Enhancement of Bone Microphone Speech Using Geometric Harmonics

Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions
Ronald R. Coifman, Stéphane Lafon

Presented by David Hexner
Outline

➤ Introduction
➤ Geometrics harmonics
➤ Frequency approach
➤ Time domain approach
➤ Summary
➤ Questions
Introduction

- Bone-conducted speech can be used for various speech applications
- It’s main advantage is that it is not affected by external noise
- However the quality of such speech is degraded (especially in the high frequency range)
- The purpose of this work is to restore to bone conducted speech so that it could be directly applied to human hearing systems and the front end of ASR systems
Ear Bone Microphone
Spectrogram of Ear Bone Microphone

Bone microphone Speech

Regular Microphone Speech
Let $X$ and $\bar{X}$ be two sets such that $X \subset \bar{X}$

Our goal is to extend the function $f$ defined on $X$ to $\bar{X}$

Let $K(x, y)$ be a positive semi definite symmetric kernel

There exists a set of $\{\lambda_j\}$ eigenvalues and $\{\psi_j\}$ eigenfunctions which are orthonormal and span $L^2$
Geometrics Harmonics cont.

\[ \lambda_j \psi_j(x) = \int_X K(x, y)\psi_j(y)\,d\mu(y) \]

- When \( \lambda_j \neq 0 \) \( \psi_j \) can be extended to \( \overline{X} \).

\[ \Psi_j(\overline{x}) = \frac{1}{\lambda_j} \int_X K(\overline{x}, y)\psi_j(y)\,d\mu(y) \]
Reconstruction Algorithm

\[ S_\delta = \{ j \text{ such that } \lambda_j > \delta \lambda_0 \} \]
\[ f = \sum_{j \in S_\delta} \langle f, \psi_j \rangle \psi_j \]

Extension algorithm:
\[ f(\bar{x}) = \sum_{j \in S_\delta} \langle f, \psi_j \rangle \Psi_j \]
Bone Conducted Speech Enhancement
Frequency Approach

- We assume that high frequency content is corrupted
- Try to reconstruct the high frequencies based on the low frequencies
Frequency Approach Training

- Let $A[K \times N]$ be the STFT matrix of speech from a regular microphone

$$W[k, j] = \frac{1}{N} \sum_n \log(A[k, n]) \log(A[j, n])$$

- The logarithm was used to turn multiplicative disturbances to additive
Frequency Approach Training Cont.

- \( W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \)

- \( W_{11} [\widetilde{K} \times \widetilde{K}] \) represents low frequencies
  - \( \widetilde{K} \) is chosen according to the low frequencies without disturbances

- Calculate eigenvalues and eigenvectors of \( W_{11} \), \( \{ \lambda_j \} \) and \( \{ \psi_j \} \) respectively

- Define \( \Psi_j = \frac{1}{\lambda_j} \begin{bmatrix} W_{11} \\ W_{21} \end{bmatrix} \psi_j \)
Frequency Approach Reconstruction

- Let $F[K \times M]$ be the STFT matrix of speech from a bone microphone.
- Let $\tilde{F}[K - \tilde{K} \times M]$ be the low frequency band matrix of $F$.

$$
F = \begin{bmatrix}
    f_{1,1} & \cdots & f_{1,N} \\
    \vdots & & \vdots \\
    f_{\tilde{K},1} & \cdots & f_{\tilde{K},N} \\
    \vdots & \ddots & \vdots \\
    f_{K,1} & \cdots & f_{K,K}
\end{bmatrix}
$$
Frequency Approach Reconstruction
Cont.

Let $\bar{F}[K \times M]$ be the reconstructed matrix

$\bar{f}_i = \sum_j \Psi_j \psi_j^T \tilde{f}_i$, $\tilde{f}_i$ and $\bar{f}_i$ columns of matrixes $\tilde{F}$ and $\bar{F}$ respectively

$\bar{F} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ f_{\tilde{K}+1,1} & \cdots & f_{\tilde{K}+1,N} \\ \vdots & & \vdots \\ f_{K,1} & \cdots & f_{K,K} \end{bmatrix}$
Frequency Approach Reconstruction Cont.

- In order not to lose the high frequency content that does exist in the bone speech, the original high frequency content is returned

\[ F_{enhanced} = 10^\bar{F} + \bar{F} \]
Frequency Approach Reconstruction

Bone microphone Speech

Enhanced Bone Microphone Speech

Air Microphone Speech
![](image)

Frequency Approach Reconstruction – Compression Point of View

- The algorithm described, can reconstruct the full spectrum speech from about 30%-40% of the spectrum (that would be the compression rate)
Frequency Approach – Compression
Time Domain Approach

- Let $X[K \times N]$ be the STFT matrix of training data of the bone microphone.
- Let $F[K \times N]$ be the STFT matrix of the training set of the air microphone recorded simultaneously.
- Let $\tilde{X}[K \times M]$ be the STFT matrix of new data from a bone microphone.
- We model the air microphone matrix as a function of the bone microphone data:
- We know the values at $\tilde{f}(\tilde{x})$ and want to extend to $\tilde{f}(\bar{x})$.
Time Domain Approach - Training

- Let $W[N \times N]$ be an affinity matrix based on a Gaussian kernel of $X$:

$$W[j, k] = \exp(-\frac{||\log(x_i) - \log(x_k)||^2}{\sigma^2})$$

- Calculate eigenvalues and eigenvectors of $W$, $\{\lambda_j\}$ and $\{\psi_j\}$ respectively.
Time Domain Approach - Reconstruction

- Let $\bar{W}[M \times N]$ be a Gaussian kernel matrix

  \[
  \bar{W}[j, k] = \exp\left(\frac{-||\log(x_j) - \log(x_k)||^2}{\sigma^2}\right)
  \]

- $\bar{W}_j = \frac{1}{\lambda_j} \bar{W}\psi_j$

- $\bar{f}_i(\bar{x}) = \sum_{j \in S_\delta} f_i \psi_j \psi_j^T$

- Return the phase from bone microphone signal
Time Domain Approach - Reconstruction

Bone microphone Speech

Enhanced Speech
Summary

- We compared two applications of geometric harmonics for enhancement of bone microphone speech
- Reconstructing of high frequencies based on low frequencies
- Reconstruction via extension algorithm on time domain