



UNIVERSITY OF  
BIRMINGHAM

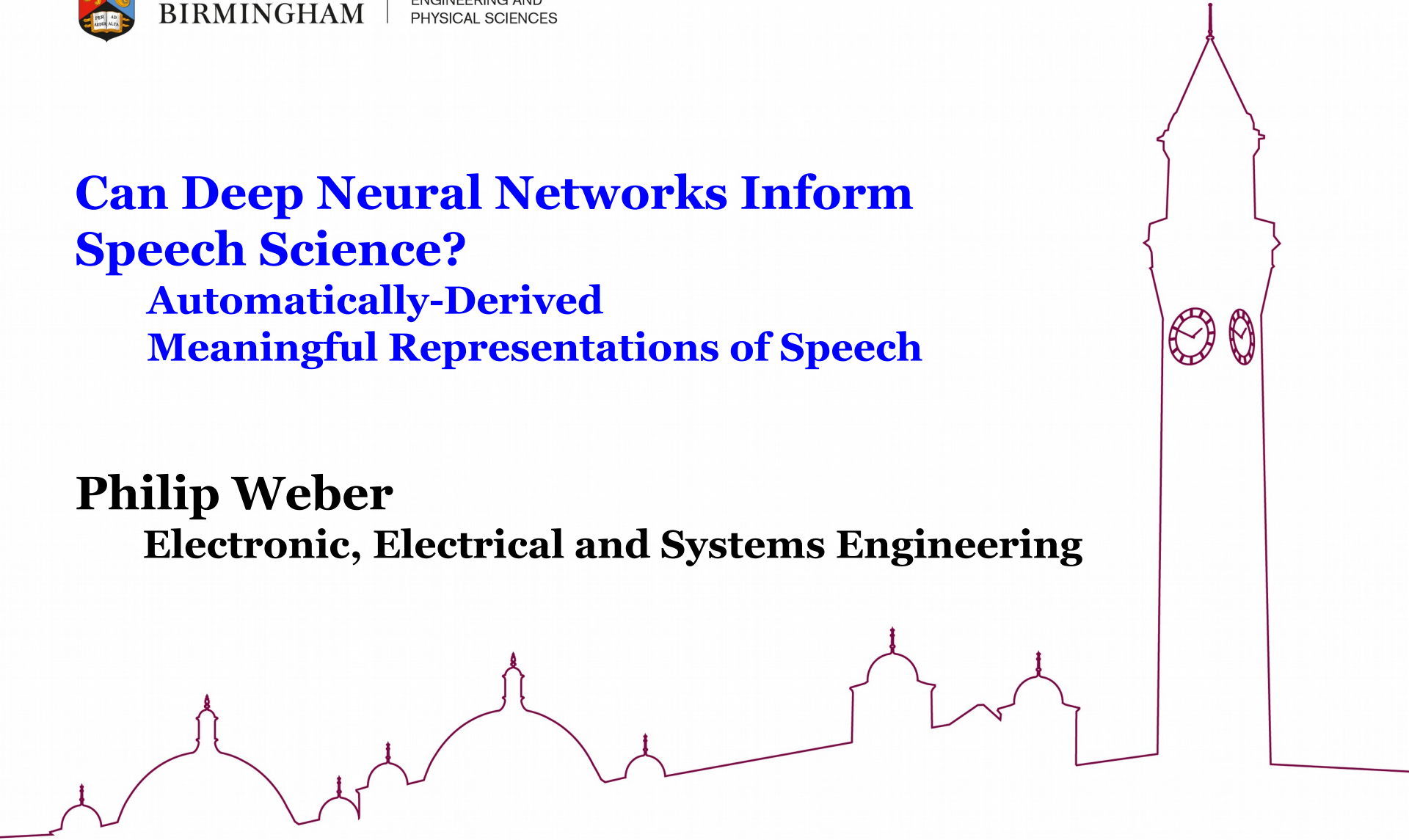
COLLEGE OF  
ENGINEERING AND  
PHYSICAL SCIENCES

# Can Deep Neural Networks Inform Speech Science?

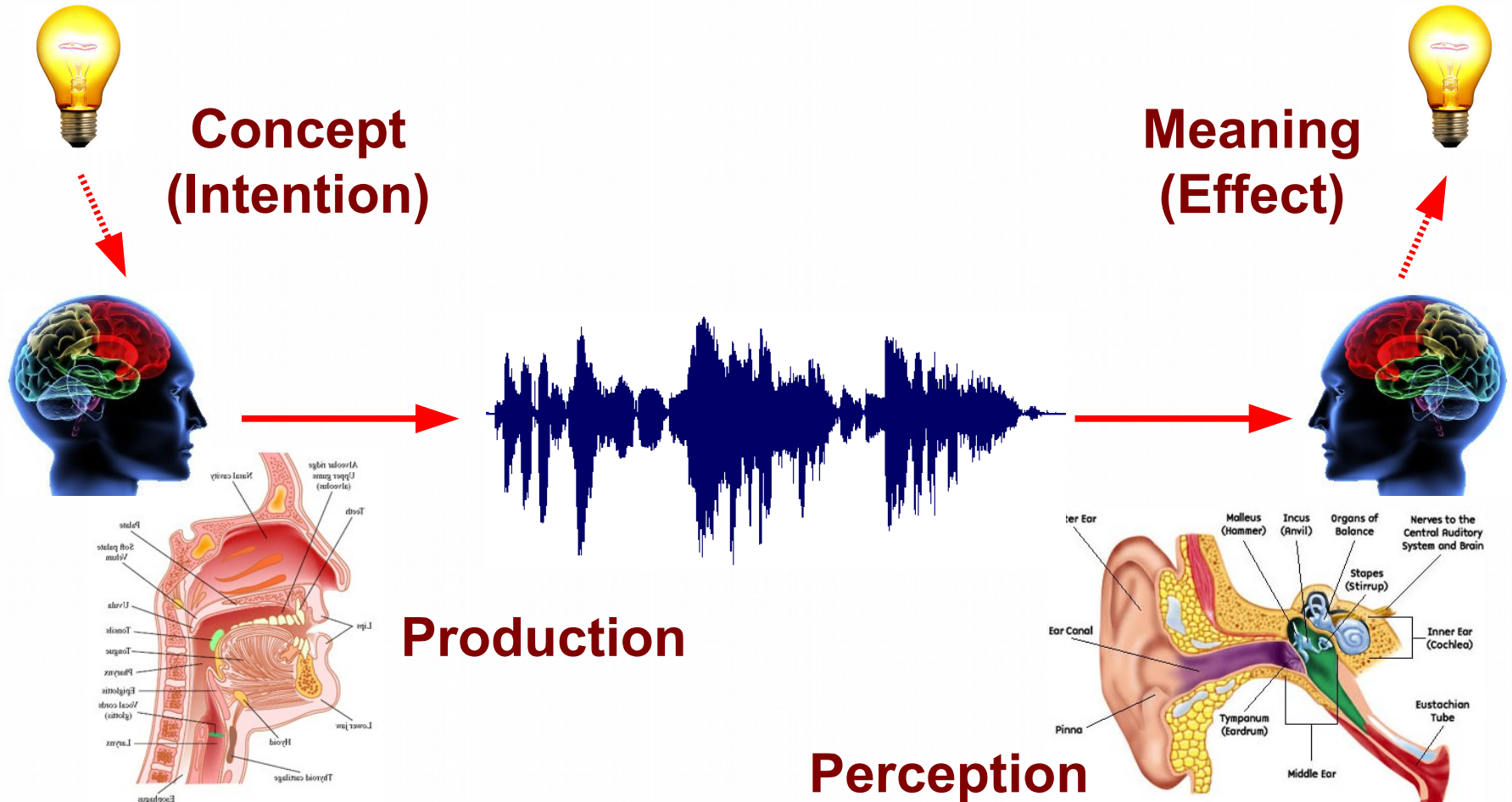
**Automatically-Derived  
Meaningful Representations of Speech**

**Philip Weber**

**Electronic, Electrical and Systems Engineering**



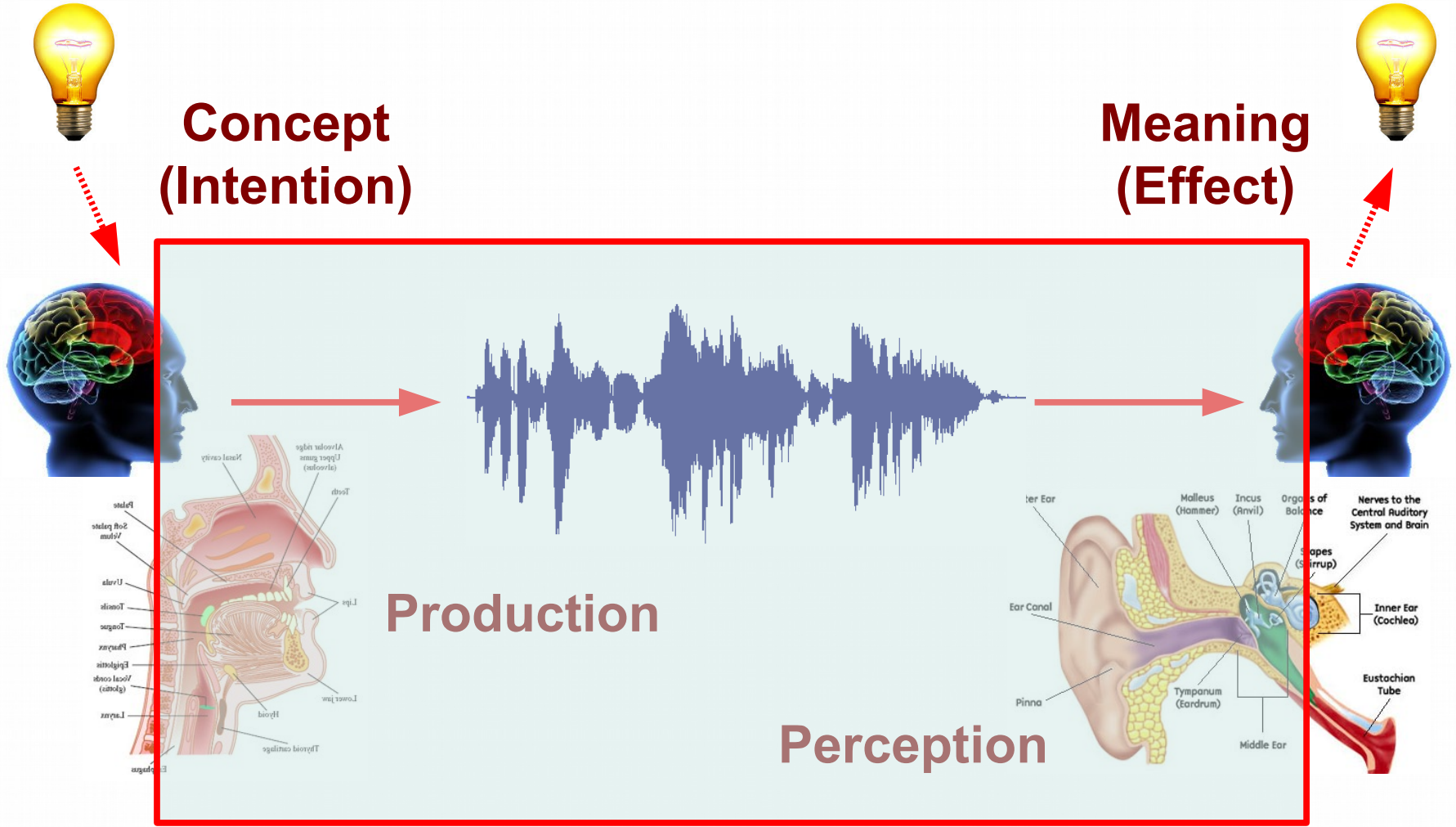
# Human Speech (Recognition)



UNIVERSITY OF  
BIRMINGHAM

COLLEGE OF  
ENGINEERING AND  
PHYSICAL SCIENCES

# Automatic Speech Recognition (Speech to Text)



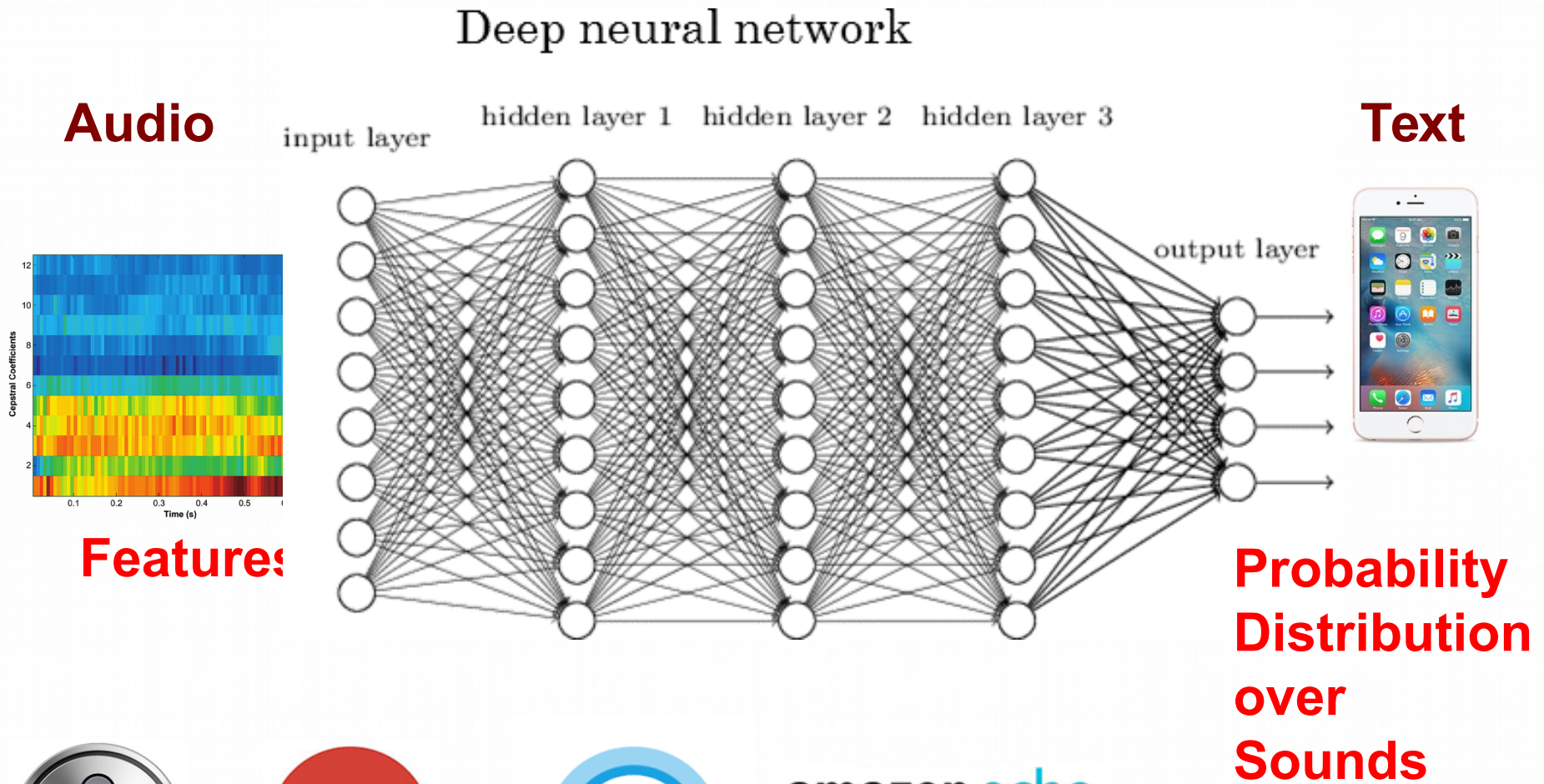
UNIVERSITY OF  
BIRMINGHAM

COLLEGE OF  
ENGINEERING AND  
PHYSICAL SCIENCES

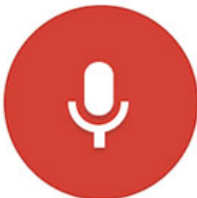




# Deep Learning – Statistical Models



IVEF  
RMIN

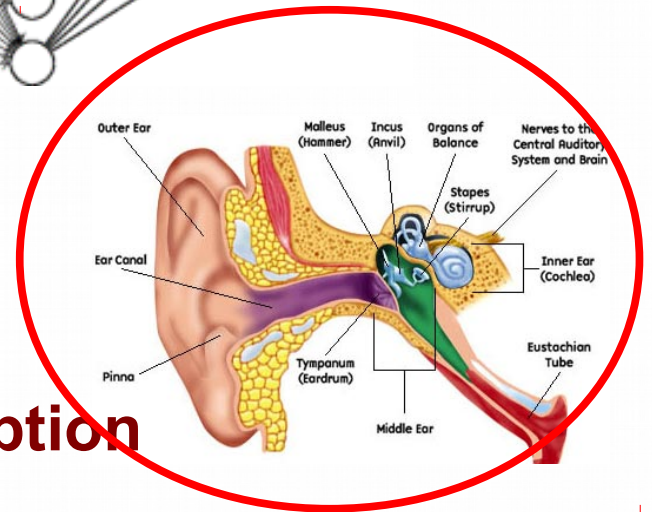
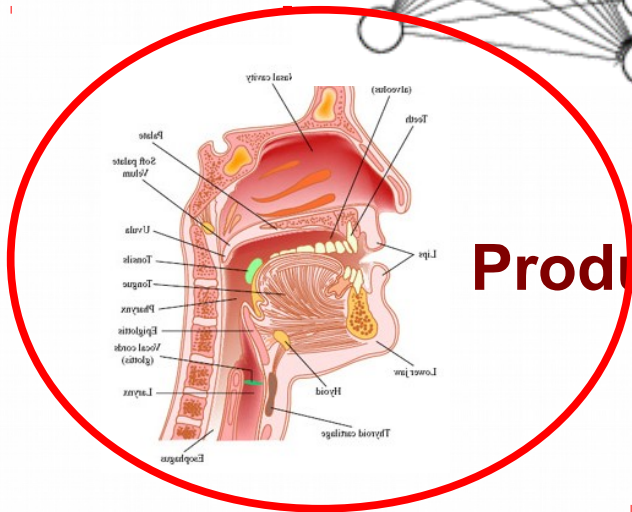
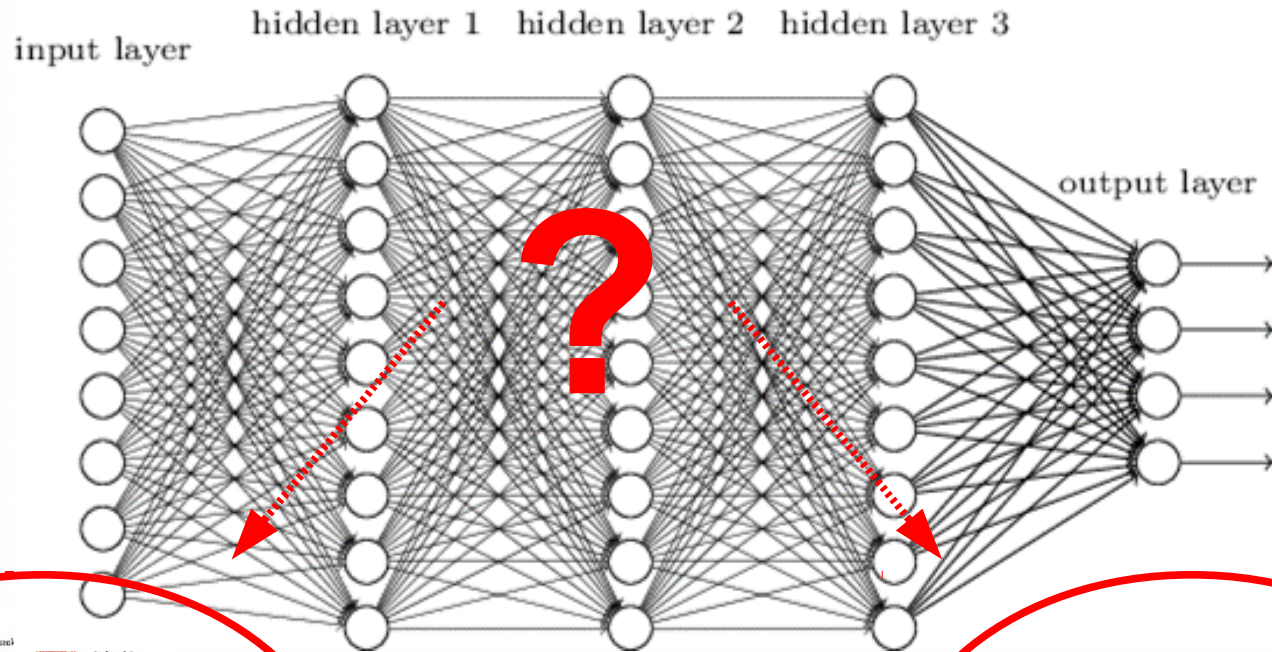


COLLEGE OF  
ENGINEERING  
PHYSICAL



amazon echo

# Deep neural network



UNIVERSITY OF  
BIRMINGHAM

COLLEGE OF  
ENGINEERING AND  
PHYSICAL SCIENCES

# Meanwhile ... Speech Science

---

***“the study of production, transmission  
and perception of [human] speech”***

**[Wikipedia]**

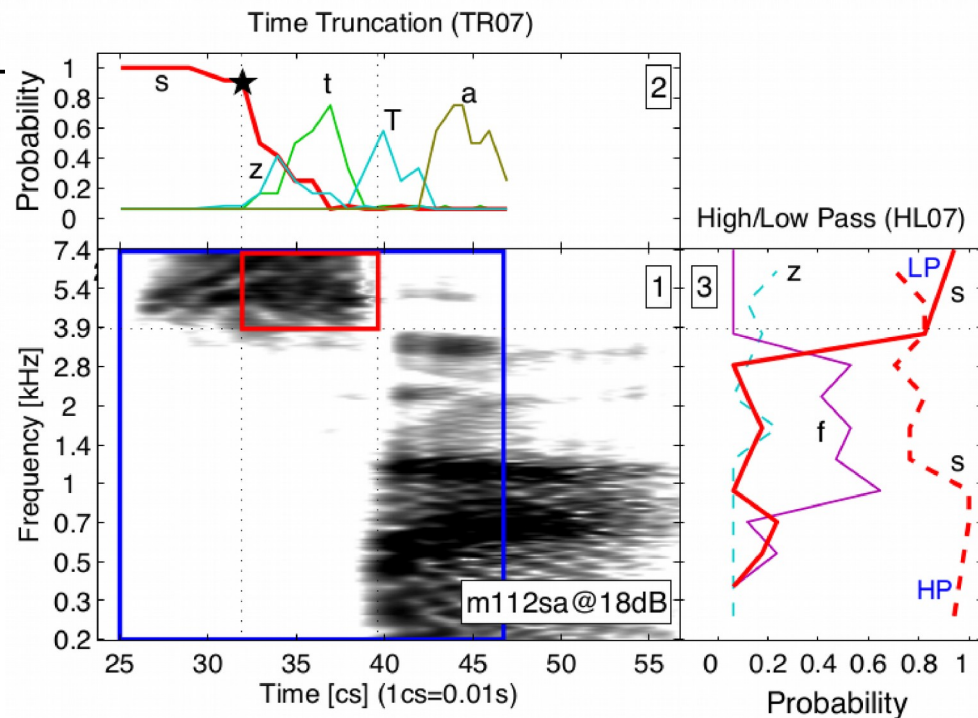
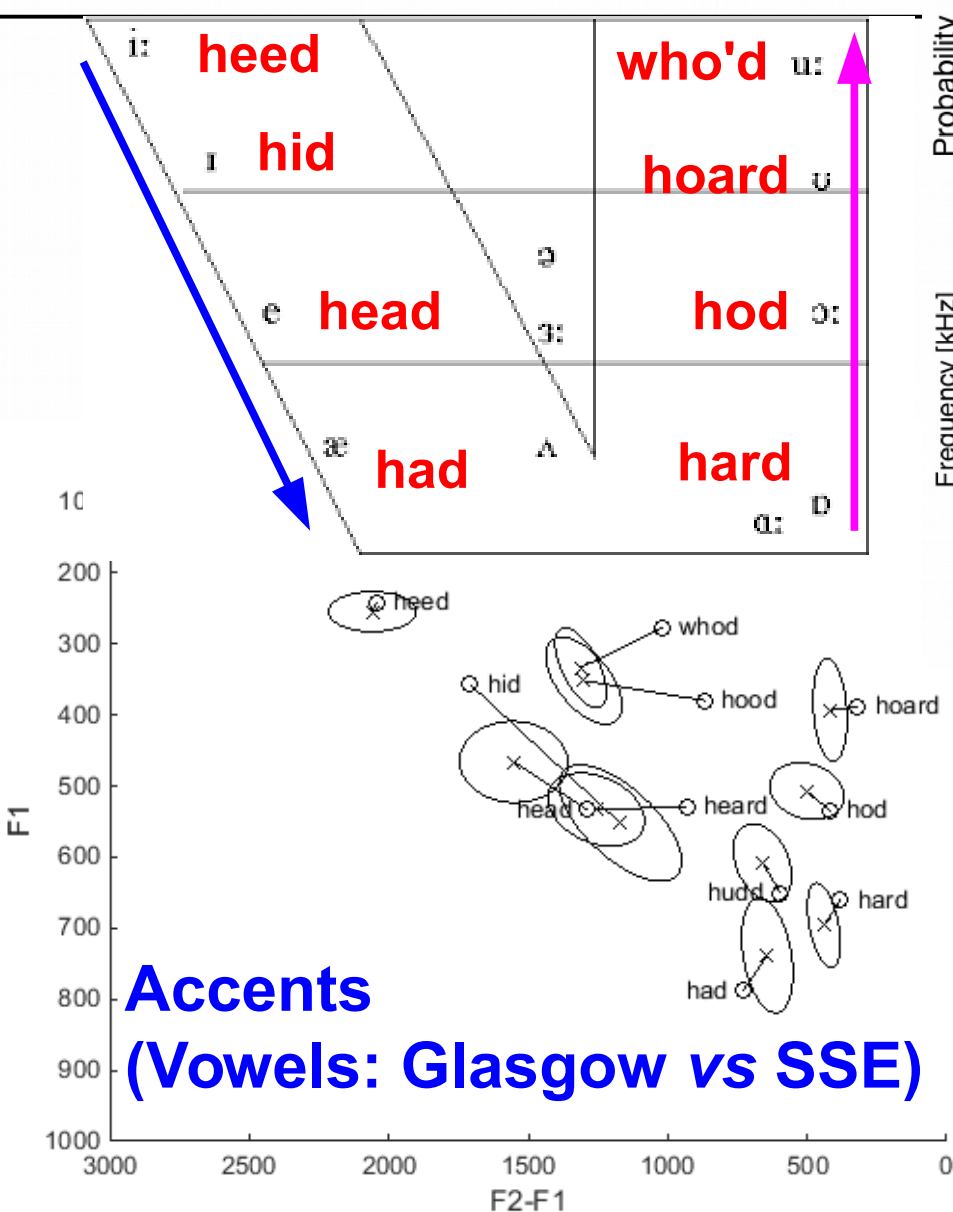


UNIVERSITY OF  
BIRMINGHAM

COLLEGE OF  
ENGINEERING AND  
PHYSICAL SCIENCES



# Speech Science



**Consonant Perception**

**Accents**  
**(Vowels: Glasgow vs SSE)**



# Speech Group (EESE)

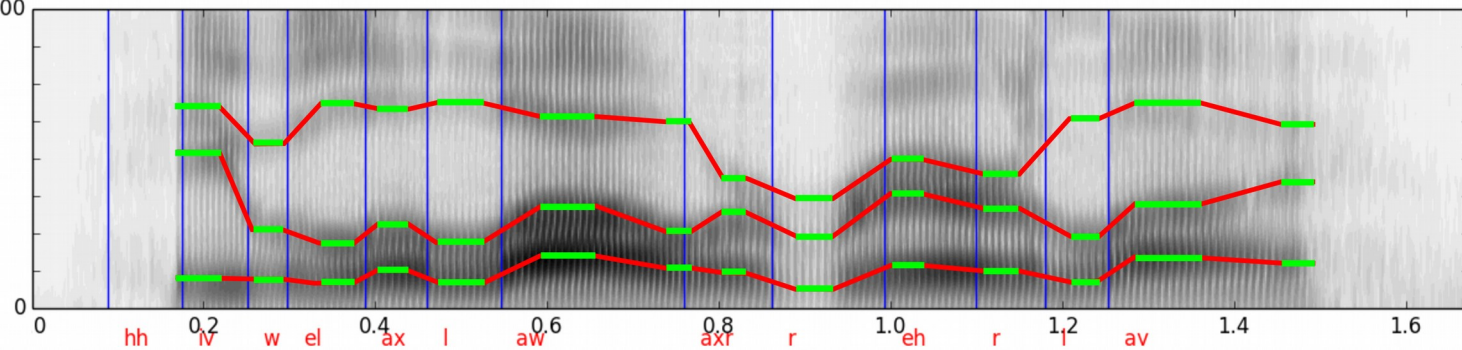
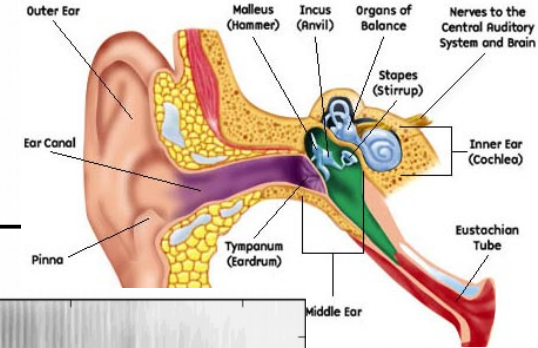
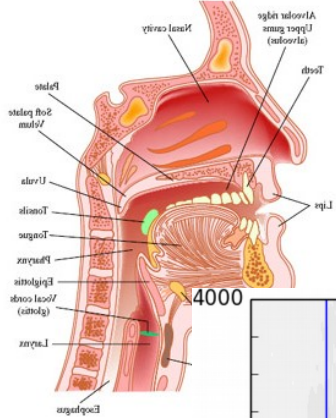
---

## What do we do?

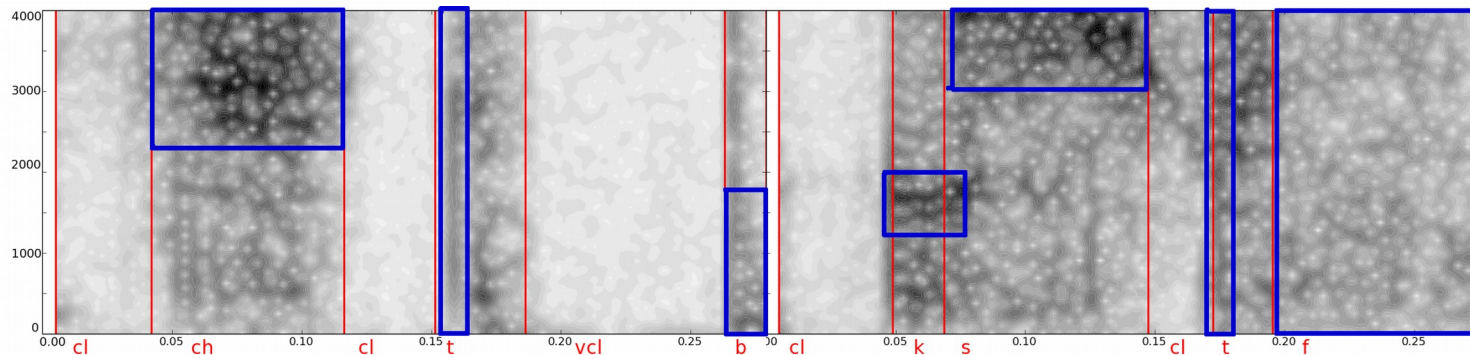
- **“Faithful” models of speech**
- **Speech models and representations**
- **Children's speech**
- **Accented speech**
- **...**



# Low-dimensional Representations of Speech



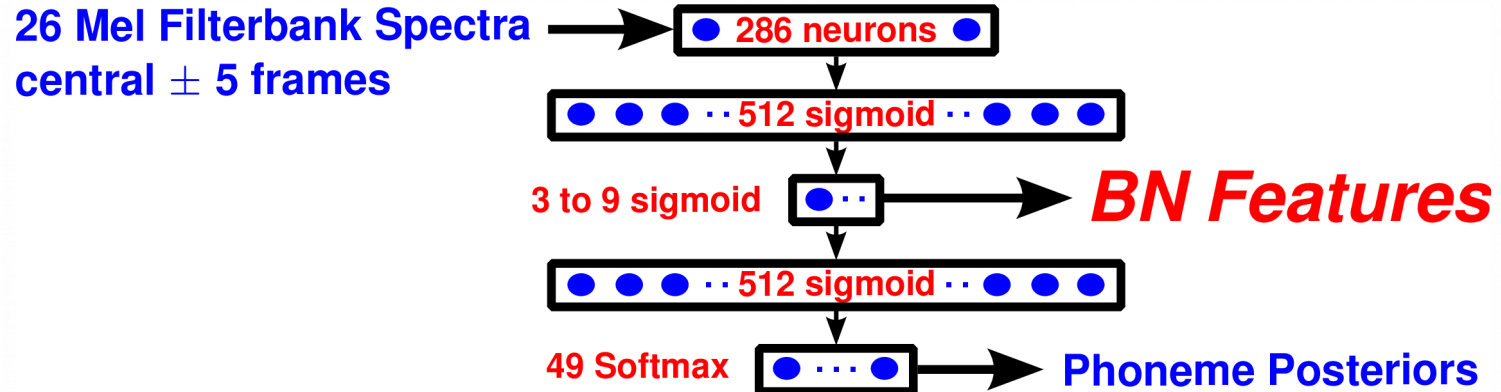
**Voiced Sounds – Formants**



**Unvoiced Sounds – Spectral Energies**



# Automatically-Derived “Bottleneck” Features



## Some Recognition Results:

Features	Dimension	% Err	Model	# Parameters
Standard MFCC	39	29.1	“Standard”	$1.4 \times 10^7$
Bottleneck	9	29.4	“Standard”	$2.3 \times 10^5$
Bottleneck	9	38.1	“Faithful”	535





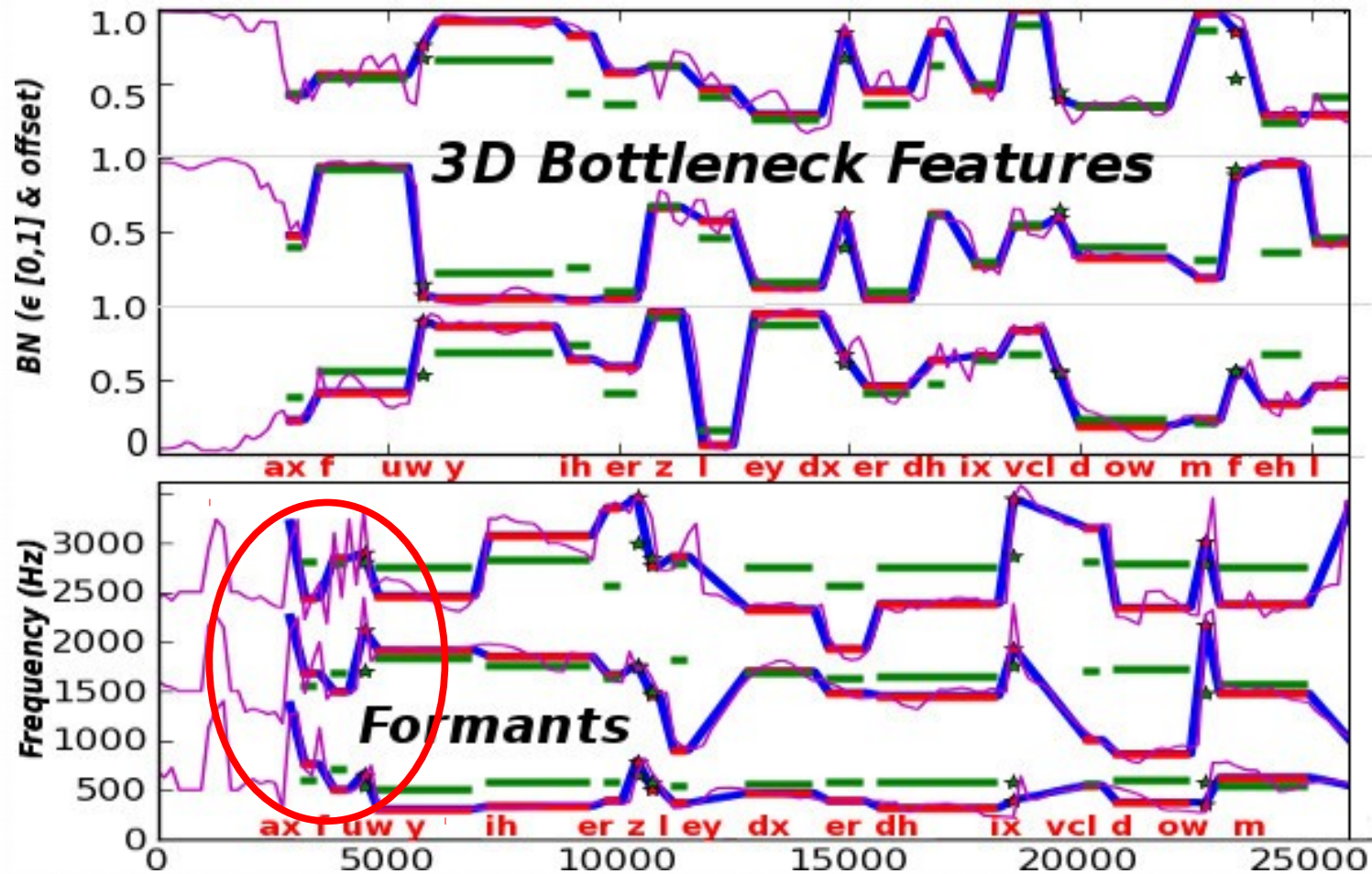
**But what are these  
bottleneck features?**

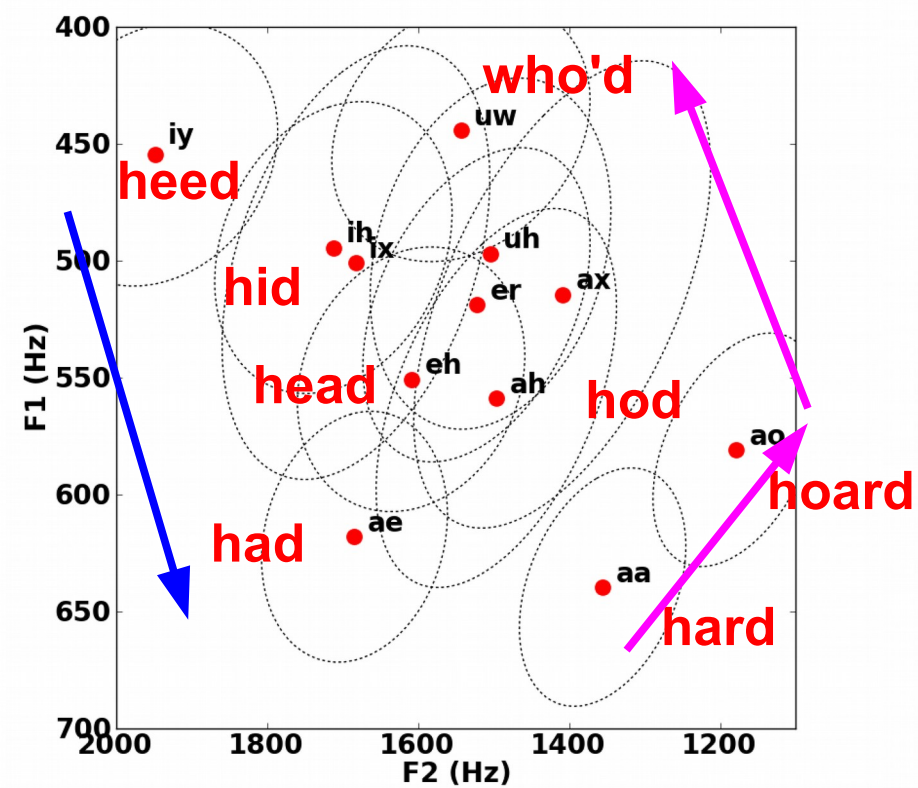
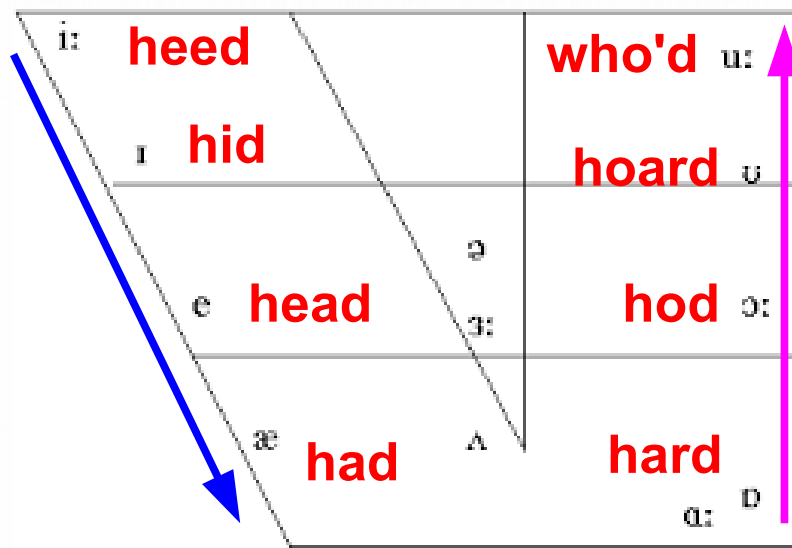


UNIVERSITY OF  
BIRMINGHAM

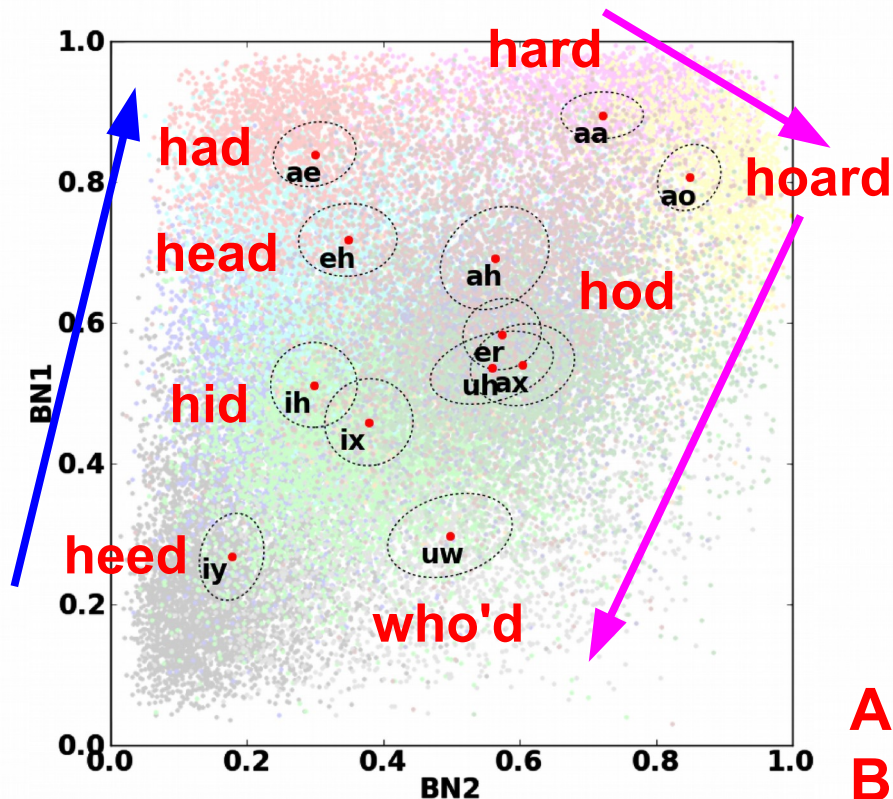
COLLEGE OF  
ENGINEERING AND  
PHYSICAL SCIENCES

# Interesting Dynamics for all Sounds





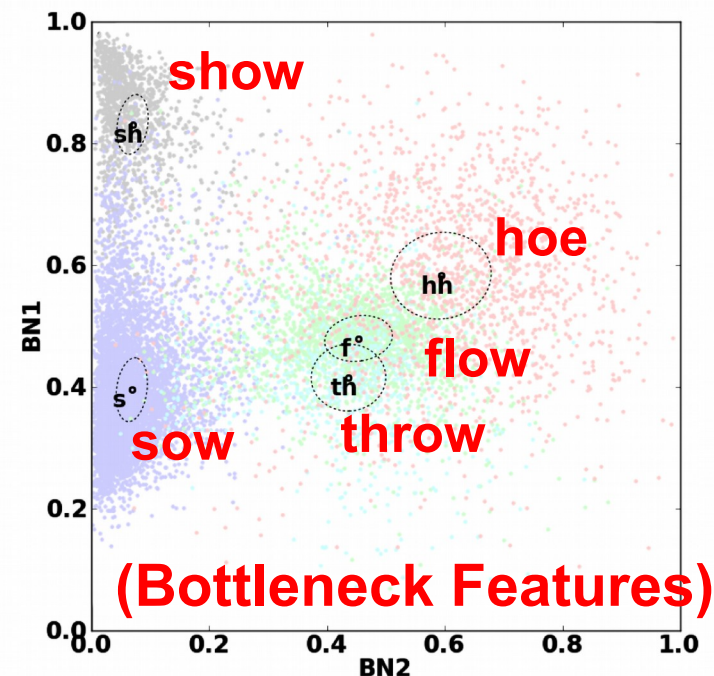
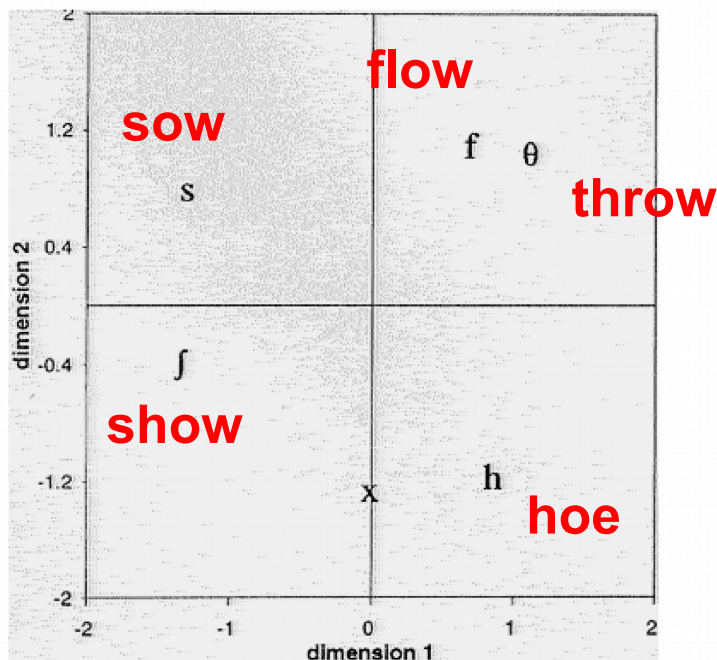
**Vowel Space Diagrams  
(formants)**



**Automatically-derived  
Bottleneck Features**

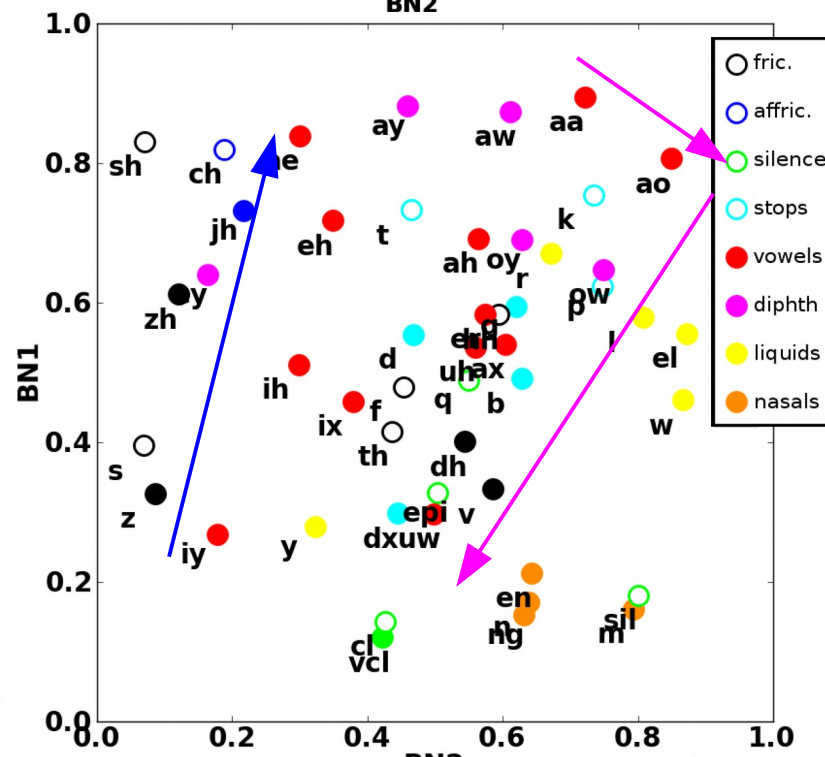
**Interesting Relationships  
with “Natural” Features**





## Fricatives [Choo et al. 1997]

A Representation for analysis  
of all Sounds?



UNIVERSITY OF  
BIRMINGHAM

COLLEGE OF  
ENGINEERING AND  
PHYSICAL SCIENCES

# Speculation and Collaboration

---

- **Robust visualisation methods to investigate speech phenomena?**
- **For more than just vowels?**
  - **Children's speech acquisition / errors?**
  - **Accents and how they change?**
  - **Language learning?**
  - ...
- **Novel methods to improve mainstream ASR?**
- ...



# Speculation and Collaboration

- **Robust visualisation methods to investigate speech phenomena?**
- **For more than just vowels?**
  - **Children's speech acquisition / errors?**
  - **Accents and how they change?**
  - **Language learning?**
  - ...
- **Novel methods to improve mainstream ASR?**
- ...

**Thank You!**



UNIVERSITY OF  
BIRMINGHAM

COLLEGE OF  
ENGINEERING AND  
PHYSICAL SCIENCES