p^{HCM}(\mathbf{x}^T|Q^N) = \prod_{n=1}^{N} b_{\ell(n-1),\ell(n)} p(\mathbf{x}^\ell(n)|Q(n)),$$

where the *chunk* $\mathbf{x}^\ell(n) = \{\mathbf{x}(1,n), \cdots, \mathbf{x}(\ell,n)\}$ consists of a sequence of feature vectors with a random length $\ell$ aligned to a segment $Q(n)$ as illustrated in Figure 1. $Q(n)$ takes a value from the set of $N_Q$ segments $\mathbb{Q} = \{Q_i\}_{i=1}^{N_Q}$ containing all possible segment outcomes. Consequently, $Q^N = \{Q(1), \cdots, Q(n), \cdots, Q(N)\}$ denotes a sequence of segments with $N << T$. The term $b_{j,k}$ denotes the chunk length transition probability and $\ell(n)$ assigns the chunk length $\ell = 1 \ldots L$ at time instant $n$. The term $p(\mathbf{x}^\ell(n)|Q(n))$ or simply $p(\mathbf{x}^\ell|Q_i)$ denotes the chunk emission probability as highlighted in [8].

The segment $Q_i$ denotes the symbolic representation of a sound. In HMM technology, a segment is a *state* and a phoneme is defined by a *triphone*. This implies that a segment is the symbolic representation of the sounds produced at the beginning, middle, or end of a phoneme. Furthermore, an utterance $W^*$ is defined in the symbolic domain by a sequence of segments $Q^N$ and in the acoustic domain by a sequence of chunks $\{\mathbf{x}^\ell(1), \cdots, \mathbf{x}^\ell(n), \cdots, \mathbf{x}^\ell(N)\}$ constituting the sequence of feature vectors.

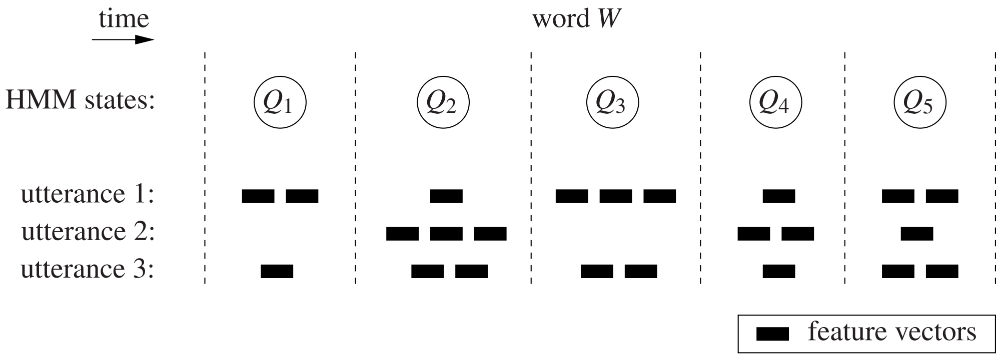



**Figure 1** - Alignment of the feature vectors to the segments $Q_1, \cdots, Q_5$ for three utterances of the same word $W$.

Assuming the feature vectors composing a chunk are statistically independent, we get the following approximation for the chunk probability:

$$p(\mathbf{x}^\ell|Q_i) \approx \prod_{v=1}^{\ell} p(\mathbf{x}(v,\ell)|Q_i). \tag{1}$$

The extended emission probability $p(\mathbf{x}(v,\ell)|Q_i)$ depends on the length $\ell$ and on the position $v = 1 \ldots \ell$ of the feature vector within a chunk. We denote the probability as the HCM-adapted emission probability (HCM-EP) modeled with a Gaussian mixture as done for continuous Gaussian HMMs. The relation between the chunk and the feature vector is thus defined as

$$\mathbf{x}^\ell = \{\mathbf{x}(v,\ell)\}_{v=1}^{\ell},$$

where the time index $n$ has been dropped for notational simplicity.

The performance of the proposed model is determined by the segment error rates $\mathrm{SER}_\ell$ and $\mathrm{SER}_{v,\ell}$ which are evaluated for segments represented by chunks for each length $\ell$ and for feature vectors represented by the position and length $(v,\ell)$, respectively. Both error rates are defined as

$$\mathrm{SER}_\ell = \int_{\mathbb{X}^\ell} \Big[ 1 - \max_{Q_i} \Big( p(\mathbf{x}^\ell|Q_i)\, P_\ell(Q_i) \Big) \Big]\, d\mathbf{x}^\ell, \tag{2}$$

$$\mathrm{SER}_{v,\ell} = \int_{\mathbb{X}(v,\ell)} [\, 1 - \max_{Q_i} \Big( p(\mathbf{x}(v,\ell)|Q_i)\, P_\ell(Q_i) \Big) \,]\, d\mathbf{x}(v,\ell), \tag{3}$$

where $\mathbb{X}^\ell$ denotes the set of all chunks of length $\ell$ and $\mathbb{X}(v,\ell)$ denotes the set of all feature vectors having the position $v$ in a chunk of length $\ell$. $P_\ell(Q_i)$ denotes the probability of occurrence of segment $Q_i$ on the set $\mathbb{X}^\ell$ and $\max_{Q_i}\ p(\cdot|Q_i)P_\ell(Q_i)$ is the maximum *a posteriori* (MAP) estimator. An analysis utilizing the Shannon's conditional entropies [11], i.e., $H(Q|\mathbf{X}^\ell)$ and $H(Q|\mathbf{X}(v,\ell))$, is also given which is closely related to the $\mathrm{SER}_\ell$ and $\mathrm{SER}_{v,\ell}$, respectively.

The remainder of this paper is organized as follows. The statistical framework to evaluate the proposed chunk model is described in Section 2. It merely concerns the way the entropy values are estimated. Section 3 presents the experimental framework and results followed by the Conclusion in the last section.

## 2 The Statistical Framework

### 2.1 Expectations of Functions

In the following we assume that the chunk $\mathbf{X}^\ell = \mathbf{x}^\ell$ and feature vector $\mathbf{X}(v,\ell) = \mathbf{x}$ represent ergodic random variables [9]. Given a function $F(\cdot)$ which operates on both variables, we can define the corresponding expectation operator $\mathbb{E}[\cdot]$. Based on the ergodic property of the feature vectors, the expectations of functions $F(\mathbf{x}^\ell)$ and $F(\mathbf{x})$ can be derived as

$$\begin{aligned}
\mathbb{E}[F(\mathbf{x}^\ell)] &\triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} F(\mathbf{x}^\ell(n)); \quad \mathbf{x}^\ell(n) \in \mathbb{X}^\ell, \\
\mathbb{E}[F(\mathbf{x})] &\triangleq \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} F(\mathbf{x}(t)); \quad \mathbf{x}(t) \in \mathbb{X}(v,\ell).
\end{aligned}$$

The above method needs infinitely large sets, in this paper however, we are using a restricted size of the sets formulated as

$$\begin{aligned}
\mathbb{E}[F(\mathbf{x}^\ell)] &= \frac{1}{|\mathbb{X}^\ell|} \sum_{n=1}^{|\mathbb{X}^\ell|} F(\mathbf{x}^\ell(n)); \quad \mathbf{x}^\ell(n) \in \mathbb{X}^\ell, &(4)\\
\mathbb{E}[F(\mathbf{x})] &= \frac{1}{|\mathbb{X}(v,\ell)|} \sum_{t=1}^{|\mathbb{X}(v,\ell)|} F(\mathbf{x}(t)); \quad \mathbf{x}(t) \in \mathbb{X}(v,\ell), &(5)
\end{aligned}$$

where $|\cdot|$ denotes the cardinality of the set. This method is called the stochastic sampling or the Monte Carlo method.

### 2.2 Entropies

To evaluate the Shannon's conditional entropies $H(Q|\mathbf{X}^\ell)$ and $H(Q|\mathbf{X}(v,\ell))$, the relations

$$\begin{aligned}
H(Q|\mathbf{X}^\ell) &= H_\ell(Q) - I(\mathbf{X}^\ell;Q), &(6)\\
H(Q|\mathbf{X}(v,\ell)) &= H_{v,\ell}(Q) - I(\mathbf{X}(v,\ell);Q), &(7)
\end{aligned}$$

are used [9], where

$$\begin{aligned}
I(\mathbf{X}^\ell;Q) &= H(\mathbf{X}^\ell) - H(\mathbf{X}^\ell|Q), &(8)\\
I(\mathbf{X}(v,\ell);Q) &= H(\mathbf{X}(v,\ell)) - H(\mathbf{X}(v,\ell)|Q). &(9)
\end{aligned}$$

The terms $I(\mathbf{X}^\ell;Q)$ and $I(\mathbf{X}(v,\ell);Q)$ are called the mutual information (MI) between the segments and the assigned chunks or the position dependent feature vectors, respectively. The value of MI determines the information extracted from the chunks $\mathbf{X}^\ell$ and from the feature vectors $\mathbf{X}(v,\ell)$ for recognizing the segments $Q_i$. The segment entropy $H_\ell(Q)$ denotes, how much information is needed to decode the segments $Q_i$. Thus, Shannon's entropies reveal how much information - measured in [*bit*] - is missing to recognize the segments without errors.

The segment entropies $H_\ell(Q)$ and $H_{v,\ell}(Q)$ in (6) and (7), respectively, are identical since the probability of occurrence of segment $Q_i$ is exactly the same on the sets $\mathbb{X}^\ell$ and $\mathbb{X}(v,\ell)$, i.e., $P_\ell(Q_i) = P_{v,\ell}(Q_i)$. They are calculated as

$$H_\ell(Q) = -\sum_{i=1}^{N_Q} P_\ell(Q_i)\,\mathrm{ld}\,P_\ell(Q_i). \tag{10}$$

The statistical dependency between adjacent segments is not modeled. The entropies in (8) and (9) used to determine the mutual information are defined as follows:

$$
\begin{aligned}
H(\mathbf{X}^\ell) &\triangleq -\int_{\mathbb{X}^\ell} p(\mathbf{x}^\ell)\,\mathrm{ld}\,p(\mathbf{x}^\ell)\; d\mathbf{x}^\ell, \\
H(\mathbf{X}^\ell|Q) &\triangleq -\sum_{i=1}^{N_Q} \int_{\mathbb{X}^\ell|Q_i} p(\mathbf{x}^\ell, Q_i)\,\mathrm{ld}\,p(\mathbf{x}^\ell|Q_i)\; d\mathbf{x}^\ell \\
&= -\sum_{i=1}^{N_Q} P_\ell(Q_i) \int_{\mathbb{X}^\ell|Q_i} p(\mathbf{x}^\ell|Q_i)\,\mathrm{ld}\,p(\mathbf{x}^\ell|Q_i)\; d\mathbf{x}^\ell, \\
H(\mathbf{X}(v,\ell)) &\triangleq -\int_{\mathbb{X}(v,\ell)} p(\mathbf{x}(v,\ell))\,\mathrm{ld}\,p(\mathbf{x}(v,\ell))\; d\mathbf{x}(v,\ell), \\
p(\mathbf{x}(v,\ell)) &\triangleq \sum_{i=1}^{N_Q} P_\ell(Q_i) \int_{\mathbb{X}(v,\ell)|Q_i} p(\mathbf{x}(v,\ell)|Q_i)\; d\mathbf{x}(v,\ell), \\
H(\mathbf{X}(v,\ell)|Q) &\triangleq -\sum_{i=1}^{N_Q} \int_{\mathbb{X}(v,\ell)|Q_i} p(\mathbf{x}(v,\ell), Q_i)\,\mathrm{ld}\,p(\mathbf{x}(v,\ell)|Q_i)\; d\mathbf{x}(v,\ell) \\
&= -\sum_{i=1}^{N_Q} P_\ell(Q_i) \int_{\mathbb{X}(v,\ell)|Q_i} p(\mathbf{x}(v,\ell)|Q_i)\,\mathrm{ld}\,p(\mathbf{x}(v,\ell)|Q_i)\; d\mathbf{x}(v,\ell),
\end{aligned}
\tag{11}
$$

where the sets $\mathbb{X}^\ell|Q_i$ and $\mathbb{X}(v,\ell)|Q_i$ denote the sets of chunks $\mathbb{X}^\ell$ and feature vectors $\mathbb{X}(v,\ell)$, respectively, assigned to a particular segment $Q_i$. The chunk probability approximation using the HCM-EP as in (1) can be used to evaluate (11). The HCM-EPs are modeled by multimodal Gaussians using a globally pooled covariance matrix. Using an LDA in the feature analysis a globally pooled covariance matrix has proven to give good results [4].

## 2.3 Scores Related to Entropies

As shown in Section 2.2 several entropy terms have to be calculated to finally obtain the mutual informations in (8) and (9). An estimate to the entropy term is proposed employing a scoring method. Let's define a score function $\theta(\cdot) = -\ln p(\cdot)$ yielding

$$\theta(\mathbf{x}^\ell|Q_i) \triangleq -\ln p(\mathbf{x}^\ell|Q_i), \tag{12}$$

$$\theta(\mathbf{x}(v,\ell)|Q_i) \triangleq -\ln p(\mathbf{x}(v,\ell)|Q_i). \tag{13}$$

It can be shown easily that the following relations hold:

$$
\begin{aligned}
E[\theta(\mathbf{x}^\ell|Q_i)] &= \ln 2\, H(\mathbf{X}^\ell|Q_i), \\
E[\theta(\mathbf{x}(v,\ell)|Q_i)] &= \ln 2\, H(\mathbf{X}(v,\ell)|Q_i).
\end{aligned}
$$

Using the stochastic sampling as shown in (4) and (5) we get

$$
\begin{aligned}
\widetilde{H}(\mathbf{X}^\ell|Q_i) &= \frac{1}{\ln 2 \cdot \left|\mathbb{X}^\ell|Q_i\right|} \sum_{n=1}^{\left|\mathbb{X}^\ell|Q_i\right|} \theta(\mathbf{x}^\ell(n)|Q_i), \\
\widetilde{H}(\mathbf{X}^\ell|Q) &= \sum_{i=1}^{N_Q} P_\ell(Q_i)\widetilde{H}(\mathbf{X}^\ell|Q_i),
\end{aligned}
\tag{14}
$$

$$\widetilde{H}(\mathbf{X}(v,\ell)|Q_i) = \frac{1}{\ln 2 \cdot \left|\mathbb{X}(v,\ell)|Q_i\right|} \sum_{t=1}^{\left|\mathbb{X}(v,\ell)|Q_i\right|} \theta(\mathbf{x}(v,\ell,t)|Q_i),$$

$$\widetilde{H}(\mathbf{X}(v,\ell)|Q) = \sum_{i=1}^{N_Q} P_\ell(Q_i)\widetilde{H}(\mathbf{X}(v,\ell)|Q_i).$$

The terms $\widetilde{H}(\mathbf{X}^\ell)$ and $\widetilde{H}(\mathbf{X}(v,\ell))$ can be estimated using the same method employing

$$\theta(\mathbf{x}^\ell) = -\ln p(\mathbf{x}^\ell) \quad \text{and} \quad \theta(\mathbf{x}(v,\ell)) = -\ln p(\mathbf{x}(v,\ell)). \tag{15}$$

### 2.4 Bounds of Error Rates

Based on the following relation in [9]:

$$-\int_{\mathbb{X}} p(x)\,\mathrm{ld}\,p(x)dx \leq -\int_{\mathbb{X}} p(x)\,\mathrm{ld}\,f(x)dx,$$

which holds for any probability function $f(x)$ constrained to $f(x) \geq 0$ and $\int_{\mathbb{X}} f(x)dx = 1$, yields

$$H(\mathbf{X}^\ell|Q_i) \leq \widetilde{H}(\mathbf{X}^\ell|Q_i) \quad \text{and} \quad H(Q|\mathbf{X}^\ell) \leq \widetilde{H}(Q|\mathbf{X}^\ell).$$

Thus better modeling of the chunk probabilities and of the HCM-EPs leads to better estimates of the entropies until the lower limit given by the *real* entropy terms is reached. Certainly, this also holds for the average values of the scores as shown in Section 2.3. It is well known in HMM technology that better *fitting* scores lead to lower error rates. Given the Shannon's entropy values, the lower and upper bounds of the error rates are given by the Fano [2] and Golić [3] bounds, respectively.

## 3 Experiments and Results

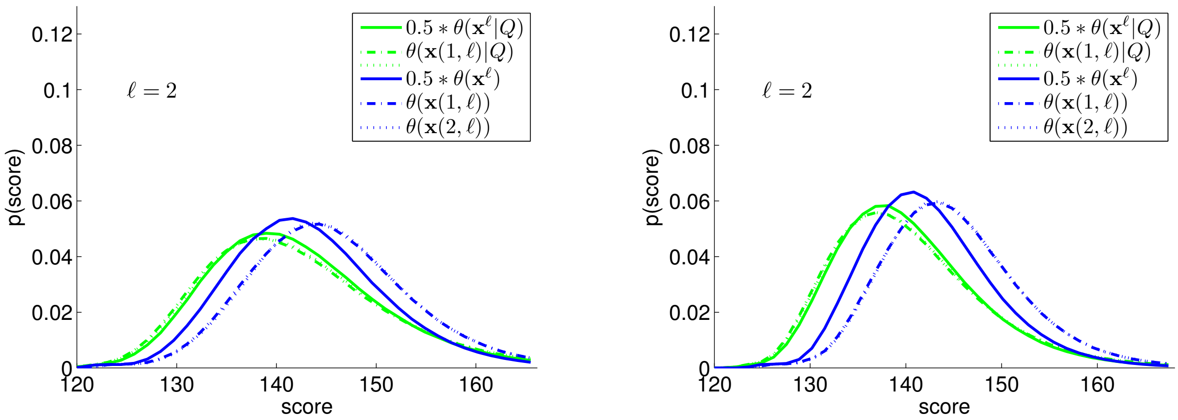
### 3.1 Experimental Set Up

The feature vectors are derived in a frame rate of 15 *ms* from in-car recorded speech sampled at 11.025 *kHz* [7]. To improve the recognition performance in noisy environments, noise reduction techniques have been applied [10]. From the noise reduced spectra, MFCCs are derived and extended to super vectors, which are transformed by an LDA [4] to obtain the final feature vectors. The basic design parameters of the initial HMMs used to classify the segments are shown in Table 1.

**Table 1** - Design parameters of the HMMs.

| # Gaussians | $D$ Dimension of feature vector | $N_Q$ Number of segments |
|---|---|---|
| 20 000 | 24 | 607 |

The same design parameters including the amount of Gaussians used for each segment were copied to model the HCM-EPs. The HCMs training for a specific chunk $\ell$ is done following the EM-algorithm in [6] taking into account all feature vectors in the set $\mathbb{X}^\ell$. We regard the chunks of length $\ell = 1,2,3$. The chunks with $\ell > 3$, which are rather rare except for non-speech segments, have been truncated to the first three feature vectors appearing in the sequence and are added to the chunks with length three. The values of the probabilities $P(\ell)$ and the total number of chunks are listed in Table 2.




(a) Original HMMs with $|\mathbb{X}^\ell| = 5\ 089\ 251$.

(b) HCM-EPs with $|\mathbb{X}^\ell| = 7\ 871\ 342$.

**Figure 2** - The distribution of scores for HMMs and HCM-EPs for $\ell = 2$.

**Table 2** - Probabilities $P(\ell)$ and the total number of chunks.

| $P(\ell=1)$ | $P(\ell=2)$ | $P(\ell=3)$ | $\sum_{\ell=1}^{3} \lvert\mathbb{X}^\ell\rvert$ |
|---|---|---|---|
| 0.29 | 0.46 | 0.25 | 33 879 857 |

The entropies $H_\ell(Q)$ as in (10) are given in Table 3. For equally distributed segments ($P_\ell(Q_i) = 1/N_Q$) the entropy would take the value of 9.25 *bits*. According to Table 3, the segments are not equally distributed and the distributions depend on $\ell$.

## 3.2 Distributions of Scores

In this section we show the distributions of the scores as described in Section 2.3. The distributions are determined first for each segment $Q_i$ taking the feature vectors from the sets $\mathbb{X}^\ell$ and $\mathbb{X}(v,\ell)$. In the second step the distribution of all segments is determined by averaging the segment specific distributions with the weight defined as the probability of each segment $P_\ell(Q_i)$. These distributions are the basis to estimate the entropies.

For chunks of length $\ell = 2$, Figure 2(a) shows the distributions of the scores based on the initial HMMs. In Figure 2(b) the scores of HCM-EPs depending on $\ell$ are plotted. In both figures the green scores denoted with $\theta(\mathbf{x}^\ell|Q)$ correspond to the scores $\theta(\mathbf{x}^\ell|Q_i)$ in (12) averaged over all segments, whereas the scores $\theta(\mathbf{x}(v,\ell)|Q)$ for $v = 1,2$ correspond to the averaged scores of $\theta(\mathbf{x}(v,\ell)|Q_i)$ in (13). The blue scores correspond to $\theta(\mathbf{x}^\ell)$ and $\theta(\mathbf{x}(v,\ell))$ as defined in (15). The distributions in Figures 2(a) and 2(b) show that the HCM-EPs have lower scores. Furthermore, the distributions of the scores for different $v$ are quite similar hinting that the distributions HCM-EPs for different position $v$ might be quite similar.

**Table 3** - Segment entropies $H_\ell(Q)$ $[bits]$ of chunks with length $\ell$.

| $H_1(Q)$ | $H_2(Q)$ | $H_3(Q)$ |
|---|---|---|
| 7.78 *bits* | 8.35 *bits* | 7.37 *bits* |

### 3.3 Entropies and Error Rates

In the following, the SERs given in (2) and (3) and the approximation of Shannon's entropies using the method of stochastic sampling according to Section 2.3 are presented. Tables 4 and 5 show that the HCM-EP leads to lower error rates compared to the original HMM. Furthermore, the SERs decrease with increasing length of chunks. The error rates are quite similar for feature vectors with different positions $v$ within chunks of the same length. How close the approximated values of the mutual information are to their real values is still an open issue.

**Table 4** - Segment error rates and mutual informations derived from feature vectors.

| $\mathbf{X}(v,\ell)$ | | $\lvert\mathbb{X}(v,\ell)\rvert$ | $\mathbf{SER}_{v,\ell}$ [%] HCM-EP / HMM | $\widetilde{I}(\mathbf{X}(v,\ell);Q)$ [$bits$] HCM-EP / HMM |
|---|---|---|---|---|
| $v=1$ | $\ell=1$ | 3 683 738 | 65.9 / 81.1 | 3.1 / -0.7 |
| $v=1$ | $\ell=2$ | 5 109 251 | 58.1 / 68.2 | 4.8 / 3.4 |
| $v=2$ | | | 58.2 / 68.4 | 4.8 / 3.3 |
| $v=1$ | $\ell=3$ | 3 464 567 | 41.2 / 51.0 | 4.7 / 3.8 |
| $v=2$ | | | 38.2 / 46.0 | 4.9 / 4.2 |
| $v=3$ | | | 44.0 / 48.3 | 4.5 / 4.0 |

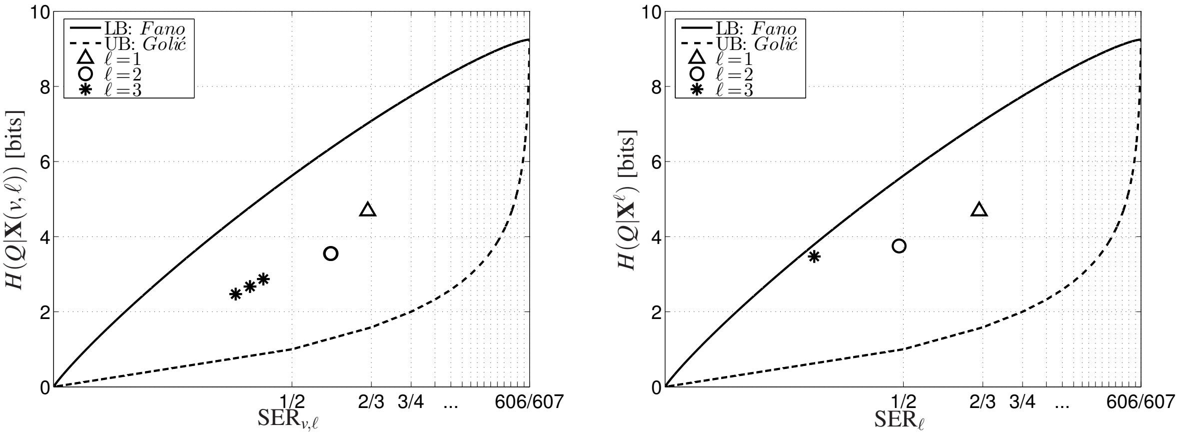
**Table 5** - Segment error rates and mutual informations derived from chunks.

| $\mathbf{X}^\ell$ | $\lvert\mathbb{X}^\ell\rvert$ | $\mathbf{SER}_\ell$ [%] HCM-EP / HMM | $\widetilde{I}(\mathbf{X}^\ell;Q)$ [$bits$] HCM-EP / HMM |
|---|---|---|---|
| $\ell=1$ | 3 683 738 | 65.9 / 81.1 | 3.1 / -0.7 |
| $\ell=2$ | 5 109 251 | 49.1 / 61.1 | 4.6 / 2.1 |
| $\ell=3$ | 3 464 567 | 31.3 / 41.6 | 3.9 / 2.7 |

Figure 3 shows the Fano and Golić bounds together with the SERs related to the approximated Shannon's entropy. Especially for $\ell=3$, it seems that the approximation is quite inaccurate due to the assumption of statistical independence of adjacent feature vectors within a chunk. Note that the values in $y$-axis are calculated following (6) and (7) by taking the entropy and mutual information values in Tables 3, 4, and 5.

## 4 Conclusions

The use of chunks leads to new acoustical models of segments, where the length of the chunks assigned to a segment is regarded as an intrinsic property of a segment. This is in contrast to HMM technology, where the length is treated as a duration effect modeled by repeating a state (segment) with the same statistical properties. The HMM and the HCM approach assumes that the feature vectors are statistically independent. The analysis of entropy shows that this assumption is wrong especially for chunks with $\ell=3$. As the chunks have a fixed and treatable dimension, chunk probabilities modeling statistical dependencies could be derived [5]. Those extended HCMs could reach the theoretical lower limit in SER. Still the statistical dependency between chunks is an open issue for recognition tasks where larger phonetic units than segments have to be recognized.




(a) $H(Q|\mathbf{X}(v,\ell))$. (b) $H(Q|\mathbf{X}^{\ell})$.

**Figure 3** - Relation between SER and Shannon's Entropy for HCM-EP model.

# References


[1] DUDA, R. O. and P. E. HART: *Pattern Classification and Scene Analysis*. John Wiley & Sons, June 1973.

[2] FANO, R. M.: *Transmission of Information: A Statistical Theory of Communications*. The MIT Press and John Wiley & Sons, Inc. New York, London, 1961.

[3] GOLIĆ, J.: *On the Relationship Between the Information Measures and the Bayes Probability of Error*. IEEE Transactions on Information Theory, IT-33(5):681–693, September 1987.

[4] HAUENSTEIN, A. and E. MARSCHALL: *Methods for Improved Speech Recognition over Telephone Lines*. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 425–428, May 1995.

[5] HÖGE, H.: *Second-Order Models for Hidden Chunk Models*. In *Proc. ITG Fachtagung Speech Communication*, October 2010.

[6] HUANG, X. D., Y. ARIKI and M. A. JACK: *Hidden Markov Models for Speech Recognition*. Edinburgh University Press, 1990.

[7] MORENO, A., B. LINDBERG, C. DRAXLER, G. RICHARD, K. CHOUKRI, J. ALLEN and S. EULER: *SpeechDat-Car: A Large Speech Database for Automotive Environments*. In *Proc. International Conference on Language Resources and Evaluation (LREC)*, June 2000.

[8] OSTENDORF, M., V. V. DIGALAKIS and O. A. KIMBALL: *From HMM's to Segment Models: A Unified View of Stochastic Modeling for Speech Recognition*. IEEE Transactions on Speech and Audio Processing, 4(5):360–378, September 1996.

[9] PAPOULIS, A.: *Probability, Random Variables, and Stochastic Processes, Third Edition*. McGraw-Hill, Inc., 1991.

[10] SETIAWAN, P., C. BEAUGEANT, T. FINGSCHEIDT and S. STAN: *Least-Squares Weighting Rules Formulations in the Frequency Domain*. In *Proc. Electronic Speech Signal Processing Conference (ESSP)*, September 2005.

[11] SHANNON, C. E.: *A Mathematical Theory of Communication*. Bell System Technical Journal, 27:379–423 and 623–656, July and October 1948.