Neural Pathways of Vocal Pitch Modulation: An ongoing fMRI Study

Minkyu Kim[1], Gregory Hickok[1,2]

Auditory & Language Neuroscience Lab

[1]Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA [2]Department of Language Science, University of California, Irvine, CA 92697, USA

UCI University of California, Irvine

ABSTRACT ***Humans' ability to voluntarily control pitch*** *is central to language function. This research investigates the neural mechanisms of vocal pitch control using* ***functional MRI****, focusing on two experimental approaches. The first examines neural circuits involved in* ***pitch modulation*** *during* ***speech repetition tasks,*** *comparing full speech rehearsal to* ***humming prosodic patterns****. The second aims to differentiate* ***pitch-specific pathways*** *from general speech networks by analyzing neural responses to speech repetition under* ***Altered Auditory Feedback*** *(AAF) with perturbations in* ***fundamental frequency*** *(F0) and the* ***first formant*** *(F1). Early findings suggest the existence of* ***two anatomically distinct pathways*** *in the* ***premotor cortex****: one pitch-related and the other articulation-related. These insights have implications for understanding vocal pitch in communication and potential clinical interventions.*

Introduction

Pitch plays an essential role in communication

Not only for **conveying information** but also for expressing **emotions** and **social signals**. It is also extensively used in **singing** (musical expression).

Recent studies emphasize that **pitch and prosody** are deeply integrated with **syntax, semantics, phonological rhythm,** and **pragmatic meaning**.

Humans show a unique ability to control pitch

Humans exhibit an **unparalleled capacity** for controlling pitch, which may have links to the **evolution of language and singing** (Fitch, 2016).

While nonhuman primates possess vocal tracts capable of producing a wide range of sounds (Boë et al., 2019), they lack the necessary **neural mechanisms** to control these sounds for **speech** (Fitch, 2016).

Hypothesis for two functionally distinct speech coordination areas in premotor cortex

Dichter et al. (2018) showed that bilateral **dorsal laryngeal motor cortex** (dLMC) selectively encode **produced pitch** but not non-laryngeal articulatory movements by high-density intracranial recordings.

Hickok et al. (2022) proposed **two distinct regions** in the **precentral gyrus** (main motivation of this research):

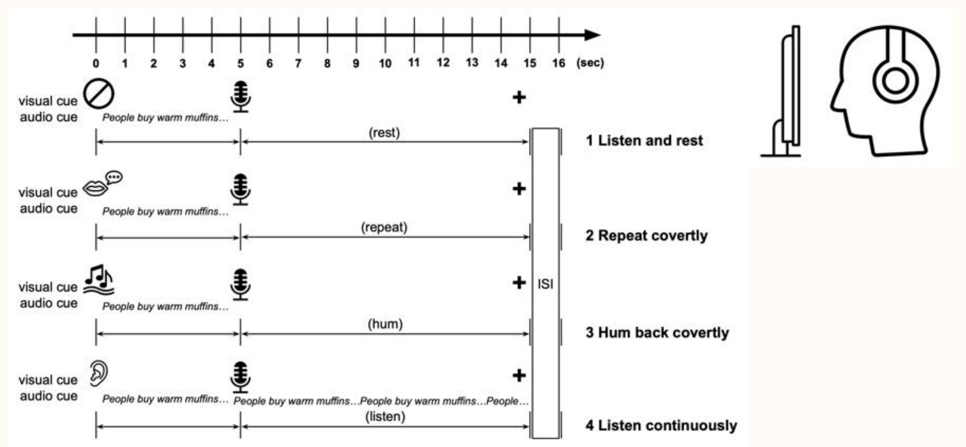
dorsal precentral speech coordination area (**dPCSA**)	**ventral** precentral speech coordination area (**vPCSA**)
auditory-weighted sensorimotor control	**somatosensory**-weighted sensorimotor control
prosodic, pitch-related, laryngeal effector	**syllabic, phonetic**-related, supralaryngeal effector

RQ: Are there distinct neural pathways for pitch modulation vs supralaryngeal articulation? Can we differentiate them?

Experiment 1
Repetition of Prosodic Patterns

TL;DR You listen to sentences and either **repeat** them or **hum** their prosodic patterns in the scanner. This will allow us to break down the **auditory-motor network** into distinct sub-networks for **pitch control** and (supralaryngeal) **articulatory movements**.

Paradigm



Each run = 25 trials (1 dummy + 3 cond × 8 trials) = 7 min. → Each session = 8 runs = 1 hr.

Each session has 64 trials per condition (**Blocked**)

Stimuli

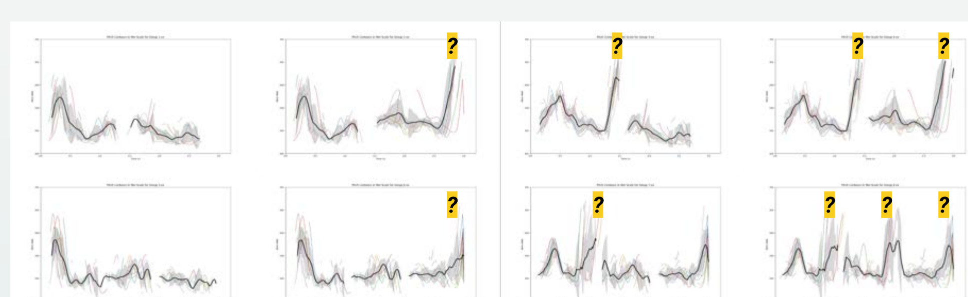
64 Machine-generated **sentence-level** speech sounds of 3–3.5 s duration with **maximized prosodic variability & phonetic variability**

Labial	*People buy warm muffins, after morning meetings.*
Coronal	*Teachers instruct subjects, through engaging lessons.*
Dorsal	*Kangaroos carry young, hopping across grasslands.*
Nasal	*Mountain meets horizon, in misty morning light.*
Stop	*Octopus attacked ships, during big ship attack.*
Fricative	*Fossil features findings, beneath the surface earth.*
Lateral	*Uncle helped little girl, telling lovely stories.*
Rhotic	*Mirror reflects bright light, under rare auroras.*

Analyzed with **eng-to-ipa panphon** & **Praat**

Pattern 1	*People buy warm muffins, after morning meetings.*
Pattern 2	*People buy warm muffins, after morning meetings?*
Pattern 3	*People buy warm muffins? after morning meetings.*
Pattern 4	*People buy warm muffins? after morning meetings?*
Pattern 5	*Baby makes mess, with messy hands, in warm bedroom.*
Pattern 6	*Baby makes mess, with messy hands, in warm bedroom?*
Pattern 7	*Baby makes mess? with messy hands, in warm bedroom?*
Pattern 8	*Baby makes mess? with messy hands? in warm bedroom?*

Written with **GPT-4o** Text-to-Speech by **Google TTS**



Data Analysis

Auditory-motor network is identified by the conjunction of **[listen > repeat]** and **[repeat > rest]**

Pre-processing (fmriprep) / GLM (AFNI)

To do: **Representational Similarity Analysis** (RSA)

Expected Outcomes

Repetition Pathway: **vPCSA + dPCSA**
Humming Pathway: **only dPCSA**

Prosodic variability correlates with the **dPCSA**
Phonetic variability correlates with the **vPCSA**

within-subject comparison

Experiment 2
Altered Auditory Feedback: F0 vs F1

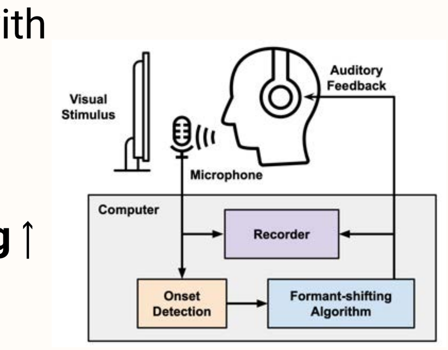
TL;DR While reading words on a screen, your voice is **altered in real time**, with either **pitch** (F0) or **vowel quality** (F1) manipulated and **fed back to your ears**. This will help us find the regions responsible for **pitch control** (source) and **vowel production** (filter).

Experimental Setup

MR-safe audio equipment with active **noise-cancellation** system (Optoacoustics)

EPI sequence modified to reduce noise: **Echo Spacing** ↑ (final noise level ≈ 65 dB)

Pitch/Formant-shifting algorithm ***Audapter*** (Cai et al., 2008; Tourville et al., 2013)



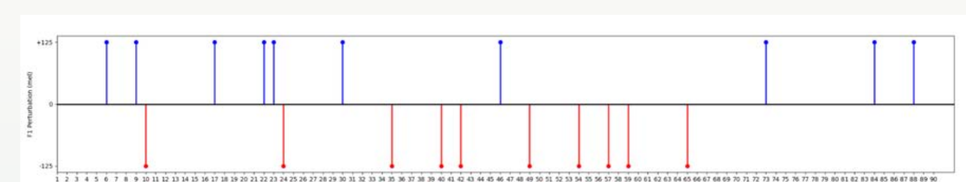
Paradigm

Following Niziolek (2021) F1 AAF Paradigm

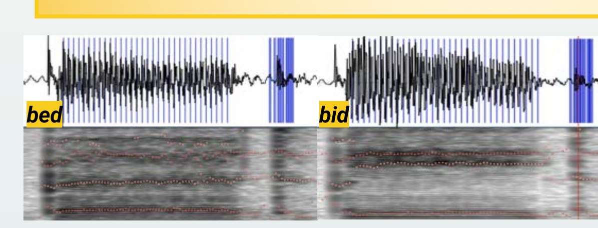
Visual stimuli: 3 English words -ɛd: *bed, dead, head*

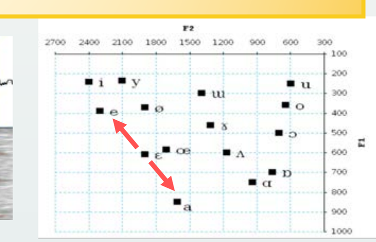
F0 perturbation ±6% / **F1 perturbation** 125 mel (45°)
⅙ up-shift + ⅙ down-shift + ⅔ control
total 90 trials per run (= 6 min) / total 8 runs
(**Rapid jittered event-related design**)

Visual feedback (colored dots) indicating whether vowel duration was too short, enough, or too long



A **formant** is a peak in the speech frequency spectrum, representing **resonant frequencies** in the vocal tract. Shifting F1 can change **vowel perception**; for example, shifting F1 up in ***bed*** /bɛd/ makes it sound like ***bad*** /bæd/, and shifting it down makes it sound like ***bid*** /bɪd/.





Data Analysis

Behavioral data to be first analyzed to confirm the **compensation** phenomena under AAF of F0 and F1

ROI to be localized using GLM (AFNI)

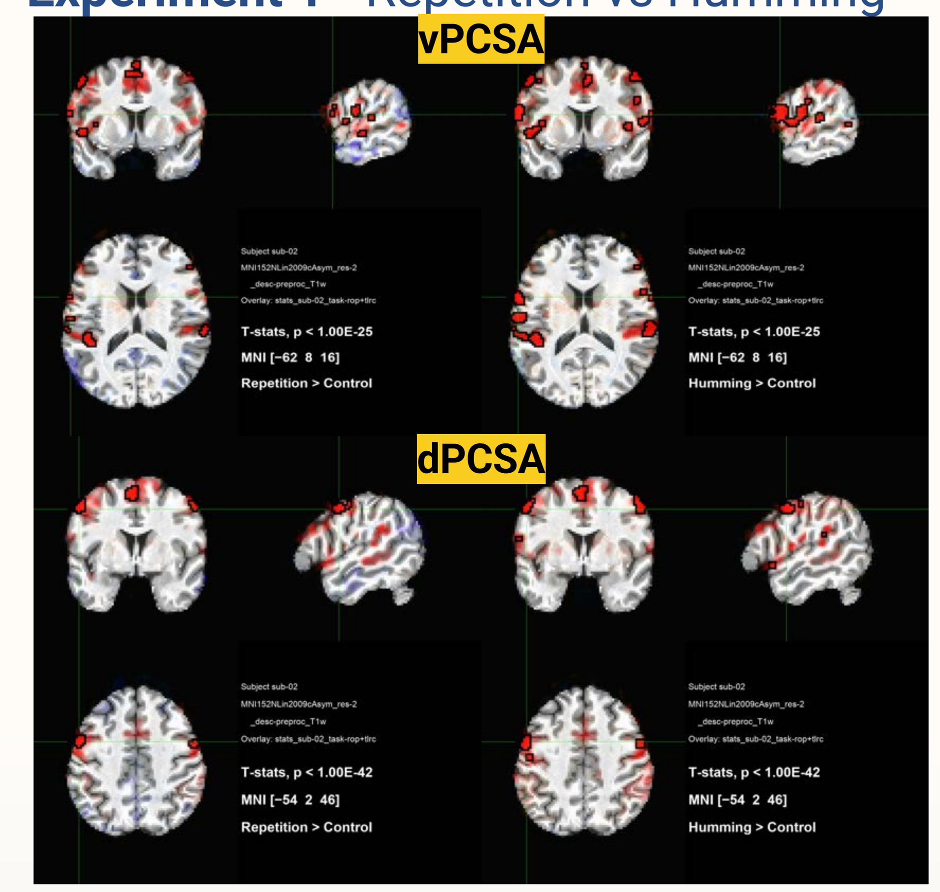
Expected Outcomes

F0 shift preferentially activate the **dPCSA** region
F1 shift preferentially activate the **vPCSA** region

Using the **same participant pool** for both experiments will enable a **within-subject comparison**, which strengthens our claims regarding the functional roles of these regions.

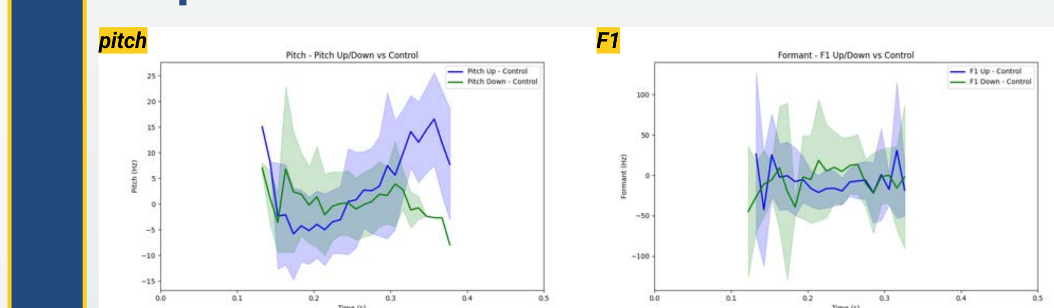
Results

Experiment 1 Repetition vs Humming



- **Localized** the regions we believe correspond to the **vPCSA** and **dPCSA**. (Individual level)
- The **humming condition** showed **stronger activation** in the **vPCSA**. (⊥ Our hypothesis)
- **Little to no difference** observed in the **dPCSA**.

Experiment 2 AAF F0 vs F1



- The **behavioral data is insufficient** to confirm a replicable **compensatory response** at this stage.

Next Steps

- Revise Experiment 1 to **reduce the vPCSA load** during the **humming condition**.
- Collect **more behavioral data** and further fine-tune the Experiment 2 paradigm.

Acknowledgement

I greatly appreciate the members of ALNS for their invaluable support, insightful discussions, and contributions to this project, with special thanks to Dr. Jon Venezia and Dr. Haleh Farahbod for their exceptional assistance.

Bibliography

Boë LJ et al., *Sci Adv.* **2019**;5(12):eaaw3916.
Cai S et al., Proceedings of the 8th ISSP (**2008**): 65-68.
Dicther BK et al, *Cell.* **2018**;174(1):21-31.e9.
Fitch WT et al., *Sci Adv.* **2016**;2(12):e1600723.
Hickok G et al., *Brain.* **2023**;146(5):1775-1790.
Niziolek CA et al., *J Speech Lang Hear Res.* **2021**;64(6S):2169-2181.
Tourville JA et al., *Lang Cogn Process.* **2011**;26(7):952-981.

visit my website!