ON THE MUTUAL INFORMATION BETWEEN FREQUENCY BANDS IN SPEECH

Mattias Nilsson, Søren Vang Andersen and W. Bastiaan Kleijn

Department of Speech, Music and Hearing
KTH (Royal Institute of Technology)
100 44 Stockholm, Sweden

ABSTRACT

In this paper we investigate the mutual information in speech between the spectral envelope of the high frequency band and low frequency bands of various widths. Direct methods on the computation of the mutual information often result in an excessive amount of data required even for modest situations. We reduce the required amount of data by quantizing the low band leading to a lower bound expression on the mutual information. We indicate by simulation that this lower bound is in the same order of magnitude as the true mutual information. Simulations on speech show that we have no less than 0.1 bit of shared information between the slope of the high band and the low frequency band from 0 - 4 kHz. Performing the analogous simulation with the gain of the high band we obtained no less than 0.45 bit of mutual information.

1. INTRODUCTION

In recent years, there have been a significant number of publications on bandwidth expansion of speech signals [1, 2, 3, 4]. The bandwidth expansion algorithms are generally aimed at recovering the spectral envelope for frequencies up to 8 kHz, given a speech signal with frequency contents below 3.6 kHz. This processing removes the muffled sound quality, which is introduced when the speech signal is band limited to 4 kHz.

The motivation for all bandwidth expansion methods is the fact that the spectral envelope of the lower and higher frequency bands of the speech signal are dependent, i.e., the low band part of the speech spectrum provides information about the spectral shape of the high band part. This results from speech being created by a physical source.

If the logarithmic spectral energies of frequency bands had a Gaussian distribution, their relation could be described with a correlation function. However, it is well known [5] that the logarithmic spectral energies of the speech signal are non-Gaussian and thus have statistical moments of order higher than two, which would not be accounted for with a correlation measure. In this situation, mutual information is an appropriate measure.

Mutual information as a measure of dependency has been used in connection with automatic speech recognition to study the distribution in time and frequency of information relevant for phonetic classification [5] and in estimating the information contained in the so-called feature-vector joint distribution [6]. By observing only regions having densely informative content, it is possible to reduce the size of the data set used for training maximum-likelihood based speech recognition systems.

This work is a first attempt to investigate the amount of information that is shared between the low and high band in speech. The objective is to determine an approximate value of the mutual information between the high band and various widths of the low band of spectral envelopes of speech. We have in this paper only considered the mutual information between spectral envelopes. Our results should provide information on whether there is a frequency region in the low band, that contains almost all information about the high band. If there is, we could claim that speech coders coding the high band independently of the low band are wasting bits in representing something which is predictable from the low frequency band. In this work we are using only the slope respectively the gain of the high band spectral envelope to capture the behavior of the high band.

The slope of the high band conveys partial information on whether a speech segment is voiced (v) or unvoiced (uv). A v/uv decision can also be made from a low frequency band 0 - 4 kHz. This suggest that the low band contains information about the high band slope in the same order of magnitude as the entropy of a v/uv classification. If we assume 80 % of the speech to be voiced, this results in an entropy of 0.72 bit. If the slope contained full information about the v/uv classification, 0.72 bit would be a lower bound on the mutual information between the slope and the low band. However, we do not suggest that this is the case since only partial information on a voiced/unvoiced classification is conveyed by the slope.

Looking at the LPC spectrum at the high frequency band we can see that there are some peaks and valleys, which will not be captured using only the slope and gain. However, this work is a first step in the direction of finding the true mutual information between the low and high frequency bands.

This paper is organized as follows. In section 2, we derive a lower bound on the mutual information, which requires a smaller amount of data compared to direct calculations of the mutual information. Simulation procedures and results for both synthetic and speech data are presented in section 3. In section 4 conclusions from this work are drawn.

2. MUTUAL INFORMATION

The mutual information between two continuous variables X and Y is given by [7]:

$$I(X;Y) = h(Y) - h(Y|X), \tag{1}$$

where $h(Y)$ is the differential entropy of Y and is defined by an integration over the value space Ω_Y of Y:

$$h(Y) = -\int_{\Omega_Y} f_Y(y) \log_2(f_Y(y)) dy \tag{2}$$

with $f_Y(y)$ being the probability density function (*pdf*) of Y. The conditional differential entropy $h(Y|X)$ of Y given X is defined as:

$$h(Y|X) = -\int_{\Omega_X} \int_{\Omega_Y} f_{Y,X}(y,x) \log_2(f_{Y|X}(y|x)) dy dx, \tag{3}$$

where Ω_X is the value space of X and $f_{Y,X}(y,x)$ is the joint *pdf* of X and Y. Throughout this paper, X is a coefficient vector representing the low frequency band and Y is a coefficient scalar or vector representing the high frequency band.

2.1. Reduction of the required amount of data

In practice we do not have access to either the true value spaces Ω_X and Ω_Y or to the true *pdf*s. Thus, the differential entropies in (2) and (3) have to be estimated from the observed data. Estimation of these differential entropies is problematic. We would like Y to be the log amplitude of the spectrum at m frequency bins covering the high band frequencies, $Y = \{Y_1, Y_2, \cdots, Y_m\}$. This requires an estimate of the joint *pdf* $f_{Y_1,Y_2,\cdots,Y_m}(y_1, y_2, \cdots, y_m)$ to compute the differential entropy. However, the determination of this joint *pdf* with sufficient accuracy demands an extremely large amount of data, even for modest values of m.

To reduce the amount of data required and to make the computations tractable, we describe the high band with two parameters: the slope S, and the gain G. The slope and gain are both derived from a first order model of the high frequency band. This allows estimation of the differential entropies $h(S)$, $h(G)$, $h(S|X)$ and $h(G|X)$. In our work, the low band representation X consists of 32 mel-frequency cepstral coefficients (MFCC).

To further reduce the amount of data required, we constrain X to be a fixed number of possible low band spectral envelope representations, i.e., we quantize our low band X with a fixed size codebook. When we quantize the low band, we tile the space of X into a fixed number of regions. The random index of these regions is denoted by K_X. We used a vector quantizer to cluster the MFCCs representing the low band using the Generalized Lloyd Algorithm (GLA) [8]. For every set of MFCCs mapped to a certain codebook entry, the slope and gain are calculated and averaged. This results in approximate MMSE-estimators of S and G, given K_X:

$$\hat{S}(K_X) = \hat{E}[S|K_X], \tag{4}$$

$$\hat{G}(K_X) = \hat{E}[G|K_X], \tag{5}$$

respectively. Here, $\hat{E}$ denotes the approximation of the true expectation operator E by the sample mean. The estimator is in practice a codebook look-up. In the remainder of this section, we describe the processing concerning the slope S, since the processing for the gain G is completely analogous.

By using a quantizer for the low band we obtain an upper bound for the conditional differential entropy term in (3):

$$h(S|X) \leq h(S|K_X), \tag{6}$$

where the inequality results from the fact that the quantized low band K_X provides equal or less information about S than the original low band X.

From our MMSE estimate $\hat{S}$ of S we now form the estimation error:

$$N_S = S - \hat{S}. \tag{7}$$

Inserting (7) in (6) yields,

$$h(S|K_X) = h(\hat{S}(K_X) + N_S|K_X), \tag{8}$$

which, since $\hat{S}(K_X)$ is a deterministic function of K_X, can be reduced to

$$h(S|K_X) = h(N_S|K_X). \tag{9}$$

We can now write (9) as an upper bound on the conditional differential entropy:

$$h(S|K_X) \leq h(N_S) \tag{10}$$

with equality when the indexing process K_X is independent of the estimation noise N_S.

Finally, from equations (1) and (10) we have a lower bound on the mutual information between the spectral envelope of the low band and slope of the high band:

$$I(X;S) \geq h(S) - h(N_S). \tag{11}$$

This lower bound requires only the determination of the differential entropies of the slope and the slope estimation noise.

2.2. Estimation of differential entropy via histogram

From a histogram of S we can approximate the *pdf* of the slope with a probability mass function (*pmf*). If we divide the range of the random variable S into bins of length Δ_S and denote the random index K_S, we can then approximate the differential entropy as [7]:

$$h(S) \approx H(K_S) + \log_2(\Delta_S), \tag{12}$$

where $H(K_S)$ is the entropy of K_S. This method applies to the estimation noise N_S as well.

3. SIMULATIONS AND RESULTS

This section describes simulations on both synthetic and speech data. The results are presented at the end of the section.

3.1. Simulation with synthetic data

To investigate the closeness of the lower bound (11) to the true value of the mutual information, we performed a simulation on synthetic data with known mutual information. Two zero mean unit variance Gaussian distributed processes S and D were constructed. We then formed a new process $S_D = S + D$ for which the mutual information between S_D and S can be determined analytically [7]:

$$I(S_D; S) = \frac{1}{2}\log_2(\frac{\sigma^2_{S_D}}{\sigma^2_D}) = \frac{1}{2}. \quad (13)$$

For the simulations we constructed 200000 realizations of the scalar processes S, D and S_D yielding the data sets $\{s\}$, $\{d\}$ and $\{s_d\}$, respectively. A synthetic training set $\{x\}$ was then constructed: for every element s_d in $\{s_d\}$ we formed a vector of dimension 10 consisting of s_d and a zero mean Gaussian vector of length 9 samples, each with variance two. Vector quantization was then performed on the training set $\{x\}$ and an approximate MMSE-estimator was calculated from the data set $\{s\}$ as described in section 2.1. The codebook size used was 1024. We then used the codebook to extract an estimate $\hat{s}$ of the true s and determine the estimation noise $n = s - \hat{s}$. All estimation noise samples formed the data set $\{n\}$. The differential entropies $h(S)$ and $h(N)$ were then calculated from the data sets $\{s\}$ and $\{n\}$, respectively, by means of histograms as described in section 2.2. The lower bound of the mutual information was then calculated using (11).

3.2. Simulation with speech data

Our data set consisted of 2200 speech files (sampled at 16 kHz) from the TIMIT data base, yielding 600000 segments of length 20 ms using 50 % overlap. The Log-Area-Ratio (LAR) was used to represent the slope parameter:

$$s = \log_{10}(\frac{1-l}{1+l}), \quad (14)$$

where l is the first reflection coefficient. The reflection coefficient was calculated from a first order LPC analysis. From the same analysis the amplification b of the LPC filter $A(z)$ was determined assuming a unit variance input:

$$A(z) = \frac{b}{1 - lz^{-1}}. \quad (15)$$

We then used the logarithm of b as our gain parameter:

$$g = \log_{10}(b). \quad (16)$$

From each speech file, high band and low band speech files were created. The low band speech file contained frequencies up to P kHz, where P ranged from 1 to 4 kHz. The high band speech file was created by first high-pass filtering the speech signal at a cut-off frequency of 4 kHz. The high-pass filtered signal was then modulated with a cosine to move the signal to the band 0 - 4 kHz, low-pass filtered at 4 kHz and finally down-sampled by a factor 2. For each speech segment the slope and the approximate MMSE estimate of the slope were found and the estimation noise was determined.

entity	*low band frequency regions* [kHz]			
	0 - 1	0 - 2	0 - 3	0 - 4
$h(S)$	0.8094	0.8094	0.8094	0.8094
$h(N_S)$	0.7844	0.7417	0.7176	0.6991
$I(X;S)$	0.0250	0.0677	0.0918	0.1103

Table 1: Results from simulation showing the differential entropy of the slope and slope estimation noise and the lower bound on the mutual information.

entity	*low band frequency regions* [kHz]			
	0 - 1	0 - 2	0 - 3	0 - 4
$h(G)$	1.2347	1.2347	1.2347	1.2347
$h(N_G)$	1.2186	1.0764	0.8861	0.7603
$I(X;G)$	0.0161	0.1583	0.3486	0.4684

Table 2: Results from simulation showing the differential entropy of the gain and gain estimation noise and the lower bound on the mutual information.

The slopes s and the slope estimation noises n_s for all speech segments were used to estimate the differential entropies $h(S)$ and $h(N_S)$, respectively, as described in section 2.2. The histograms were computed using 30 bins. To have the same resolution in the quantization of the ranges of S and N_S, the same bin width Δ_S was used in both histograms. The mutual information was then computed by subtracting the differential entropy of the slope estimation noise from the differential entropy of the slope in accordance with (11). The simulations for calculating the lower bound on the mutual information between the gain G and the low frequency bands were performed analogously.

3.3. Results

The result of the simulation on synthetic data was a mutual information lower bound equal to 0.3 as compared to the true value 0.5. This means that our numerical methods indeed give a lower bound and have the correct order of magnitude.

Using real speech data and determining the lower bound on the mutual information between the slope of the high band and various widths of the low band gave the results shown in Table 1. Table 2 shows the results from the simulation with the gain. Observing Figure 1, we see that there is no less than 0.1 bit of mutual information between the spectral envelope of the low band frequency region 0 - 4 kHz and the slope of the high band. From Figure 1 we see that largest increase in mutual information is achieved when we increase the information about the low band from representing 0 - 1 kHz to representing 0 - 2 kHz. The mutual information then seems to level out as more information about the spectral characteristics of the low band are given. One possible explanation is that we have one formant in the region 0 - 1 kHz from which alone it is hard to estimate the slope, since the total number of formants determine the slope of the spectrum at the high band, assuming an all-pole signal model. All-pole moldels form a

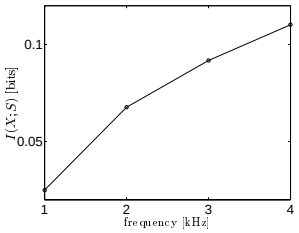

Figure 1: A lower bound on the mutual information between the slope of the high band given spectral envelope representation of the low band for regions 0 - 1, 0 - 2, 0 - 3 and 0 - 4 kHz.

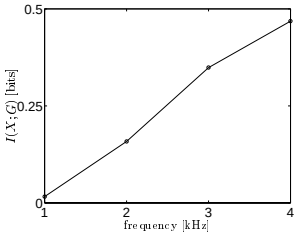

Figure 2: A lower bound on the mutual information between the gain of the high band given spectral envelope representation of the low band for regions 0 - 1, 0 - 2, 0 - 3 and 0 - 4 kHz.

good model of the vocal tract, and, thus, of the spectral envelope. However, by extending the region to 0 - 2 kHz we have a significantly better estimate of how many formants there are in the region 0 - 4 kHz and thus we obtain a better prediction of the slope.

Figure 2 shows the results obtained when we used the gain G instead of the slope S in the simulations. From Figure 2 we can see that the there is no less than 0.45 bit of mutual information between the gain and the spectral envelope of low band frequency region 0-4 kHz. The curve indicates that the shared information between the low band and the gain of the high band is related to how much of the speech signal energy we observe.

To test the sensitivity of the results towards the use of a histogram to estimate the differential entropy, we verified empirically that the results did not fluctuate with increasing number of histogram bins. We have not tested the sensitivity towards the codebook size, but using the bandwidth expansion scheme described in [3] we measured the average spectral distortion over 300 speech files from the TIMIT database with the codebook sizes 128 and 1024. The average spectral distortion was lowered by less than 0.2 dB using the codebook size of 1024 instead of 128. This experiment strongly suggests that the results would not change with finer quantization.

4. CONCLUSIONS

This paper describes a method for estimating a lower bound on the mutual information between a low band spectral coefficient vector and a high band spectral slope or gain. In the derivation of the method, we used a quantization step and an independence assumption and we showed in section 2.1 that these two assumptions are consistent with a lower bounding of the true mutual information. As a result of these steps, our method required significantly less data than a more direct approach. The sensitivity test of the results towards the number of bins used in the histograms showed that the results do not fluctuate. From the conducted simulations we can conclude that:

- there is mutual information between the low and high frequency bands;
- the mutual information is no less than 0.1 bit for the slope and 0.45 bit for the gain given the low band 0 - 4 kHz.

The 0.1 bit value we obtained for the lower bound on the mutual information between the low band and the slope of the high band is in the same order of magnitude as the entropy of the v/uv classification described in the introduction. Codebook based bandwidth expansion methods typically use between 7 and 10 bits to quantize the low band; it seems likely that these methods exploit more than 0.1 or 0.45 bit of information about the high band. Therefore, we a method for calculating mutual information that can handle a larger coefficient vector describing the high band, but still does not require an excessive amount of data, would likely reveal more mutual information between frequency bands in speech.

5. REFERENCES

[1] Y. M. Cheng, D. O'Shaugnessy, and P. Mermelstein, "Statistical recovery of wideband speech from narrowband speech," *IEEE Trans. Speech and Audio Proc.*, vol. 2, no. 4, pp. 544–548, 1994.

[2] C. Avendano, H. Hermansky, and E. A. Wan, "Beyond nyquist: Towards the recovery of broad-bandwidth speech from narrow-bandwidth speech," in *Proc. Eurospeech*, (Madrid), pp. 165–168, 1995.

[3] N. Enbom and W. B. Kleijn, "Bandwidth expansion of speech based on vector quantization of the mel frequency cepstral coefficients," in *IEEE Workshop on Speech Coding*, (Porvoo,Finland), pp. 1953–1956, 1999.

[4] J. Epps and W. H. Holmes, "A new technique for wideband enhancement of coded narrowband speech," in *IEEE Workshop on Speech Coding*, (Porvoo,Finland), pp. 174–176, 1999.

[5] H. Yang, S. v. Vuuren, and H. Hermansky, "Relevancy of time-frequency features for phonetic classification measured by mutual information," in *Proc. IEEE Int. Conf. Acoust. Speech Sign. Process.*, pp. 225–228, 1999.

[6] J. A. Bilmes, "Maximum mutual information based reduction strategies for cross-correlation based joint distributional modeling," in *Proc. IEEE Int. Conf. Acoust. Speech Sign. Process.*, pp. 469–472, 1998.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[8] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Dordrecht, Holland: Kluwer Academic Publishers, 1991.