Vowel Recognition Using Bayesian Analysis

Tilman Birnstiel

University at Albany
Physics Department

Bayesian Data Analysis and Signal Processing
12 / 12 / 2006
Outline

1. Introduction
   - The Human Vocal Mechanism
   - The Problem

2. Bayesian Solution
   - The Basic Idea
   - How To Solve It

3. Results and Conclusions
   - Results
   - Conclusions
Outline

1 Introduction
   • The Human Vocal Mechanism
   • The Problem

2 Bayesian Solution
   • The Basic Idea
   • How To Solve It

3 Results and Conclusions
   • Results
   • Conclusions
Vowels

- a vowel produces a periodic repeating signal
- several frequencies seem to be involved

Fig. 1: Signal of the vowel [a]
The air in each cavity vibrates in a characteristic frequency and harmonics of this frequency
⇒ the frequencies are the eigenmodes of the cavities

the main cavities are: oral cavity, nasal cavity and the upper throat

the tongue modulates the oral cavity
⇒ the vowel signal should be decomposable in few characteristic frequencies: the formants
Fourier Transformation

Fig. 3: Frequency spectrum of the Vowel [a]
Model 1

First model: a sum of 9 sinusoids (3 formants and 6 harmonics of the basic frequency)

\[ s(t) = \sum_{i=1}^{9} A_i \sin (2 \pi f_i (t + \delta_i)) \]  

Parameters:
- 9 amplitudes \( A_1, \ldots, A_9 \)
- 3 formant frequencies \( f_1, f_2, f_3 \)
- 9 phase shifts \( \delta_1, \ldots, \delta_9 \)

21 Parameter!
Outline

1. Introduction
   - The Human Vocal Mechanism
   - The Problem

2. Bayesian Solution
   - The Basic Idea
   - How To Solve It

3. Results and Conclusions
   - Results
   - Conclusions

Tilman Birnstiel  Vowel Recognition
Model 2

Still a sum of sinusoids, but frequencies, amplitudes and shifts are determined *successively* (see Fig. 3):

$$s(t) = \sum_{i=1}^{N} A_i \sin \left( 2\pi f_i (t + \delta_i) \right)$$

Begin with:
- search for $f_1, A_1, \delta_1$ (given that $f_1$ is at about 100 Hz) then
- search for $f_2, A_2, \delta_2$ (given $f_1, A_1, \delta_1$) then
- ...  

This is possible because of the orthogonality of the sinusoids.
The following Priors and Likelihoods are used:

- $\sigma$ of the data unknown: use Student-t-Distribution with Jeffrey’s Prior
- basic frequency at 100 Hz (broad Gaussian with $\sigma = 50$)
- use the determined frequency for the next prior:

$$f_{i+1} = f_i + f_1$$  \hspace{1cm} (3)

- for the model selection process, only $\sigma_A$ and $\bar{f}$ are known  
  $\Rightarrow$ use Gaussian (max. Entropy)
only the vowels [a] (e.g. father), [i] (e.g. he) and [o] (e.g. code)
only the first 7 frequencies are used
6 examples for each vowel are used to build the model
the basic frequency is assumed to be at about 100 Hz (see priors)
use MCMC-Algorithm by Kevin Knuth
Create Database

returns frequencies and amplitudes

calls

Model Selection

creates avg. Model for each vowel

calls calls

Model

returns value

calls

MCMC

returns best parameters

calls

logPosterior

Data

returns best parameters

new Data

calls

Frequency Search

Fig. 4: How to determine the vowel

Tilman Birnstiel Vowel Recognition
Model Selection

**Aim**: determine which of the mean vowel models fits best to the data

Use Bayes’ Theorem again. For example for vowel [a]:

\[ P([a]|\{d_i\}, I) = \frac{P(\{d_i\}|[a], I) \times P([a]|I)}{Z} \]  \hspace{1cm} (4)

Where \( Z \) is the evidence:

\[ Z = \sum_{\text{all vowels } v_i} P(\{d_i\}|v_i, I) \]  \hspace{1cm} (5)

⇒ a probability for each vowel
Outline

1. Introduction
   - The Human Vocal Mechanism
   - The Problem

2. Bayesian Solution
   - The Basic Idea
   - How To Solve It

3. Results and Conclusions
   - Results
   - Conclusions
Fig. 5: Evolution of the Parameters and the logP for a frequency search.
## Success Rate

| data | $P(a|d)$ | $P(i|d)$ | $P(o|d)$ | ✔ / ✗ |
|------|----------|----------|----------|--------|
| $a_1$ | 99.99%   | 0%       | 0.01%    | ✔      |
| $a_2$ | 1.6%     | 0%       | 98.4%    | ✗      |
| $a_3$ | 99.99%   | 0%       | 0.01%    | ✔      |
| $i_1$ | 0%       | 100%     | 0%       | ✔      |
| $i_2$ | 0%       | 100%     | 0%       | ✔      |
| $i_3$ | 0%       | 98.4%    | 1.6%     | ✔      |
| $o_1$ | 42.8%    | 0%       | 57.0%    | ✔      |
| $o_1$ | 0%       | 0%       | 100%     | ✔      |
| $o_1$ | 0%       | 0%       | 100%     | ✔      |
Achievements & Further Problems

- Algorithm works fine for 3 vowels
  - ⇒ Extend to more vowels
- Most vowels are determined successfully
  - ⇒ What went wrong with $a_2$
- It takes about 50 minutes to get a result
  - ⇒ Further optimization
  - ⇒ Sharpen priors?
Devinderjit Sivia and John Skilling.  
*Data Analysis: A Bayesian Tutorial.*  

Kevin H. Knuth.  
Lecture notes: Bayesian data analysis and signal processing, 2006.

Wikipedia.  
Vowel — wikipedia, the free encyclopedia, 2006.  
[Online; accessed 1-December-2006]  

C.R. Nave.  
The human voice, 2005.  
[Online; accessed 11-December-2006]  
http://hyperphysics.phy-astr.gsu.edu/hbase/music/voicecon.html.

Larry G. Bretthorst.  