Vowel type, speaker size, and a bit of measurement error accounts for all the information in sustained vowels

Richard Turner (turner@gatsby.ucl.ac.uk)

Gatsby Computational Neuroscience Unit, 24/08/2007
Motivation
A Gross Approximation

Three Approximations:

• **VTL growth**: All vocal tracts are the same shape. \( l_i^n = a_{n,m} l_j^m \)

• **Physics**: Formant wavelengths are all standing-wave-like. \( \lambda_i^n = \alpha_i l_i^n \)

• **Noise**: Formant frequencies can be recovered noiselessly.

Two Predictions:

• Vowel clusters should be **pencils** pointing to the origin

• **Ratios of formants** should be speaker independent
Clustering via PCA
Clustering via PCA - $\langle \theta \rangle = 4.2$
Clustering via PCA - $\langle \theta \rangle = 4.2$
Are people scaled clones?
If people aren’t clones, but the physics is simple
Is the physics standing-wave-like?

\[ \lambda_3 = 2 \, l_B \]

\[ \lambda_2 = 4 \, l_F \]

\[ \lambda_1^2 = A_B \, l_B \, l_C / A_C \]
Are the measurements noiseless?
Are the measurements noiseless?
Are the measurements noiseless?
Is this true for the P&B data?
What effect does this noise have?
What effect does this noise have?
What effect does this noise have?
What effect does this noise have?
A new model - Noisy and VT Shape Varying

\[
\begin{pmatrix}
  f_1 \\
  f_2 \\
  f_3
\end{pmatrix}
= \begin{pmatrix}
  g_1 \\
  g_2 \\
  g_3
\end{pmatrix} a + \begin{pmatrix}
  \mu_1 \\
  \mu_2 \\
  \mu_3
\end{pmatrix} + \begin{pmatrix}
  \epsilon_1 \\
  \epsilon_2 \\
  \epsilon_3
\end{pmatrix} + \sum_{i=1}^{I} \begin{pmatrix}
  h_{1,i} \\
  h_{2,i} \\
  h_{3,i}
\end{pmatrix} a_i
\]  

\(p(a) = \text{MOG}\)  
\(p(\epsilon) = \text{Norm}(0, \sigma^2)\)
Results

$\langle \theta \rangle = 4.2$
Results - $\langle \theta \rangle = 4.2$
Results - $\langle \theta \rangle = 4.2$
Results

The diagram shows the probability of occurrence for relative VTL (Vertical Text Line) values. The x-axis represents the relative VTL, with values ranging from -0.3 to 0.5. The y-axis represents the probability of occurrence, ranging from 0 to 7. The histogram and curve indicate the distribution of relative VTL values.
Conclusion

- Formant Frequency data appear to contain a **considerable contribution from observation noise**.

- Once this noise is removed, the **remaining variability** is mainly in the form of a **uniform scaling**.

- Speakers appear to **compensate** for the non-uniform growth of the VT.