Research & Reviews

ISSN: 2320-2459

Research and Reviews: Journal of Pure and Applied Physics

Estimation of Clean Spectrogram Noisy Value Functions Based on Metropolis Iterative Algorithm.

Mahdi Jalali*

Department of Electrical Engineering, Naghadeh Branch, Islamic Azad University, Naghadeh, Iran.

Short Communication

Received: 07/06/2013
Revised: 17/06/2013
Accepted: 24/06/2013

***For Correspondence**

Department of Electrical Engineering, Naghadeh Branch, Islamic Azad University, Naghadeh, Iran.

Keywords: spectrogram, metropolis, algorithm

ABSTRACT

The paper consisted of two parts. First, we estimated the clean speech signals from the estimated clean spectrograms with several values of K for one word. We then looked at the spectrograms of the estimated clean speech signals. Ideally, these two spectrograms (the estimated clean speech spectrogram and the spectrogram of the estimated clean speech) should be the same. We found that the spectrogram of the estimated clean speech signal with K=20 iterations looked closest to the estimated clean spectrogram. Next, we chose a column for which the estimated clean spectrogram and the spectrogram of the estimated clean speech signal visually differed.

INTRODUCTION

Spectrograms are more likely to have arisen from signal segments containing noise only (rather than speech and noise). For this reason, we initially define the bin delimiters in each bin. In other words, the band-dependent bin delimiters are set sequentially; the next band-dependent bin delimiter was set to the lowest gray scale value with corresponding noisy grayscale values in that range appear in the training set [1,2,3]. The denominator, 256, was chosen arbitrarily to be the maximum allowable number of bins. For some noisy grayscale value, corresponding ones in binary images become more probable than zeros. This means that for some noisy grayscale value, it is more likely that speech was present than not. The last bin is defined to contain strictly less than zeros [4,5]. For each band, the bin delimiters and the number of bins used were stored, which were not the same for the different bands and SNR values. The reason for such binning is to ensure that for each bin q, the counts in hist0 [q,b] and hist1[q,b] are at leastbcount1[b]/256c. That way, small bin counts that lead to poor estimation of the noise information probability are avoided. Several other binning methods were tested, including uniform linear binning and uniform log binning [5]. Both the uniform linear and uniform log binning contained bins whose histogram counts were small, less than 30, that is the reason that we used the much more complex binning described above.

MATERIALS AND METHODS

Short-Time Fourier Transform transforms (STFT) a finite-length, real-valued, discrete time sequence into an image. Henceforth, by sequence we mean a finite-length, real-valued discrete time sequence. To be precise, each operator $_s$in the family is determined by three positive integer parameters (L- the length of the sequence, I-the number of rows in the STFT, and s- the time skip step) and a window function W. The window function w is a sequence of 2I-2 real numbers, w[0],w[1],...,w[2I-3]. The domain is the set of sequences x of L real numbers, x[0],x[1],...,x[L-1]. The range of $_s$is the set of I_J-dimensional complex-valued images, where for i<I and j<J. Figure 1 displays how the STFT image is created from a time sequence.

Spectrograms, contain a wide range of real values. We therefore performed some post-processing of the spectrograms to get images that are useful. First a new image was created that contains the spectrogram values in decibels (dB). The decibel values were calculated from the spectrogram values. We chose the value 100 since it is a commonly used value. Next we set the value of maximum amplitude to be the largest dB value in the new image. Dynamic range is the range of viewable dB values and was set to 50 dB since that is the commonly used value. Therefore, the minimum amplitude is defined as minimum *amplitude = maximum amplitude-dynamic range*. In

the new image, all dB values below the minimum amplitude were set to the minimum amplitude. Finally the modified dB values were converted into a 0-255 integer scale by Henceforth, such a post-processed image of the spectrogram is referred to as the spectrogram image; these are the images used for display in this text. Figure 2 displays the spectrogram images of the first 1500 ms of the noise signal for 5 and 0 dB SNR. Since the two images are scaled independently, they are visually hard to differentiate. To test the accuracy of our spectrogram software we visually compared the spectrogram images produced by our software to those produced with the same parameters using [2].

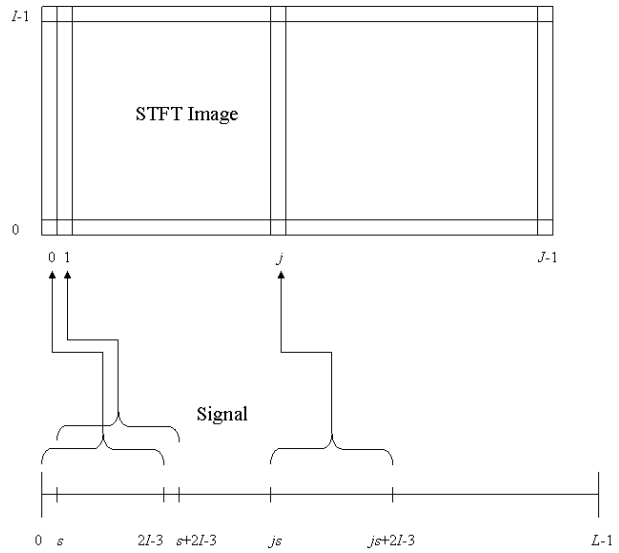

Figure 1: A display of how a STFT image is created from a time sequence.

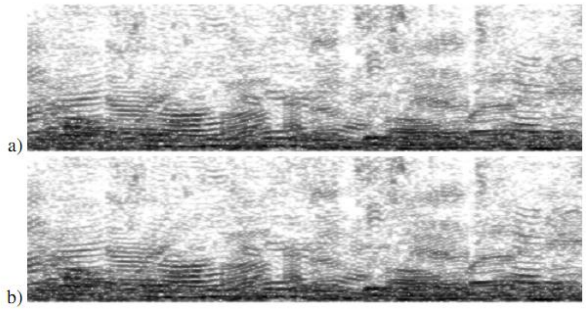

Figure 2: Spectrogram images of the first 1500 ms of the noise signal with a) 5 dB SNR, and b) 0 dB SNR.

RESULT AND DISCUSSION

We take the estimated binary column and the noisy grayscale spectrogram column to estimate the clean grayscale spectrogram column that corresponds to the unknown clean speech signal. We first describe how to "fuzzify" the estimated binary columns to create a [0,1]-valued column. We refer to this [0,1]-valued column as the soft segmentation column of the noisy spectrogram column. Similarly to the collection of soft segmentation columns as the soft segmentation image, or simply as the soft segmentation. the band-dependent posterior probability that the hard segmentation column (binary column) value at pixel is 1, given that bin and the binary values in the neighborhood configuration of the pixel are equal. Symbolically, each pixel of the soft segmentation column is set to the soft segmentation column value of pixel in row i=0 was set to zero. The meaning of such a column is that it locally estimates the conditional probability that a time-frequency location in the spectrogram contains clean speech. In the implementation, the posterior column values are set using one look-up table value from the float table. Figure 3 displays soft segmentation images for 5 and 0 dB SNR. Notice that these images are indeed smoother than the hard segmentations.

The "fuzzification" described here was more specific to the problem at hand and seemed to provide results that were better than expected with the simple "fuzzification" techniques. Next, we estimated the clean grayscale column that represents the unknown clean speech segment by multiplying the noisy grayscale spectrogram column by the soft segmentation column, pixel by pixel. This can be considered as using the soft segmentation column to mask the noisy grayscale spectrogram column. Since the soft segmentation column

ISSN: 2320-2459

is a [0,1]-valued column, the effect of the multiplication of the soft segmentation column by the noisy spectrogram column is that the spectrogram values are dampened. Notice that areas that were set to white in the hard segmentations, did not necessarily remain white in the soft segmentation. That way area in which the hard segmentation may have made the "wrong" decision, the noisy speech signal is not completely eliminated. Furthermore, areas which contained high values in the noisy spectrograms, were preserved with a lower value. This makes sense and is a desired behavior.

Figure 3: Soft segmentation images of the noisy spectrogram with a) 5dB SNR, and b) 0 dB SNR.

CONCLUSION

It is known that the probability of an image appearing in the Metropolis sequence is as likely as the probability of the image in the distribution. We, therefore, sampled the prior distribution using the Metropolis Algorithm. We then estimated the percent of black pixels in each band and compared it to the percent of black pixels in the training set. Several other neighborhoods were considered for the binary prior. We found that with our banding we could accurately estimate cliques and separator distributions containing more than five pixels from the given training set. By that we mean that for cliques containing more than five pixels, we found that there existed at least one binary clique configuration that appeared less than thirty times in the training set. Furthermore, not all collections of cliques formed neighborhoods for which Graphical Model theory could be used. These constraints were the reason for the chosen small neighborhood.

REFERENCES

1. N Roman, D Wang, GJ Brown. Speech segregate on based on sound localization. J Acoustical Soc America. 2003;114: 2236-2252.
2. BM Carvalho, GT Herman, S Matej, C Salzberg, E Vardi. Binary to mography for triplane cardiography. In A Kuba, M Samal, and A. Todd-Pokropek,editors, Information Processingin Medical Imaging, pages 29-41.Springer-Verlag,2003.
3. M Jalali, T Sedghi. Semi Supervised Feature Extraction for Filling Semantic Gap in Image Retrieval. In Proc. of IEEE, Machine Vision and Image Processing Symposium, Iran, 2011.
4. M Stridh, L Sörnmo, CJ Meurling, B Olsson. Sequential characterize tachyarrythmias based on ECG time-frequency analysis. IEEE Trans Biomed Eng. 2004;51:100-114.
5. T Fillon, J Prado. Evaluation of an ERB frequency scale noise reduction For hearing aids: A comparative study. Speech Comm. 2003;39:23-32.