Phonetic Feature Encoding in Human Superior Temporal Gyrus

Nima Mesgarani, Connie Cheung,
Keith Johnson, Edward F. Chang

BIONB4110
March 17, 2014
Presented by: Erika Glaubitz and Isabelle Stark

Science

Founded in 1880 (with the financial assistance of Thomas Edison)

Published weekly

Impact Factor: 31.027

American Association for the Advancement of Science

AAAS

National science community

Non-profit

Mission statement: To “advance science, engineering, and innovation throughout the world for the benefit of all people.”

Three journals

People

NIMA MESGARANI, PhD

PhD in Electrical Engineering from University of Maryland

Postdoc at Johns Hopkins Center for Language and Speech Processing and UCSF Neurosurgery Department

Assistant Professor of Electrical Engineering at Columbia

Interested in integrating engineering and neuroscience

http://www.ee.columbia.edu/~nima/web/Home.html

People

CONNIE CHEUNG

BS UC Berkeley College of Engineering

“Idea to IPO” (Initial Public Offering) course winner - human brain mapping algorithm

PhD candidate in Bioengineering and Biomedical Engineering at UCSF under the supervision of Edward Chang (PI)

http://www.ucsf.edu/news/2013/04/105096/innovative-course-teaches-how-turn-ideas-life-sciences-ventures

http://www.linkedin.com/pub/connie-cheung/17/360/ab1

People

KEITH JOHNSON, PhD

PhD in Linguistics from Ohio State

Professor - Department of Linguistics - UC Berkeley

Director of the UC Berkeley Phonology Lab

Helped with the linguistic aspects of the study

Currently working on phonetic studies of Cherokee and Creek

http://linguistics.berkeley.edu/~kjohnson/

People

EDWARD CHANG, MD - PI

BA Amherst; MD UCSF

Associate Professor of Neurological Surgery and Physiology - UCSF

Chief of Epilepsy and Pain Neurosurgery

Co-Director of the Center for Neural Engineering and Prostheses

Interest in epilepsy and neuroprosthetics for those with brain damage

http://www.cin.ucsf.edu/HTML/Edward_Chang.html

UCSF Center for Integrative Neuroscience

http://www.clarkconstruction.com/our-work/projects/ucsf-sandler-neurosciences-center

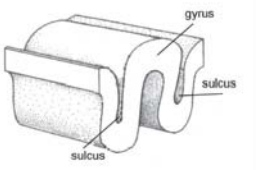
Established at UC San Francisco in 1990

Research is heavily funded by NIH and the University of California

Research focuses on how nerve cells in the brain work together to generate behavior

Superior Temporal Gyrus

Gyrus: ridge on cerebral cortex



Superior Temporal Gyrus

What is the function of the superior temporal gyrus?

http://en.wikipedia.org/wiki/Gyrus
http://www.studyblue.com/notes/note/n/2012-fall-neuroscience-exam-1/deck/3620185

Superior Temporal Gyrus

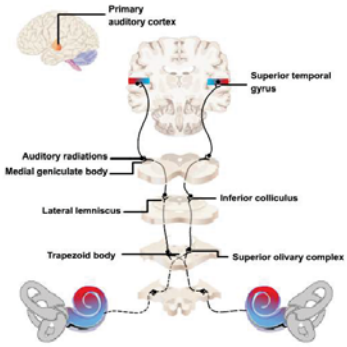
Contains the primary auditory cortex

Includes Wernicke's area

Functions include perception of emotions in facial stimuli and speech processing

http://commons.wikimedia.org/wiki/File:Cerebrum_-_superior_temporal_gyrus_-_lateral.png

Auditory Pathway



http://neuroscience.uth.tmc.edu/s2/chapter13.html

Linguistics Background

A phoneme is the smallest contrastive unit that changes the meaning of a word (e.g. /b/ and /d/ in "bad" and "dad")

Phoneme categories important in this study are *plosives*, *fricatives*, *vowels*, and *nasals*

The Plosives		
/p/	-	Push,Pin,Pat,Pop
/b/	-	bush,bin,bat,ball
/k/	-	cold,coat,coal
/g/	-	gold,goat,goal
/t/	-	tall,tusk,town
/d/	-	doll,dawn,down

The fricatives		
/f/	-	fan,phone,four
/v/	-	van,vast,vice
/θ/	-	think,method,north
/ð/	-	then,they,breathe
/s/	-	cycle,cite,site,
/z/	-	zip,zing,zoom
/ʃ/	-	ship,machine
/ʒ/	-	leisure,measure,treasure
/h/	-	hot,high,behave

The Nasals		
/m/	-	meal,meat,mean
/n/	-	neam,name,norm
/ŋ/	-	sing,ring,think,pink

http://www.english-for-students.com/English-Pronunciation-1.html

Previous Studies

Electrical stimulation of the STG disturbs speech processing ability in humans (Boatman et. al 1997)

Chang *et. al* found that neural responses in the auditory cortex showed categorical responses to continuous stimuli (*Nature Neuroscience* 2010)

Purpose

To answer:

How do neurons in the STG respond to speech sounds with different phonetic characteristics?

What features are speech-processing neurons in the STG selective for?

How is selectivity for those features organized in the STG?

Participants

6 human participants

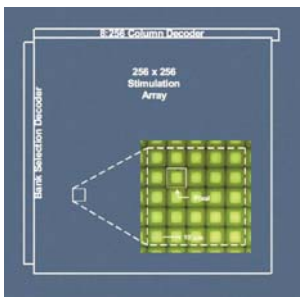
Clinical evaluations for epilepsy surgery

Patients were taken off of seizure medication to lower seizure threshold during study and to make sure medication was not affecting results

5 language dominant and 1 non-language dominant hemisphere were recorded from

The Wada test was used to determine which hemisphere was language-dominant in each subject

Electrodes



A 256-by-256 CMOS Microelectrode Array for Extracellular Neural Stimulation of Acute Brain Slices (Yuste et. al)

High-density subdural electrodes were implanted on the STG as part of a clinical epilepsy treatment procedure

Implantation sites were chosen based on response to speech versus silence determined in a t-test

Flat, planar electrode arrays were used to prevent tissue damage. No single-unit recordings.

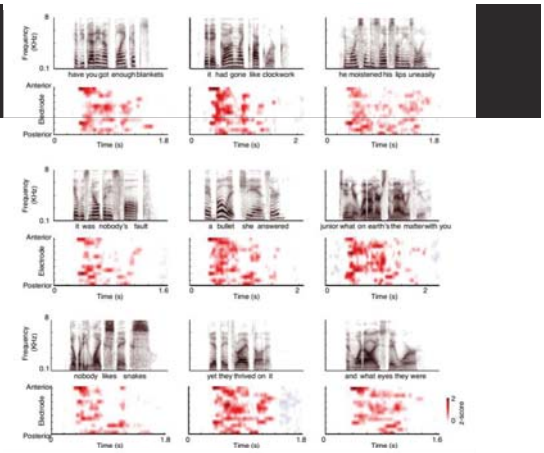
Speech Samples

The subjects listened to 500 English sentences spoken by 400 American English speakers

Speech samples were obtained from TIMIT Acoustic-Phonetic Continuous Speech Corpus, which is commonly used in speech and communication research

All TIMIT samples have been delineated in time based on acoustic-phonetic features

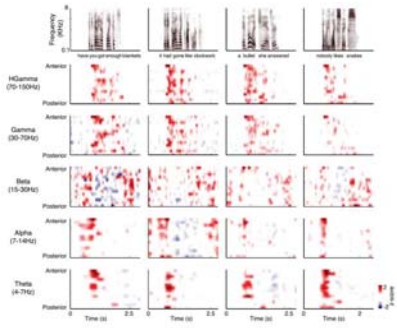
Figure S1



TIMIT sentences broken up into separate syllables to examine response to phonemes

Figure S2

High Gamma Frequency



1) Temporal resolution to distinguish between different phonemes

2) High correlation with neuronal spiking

3) Most consistent auditory response

High gamma were most effective frequency for measuring response

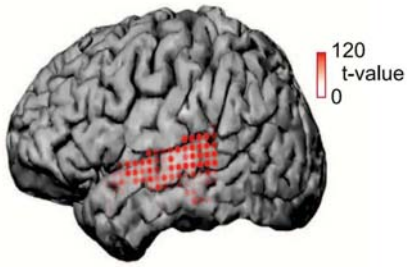
Organization of Paper

The paper is organized into 4 main figures, each showing a different broad category of the results.

Figure 1 - Selectivity

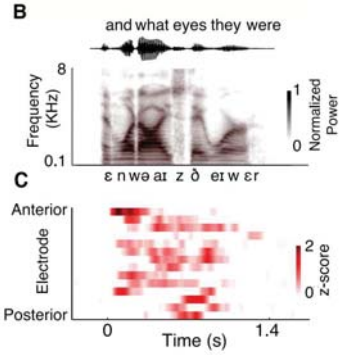
In Figure 1, the authors try to answer the question of how selectivity to speech sounds is represented in the STG by recording high-gamma frequencies from the electrodes.

Electrodes - Fig 1A



Example positioning of electrodes (red) on one participant's brain. Increased opacity corresponds to increased discrimination between speech and silence

Fig 1B and 1C



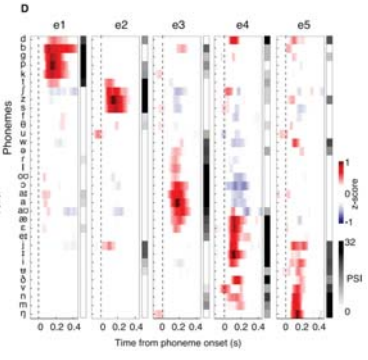
B
An example of an acoustic waveform and its spectrogram.

The horizontal axis shows the TIMIT phonetic transcription of the sentence.

C
Responses by electrodes to the same sentence.

z-score is a measure of deviation from the mean

Fig. 1 D



Mean neural responses at five example electrodes to all English phonemes

Intensity of the response is represented by the PSI and z-score

z-score is a normalized response

Phoneme selectivity index (PSI) vectors for the electrodes are shown alongside each plot

Phoneme Selectivity Index

Measure of the selectivity of each electrode

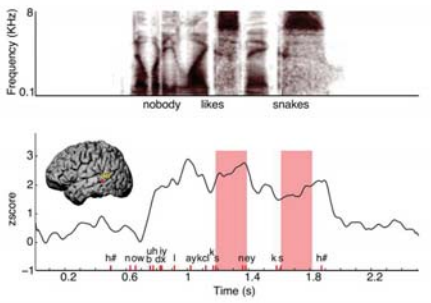
A number between 0 and 32 (because 33 total phonemes)

0: not at all selective, no other phonemes distinguishable at this electrode

32: completely selective

Based on statistical differences of response amplitudes

PSI - Fig. S3 A



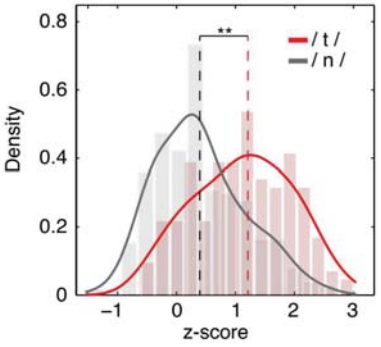
Example electrode (e5) **response** to sample sentence

Red bars indicate greatest response

Black line indicates z-score across time

Little red lines indicate timing of the sounds producing the sentence

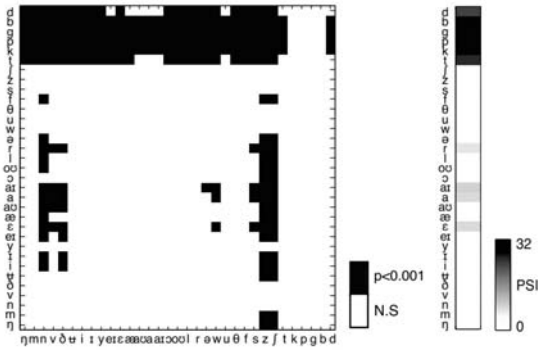
PSI - Fig. S3 C



Example of responses to the phonemes /t/ and /n/

Comparing all phonemes in this manner results in a PSI Index

PSI - Fig. S3 D and E



High Gamma and PSI

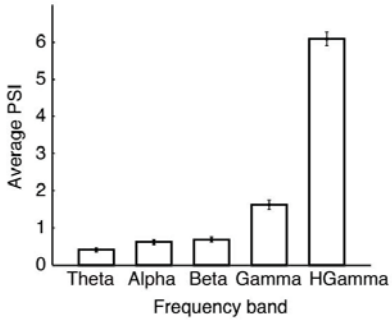


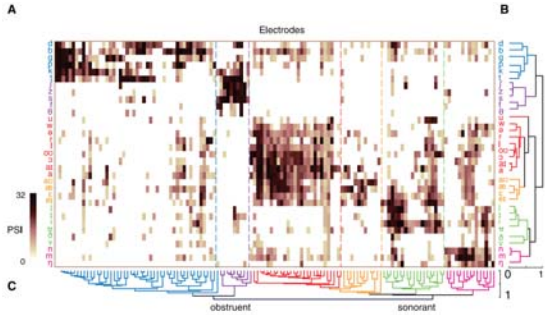
Fig. S5 C

Another reason to examine High Gamma frequencies

Fig. 2 - Clustering

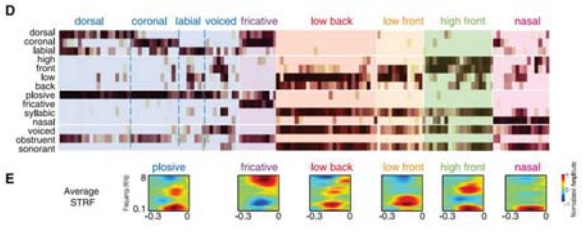
In Figure 2, the researchers determine how clusters of electrodes respond to clusters of phonemes. This provides a basic understanding of how we are able to differentiate between different sounds.

Electrode Responses Fig. 2



Electrode responses and phonemes are CLUSTERED

Responses Fig. 2 D-E



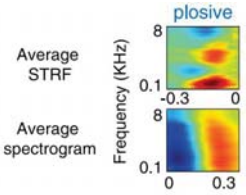
Clustering relates to FEATURES of phonemes
Same as how linguists already classified the phonemes

After finding these results, they wanted to understand which aspects of the phonemes electrodes can distinguish

Spectrotemporal Receptive Field (STRF)

Describes the response of a neuron to different frequencies of sound over time. Each neuron responds optimally to a certain frequency spectrum.

Used to predict neural responses in the STG

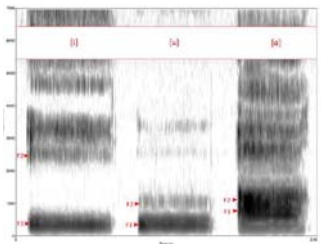


Example: Fig. 2E-F show the average STRFs of the main electrode clusters. Here, the STRF for plosive-selective electrodes is shown with time on the x-axis.

Figure 3 - Vowel encoding

In Figure 3, the researchers investigate how vowels are represented in the STG by comparing how vowels are distributed in acoustic space to how they are distributed in neural space.

Acoustic parameters of vowels

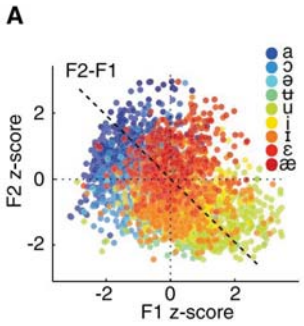


http://upload.wikimedia.org/wikipedia/commons/7/77/Spectrogram_-iua-.png

Vowels are defined by their fundamental (F0) and formant frequencies (F1-F4)

F1 and F2 are used in perception to distinguish different English vowels

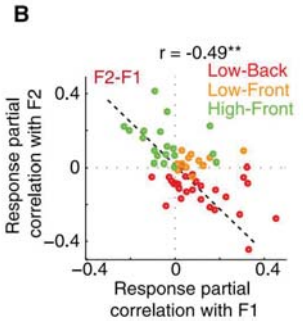
Fig. 3A



Optimal projection of the formant space of different vowels is the difference between F2 and F1

Consistent with vowel perceptual studies

Fig. 3B



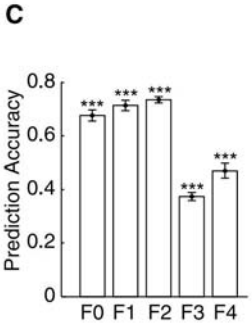
Sensitivity of STG sites to F1 was negatively correlated with sensitivity for F2 ($p < 0.01$, t test)

Dots represent electrodes

Shows an integrated response of each vowel-sensitive electrode to F1 and F2

Matches projection in Fig. 3A

Fig. 3C

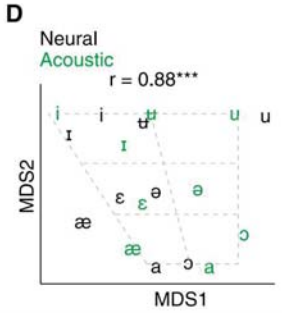


Decoding of fundamental and formant frequencies was accurate ($p < 0.001$, t test)

Authors used linear regression analysis to predict the relationship between acoustic parameters and neural responses

Error bars indicate standard error of the mean

Fig. 3D



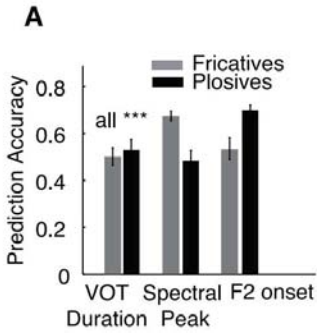
Multidimensional scaling shows that the organization of vowels based on acoustic parameters is similar to their organization in neural space ($p < 0.001$, t test)

Vowels close together in this figure produced similar neural responses, vowels far apart produced dissimilar responses

Figure 4 - Plosives and Fricatives

In Figure 4, the researchers try to determine how electrodes respond to three specific spectrotemporal features of sound and how this response differs between two different groups of phonemes.

Fig. 4A



Prediction accuracy of three plosive and fricative acoustic parameters from neural population.

Error bars indicate standard error of the mean.

Voice Onset Time

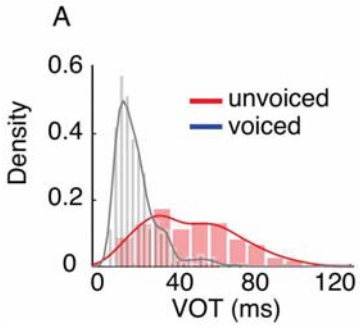


Fig. S10A

Distribution of voice-onset time for all plosive phonemes

A voiced sound is one in which the vocal cords vibrate.

Voiced plosives: /b/, /d/, /g/
Unvoiced plosives: /p/, /t/, /k/

Same analysis was done for fricative phonemes (Fig. S12)

Spectral peak

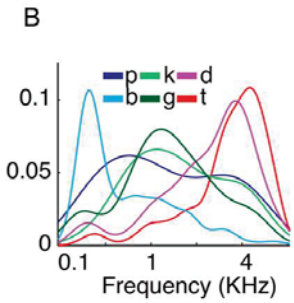


Fig. S10B

Acoustic spectrograms of phonemes were used to determine the frequency of maximum energy (spectral peak)

Distinguishes labials (/p/,/b/), coronals (/t/,/d/), and dorsals (/k/ and /g/).

F2 onset of the following vowel

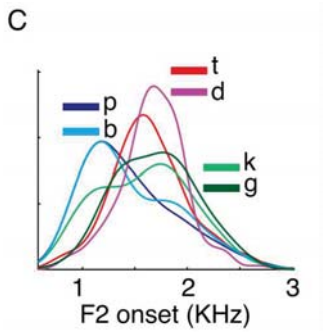
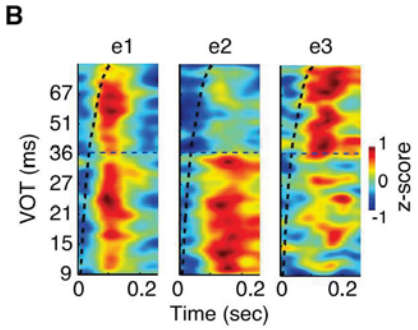


Fig. S10C

F2 onset of the vowel following each of the plosive consonants in the speech samples

Fig. 4B

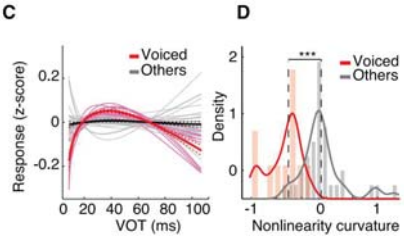


Selectivity of three electrodes for duration of voice-onset time

Plotted for all instances of plosive phonemes in the speech samples (1200)

The curved dotted line delineates voice-onset time duration

Fig. 4 C and D



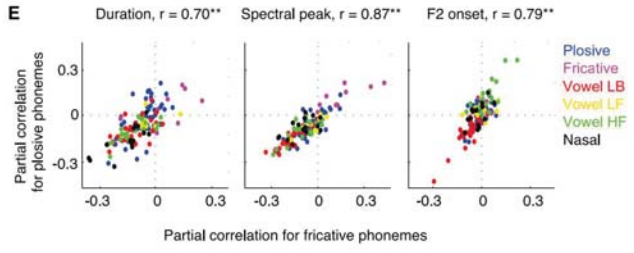
Voiced plosive electrodes have more of a bias for short VOTs than non-voiced do.

In D, negative nonlinearity curvature means there was a preference for short VOTs, while positive indicates a preference for longer VOTs.

Means there is a **specificity** for electrodes regarding VOT, and that this is related to the phoneme being voiced or non-voiced.

Contrasts with previous results in other studies

Fig. 4 E



Answers the question: does electrode response vary due to type of phoneme (fricative vs. plosive)?
NO - response is based off of the features, not based on the overall category.
So the information of the features must be synthesized to create the understanding of the overall category.

Discussion & Conclusions

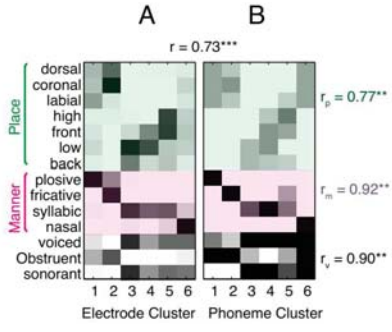


Fig. S8

Significant correlation between average PSI values for the six main electrode clusters and average PSI values for phoneme clusters

Selectivity is determined primarily by acoustic cues for manner of articulation, and secondarily by place of articulation

Discussion & Conclusions

How do neurons in the STG respond to speech sounds with different phonetic characteristics?

There is selectivity for certain features of speech sounds within populations of neurons in the STG.

What features are speech-processing neurons in the STG selective for?

Acoustic-phonetic features of sound, as well as spectrotemporal features such as voice-onset time. Selectivity is determined mainly by manner of articulation of a speech sound, and secondarily by place of articulation.

How is selectivity for those features organized in the STG?

Higher-order processes involved in speech sequencing, such as determining voice-onset time, are spatially distributed. The more basic process of distinguishing phonemes is localized.

Discussion Questions

1. Describe the superior temporal gyrus. Why did the researchers choose to focus on this area?

2. What is a phoneme? Define voice-onset time (VOT), spectral peak, and formants (F0, F1-F4).

3. Why did the authors choose to analyze high-gamma cortical surface field potentials? How do their measurements differ from single-unit recordings?

4. What are the implications of this study? Consider the fact that Edward Chang is interested in epilepsy and neuroprosthetics.

Discussion & Conclusions

Speech representation in the auditory cortex is governed by the same acoustic-phonetic features that are used to describe aspects of speech sounds in linguistics.