Pitch Perception in Modified Sinewave Speech and Cochlear Implants

Havi Pham

Jonathan Nissenbaum & Katrien Vermeire

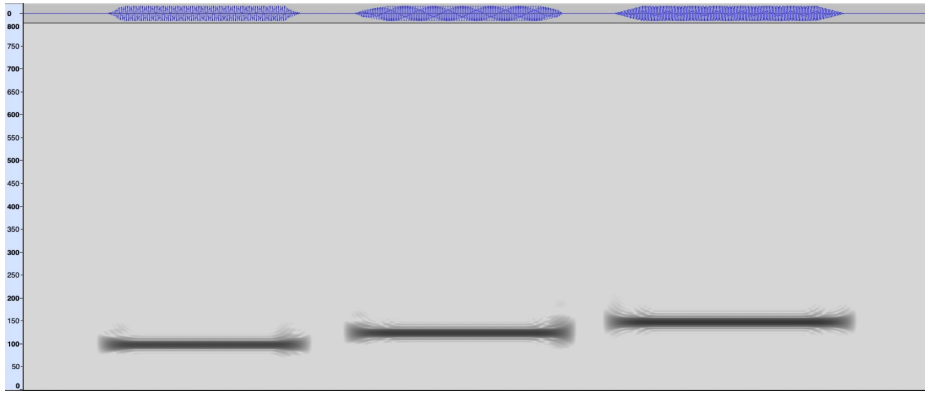
SMA Award # 2050922

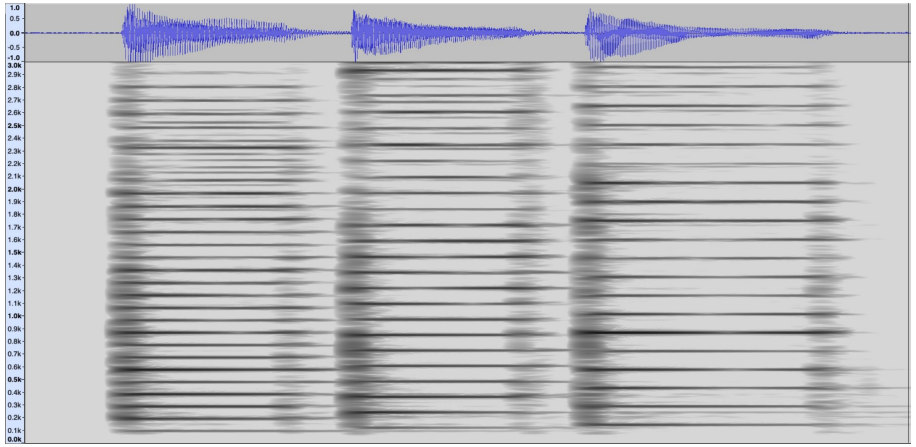
National Science Foundation and co-funded by US Department of Defense Air Force Office of Scientific Research ASSURE program

Pitch *Perception*

Physical aspects of signal that are known to correlate with pitch:

- ***fundamental frequency,*** (fundamental): the lowest frequency of a periodic waveform, or a pure tone (upper right)
- ***harmonic structure***: if the periodic wave is complex, the pattern of harmonic frequencies at multiples of the fundamental (lower right)

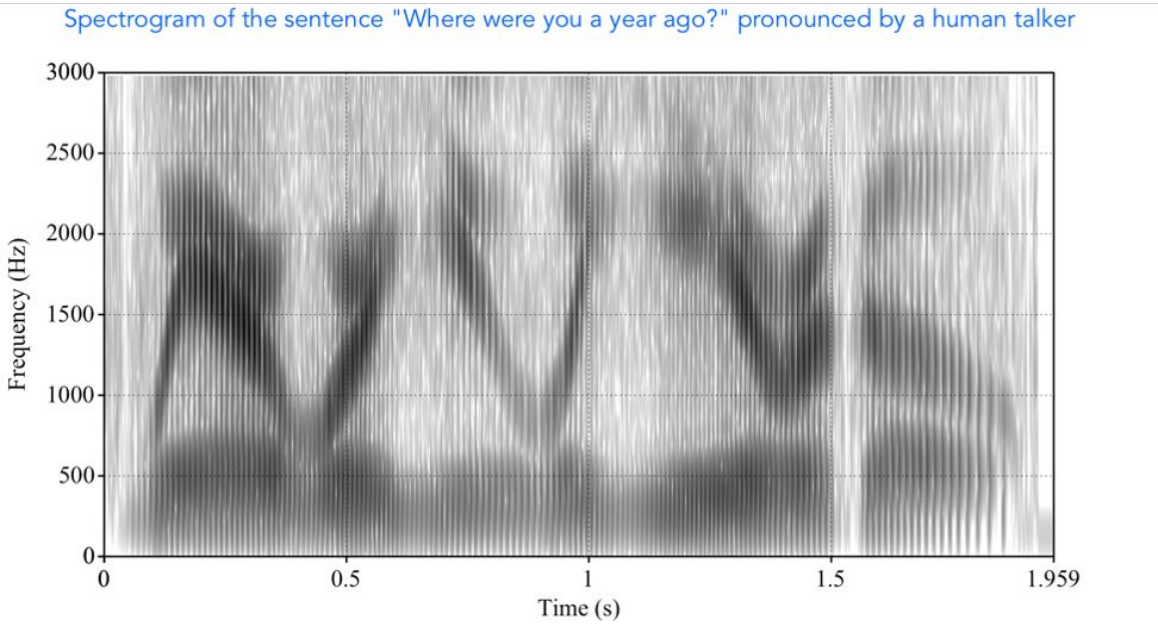




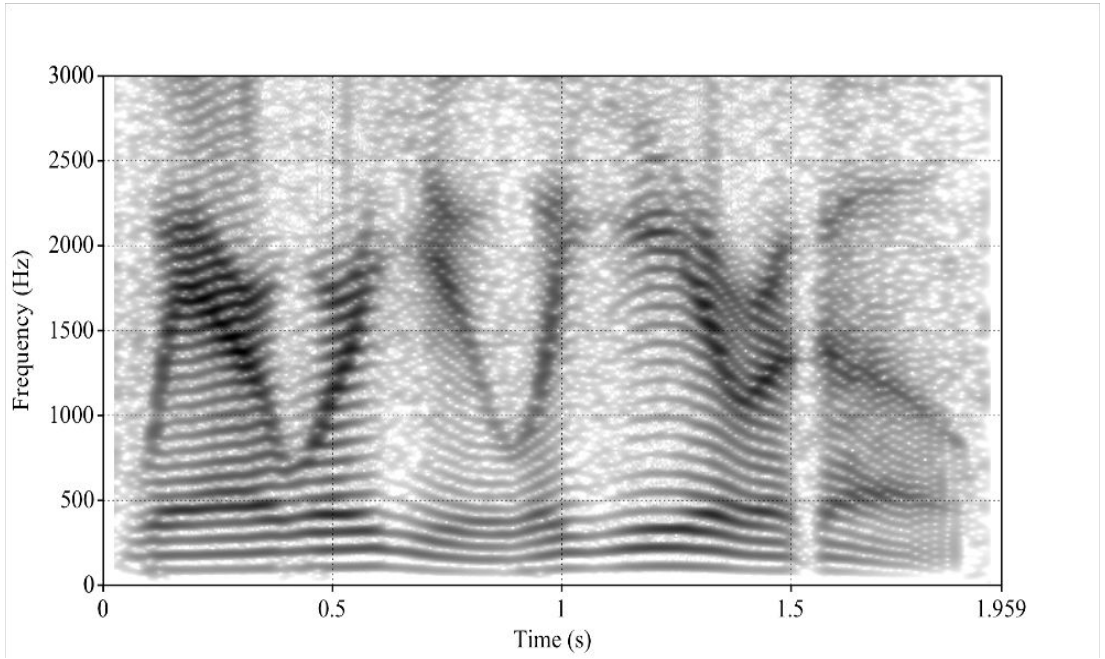
Pitch, One of Two Important Aspects of Speech:

- Vowels & Consonants Distinct from Pitch
- Pitch can be perceived independently of Vowels & Consonants, vice versa.
- Parts of the Physical signals for Consonants & Vowels --> formants, darkened energy
  - [wideband spectrogram]
- Physical signal for Pitch is the harmonic structure, distinctive from formants
  - [narrowband spectrogram]

Spectrogram Comparisons



Wide Band Spectrogram



Narrow Band Spectrogram

Introduction to Pitch Perception Research

- **Background Question:** ***What are the mechanisms that lead us to perceiving pitch?***

- Clinical Significance: CI provides strong segmental information (consonants & vowels of speech) but does a poor job of conveying pitch percept.

- ***General Research Question: What explains the lack of pitch perception in CI users?***

“Dominance Region” in Relation to Pitch Cues:

- The normal speech signal contains an abundance of harmonic information to support pitch perception.

- Could it be that cochlear implants overwhelm the auditory system by providing *too much* harmonic information?

- If so, the poor pitch perception of CI users could amount to a signal-in-noise problem

- *“dominance region”* ; range of frequencies for which the human auditory system is most attuned for pitch cues; 200 hz - 1,000 hz (Remez-Rubin)

A Different Stimulus

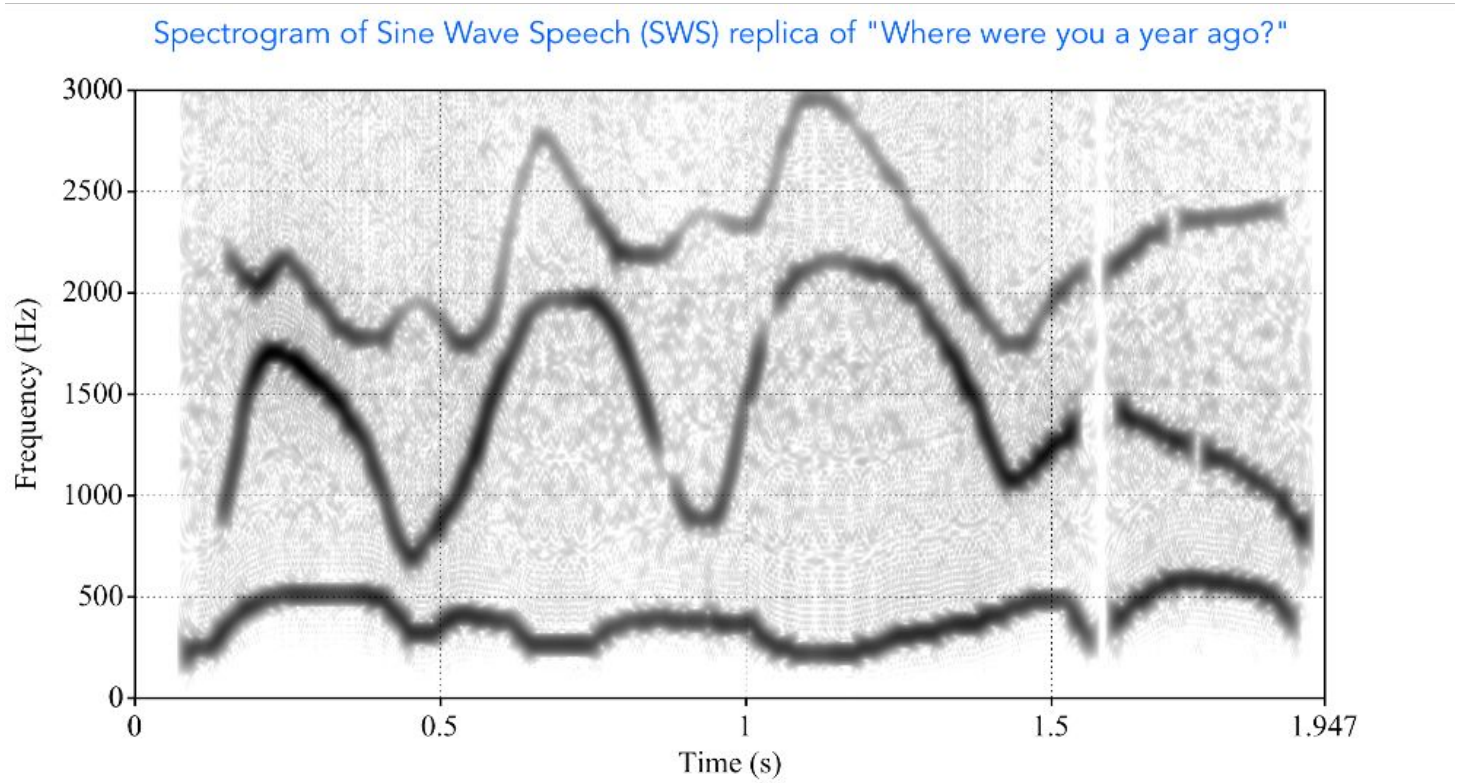
- ***Specific Research Question: Will modified SWS help CI users perceive pitch better?***

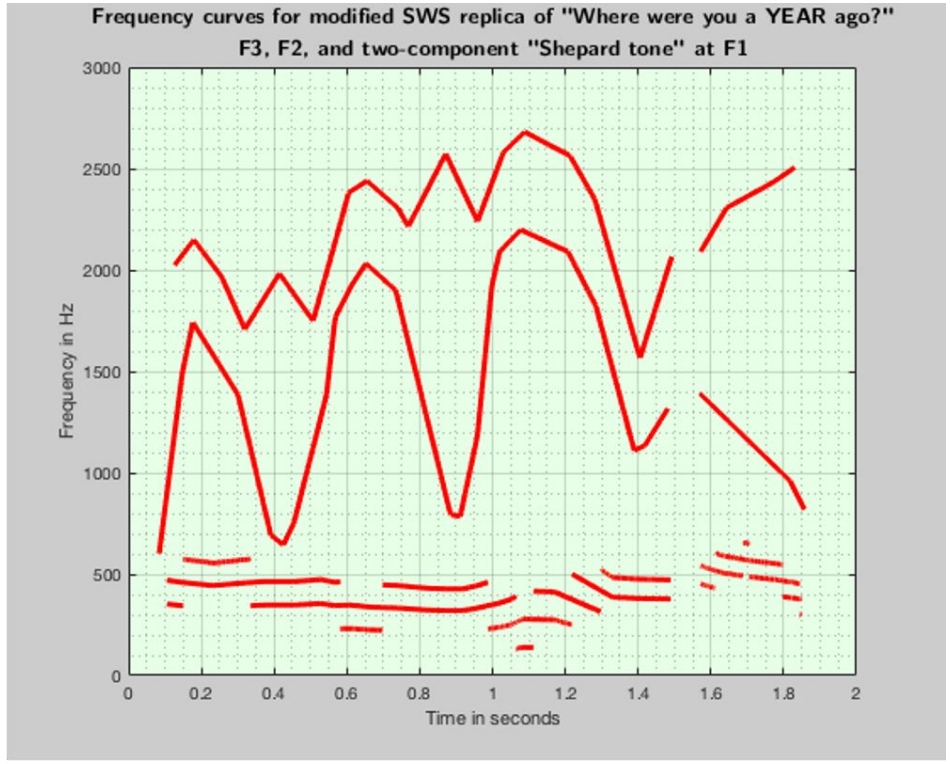
- Modified Sinewave Speech: a method for creating speech signals with a very narrow cue for pitch inside the *dominance region*
- *Hypothesis:*

  Perhaps, the overabundance of information (i.e: not being directed specifically to the dominance region) causes the lost of pitch perception information. Attunement to a narrow part of the frequency range.

SWS vs. Modified SWS

- SWS Stimulus will be outlining the lowest formant





Methodology for Study

- We used a modified SWS stimulus that has a sparse cue for pitch. Pitch cue created in range of dominance region, f1; all other harmonics are omitted.
- We synthesized modified SWS sentences and presented them to 30 normal hearing listeners; ages 18 -42
- Study Timeline:

  Phase 1:Normal Hearing (current study)

  Phase 2: Cochlear Implant Users (projected follow up)

Stimuli & Listening Tasks

Consisted of 10 sentences with pitch focus on question answer congruence, focusing adverbs and reasons & counterfactuals. (Rooth, 1996)

Question Answer Congruence	*Question Answer Congruence*
→ PETE* wants soda Pete wants SODA*	→ MARY* brought wine Mary brought WINE*
Which answers are most fitting to their questions: Who wants soda? Does Pete want water or soda?	Which answers are most fitting to their questions: Who brought wine to the party? What did Mary bring to the party?

**capitalized words indicate pitch focus on word

Stimuli & Listening Tasks (cont.)

Focusing Adverbs

- ➔ Sarah only put BOOKS* on the table
- ➔ Sarah only put books on the TABLE*
  - ◆ Yes/No Questions

e.g.:
Could Sarah have put anything else on the table?

Could Sarah have put the books on the floor?

- ➔ Joe only met MARY* for dinner
- ➔ Joe only met Mary for DINNER*

Reasons & Counterfactuals

- ➔ Don isn't speeding because someone is tailing him
- ➔ Don isn't speeding because someone is TAILING* him
  - ◆ Scenario Based Question

**capitalized words indicate pitch focus on word

Participants & Listening Tasks

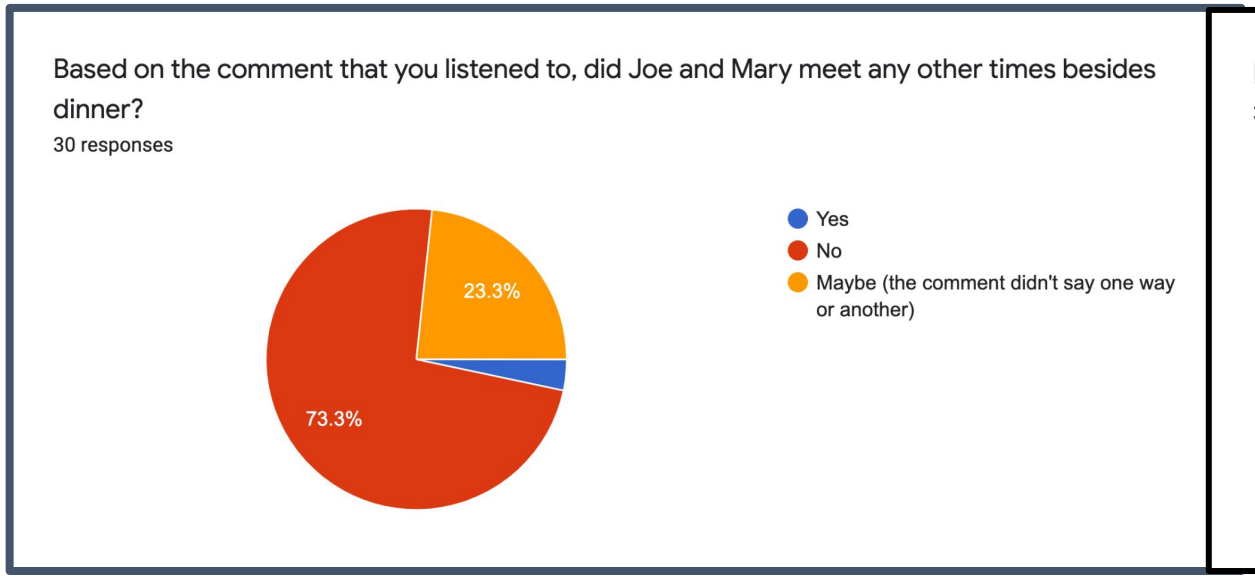
- 30 Participants between the ages of 18 - 42 completed 14 listening tasks
- Each listening task included:
  - a modified SWS recording
  - question on the screen based on the stimulus that they heard
- Experiment distributed through Google Forms

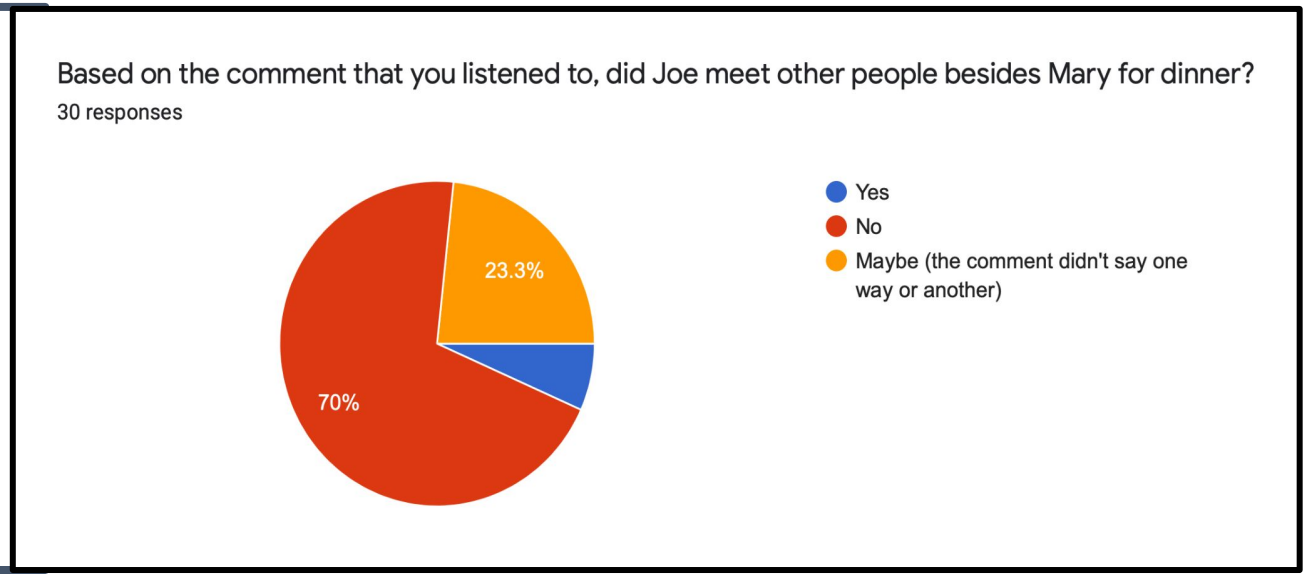
Expected Outcomes

Expected Results dependant on consistency + ‘accuracy’.

- Pitch cues expected to be heard if participants displayed a consistent answer picking from their individual surveys
- How most people did on a question informative of what a majority of people could be hearing
  - If normal hearing participants are able hear pitch cues accurately, then this gives us a baseline to measure for phase two of research.

Results (part 1)





Joe only met Mary for <u>DINNER</u>

Joe only met <u>MARY</u> for dinner

Results for two sample tasks from one set

Results (continued)

– For each of the five sets of related tasks:

- We excluded any participant from the analysis if they gave identical answers on all tasks within a set (since those answers could not give us any information)
- This method caused us to entirely exclude one of the five tasks from the analysis (due to the overabundance of identical pairs of answers, from 26 out of 30 participants(!))
- The other four sets of tasks yielded statistically significant results.

Results (continued): 2 sets of tasks using focus with ONLY

SET 1: "Joe only invited Mary for dinner" [19 participants]

Auditory Stiumulus	VISUAL STIMULUS: "Did Joe and Mary meet at any other times besides dinner?" Expected:	Actual:	VISUAL STIMULUS: "Did Joe meet any other people besides Mary for dinner" Expected:	Actual:
FOCUS on *dinner*	NO	No: 15 Maybe: 3 Yes: 1	MAYBE	No: 7 Maybe: 11 Yes: 1
FOCUS on *Mary*	MAYBE	No: 8 Maybe: 11 Yes: 0	NO	No: 12 Maybe: 5 Yes: 2

SET 2: "Sarah only put books on the table" [19 participants]

Auditory Stiumulus	VISUAL STIMULUS: "Could Sarah have also put some Books on the floor?" Expected:	Actual:	VISUAL STIMULUS: "Could Sarah have also put a box of crayons on the table?" Expected:	Actual:
FOCUS on *table*	NO	No: 12 Maybe: 6 Yes: 1	MAYBE or YES	No: 6 Maybe: 5 Yes: 8
FOCUS on *books*	MAYBE or YES	No: 6 Maybe: 5 Yes: 8	NO	No: 11 Maybe: 4 Yes: 4

Results (continued): 2 sets Question-Answer matches

SET 1: "Pete wants soda" [23 participants]

Auditory Stiumulus	"Who wants soda?"		CHOICE: "Does Pete want water or soda?"		"Either one"	
	Expected:	Actual:	Expected:	Actual:	Expected:	Actual:
FOCUS on *Pete* Task 1	YES	9	NO	5	NO	9 (!)
FOCUS on *Soda* Task 8	NO	2	YES	8	NO	13 (!)

SET 2: "Mary brought wine" [22 participants]

Auditory Stiumulus	"Who brought wine?"		CHOICE: "What did Mary bring?		"Either one"	
	Expected:	Actual:	Expected:	Actual:	Expected:	Actual:
FOCUS on *Mary* Task 12	YES	5	NO	5 (!)	NO	12 (!)
FOCUS on *wine* Task 5	NO	1	YES	10	NO	11 (!)

Discussion

The results were mixed, but mostly promising:

- For the two sets using **focus with "*only,*"** the results were largely in line with what we expected
- For the two sets of **Question-Answer matches,** the results were mixed:
  - Very few subjects matched the auditory stimuli with the "wrong" (unexpected) question.
  - However, the high number of **"Either one"** answers was unexpected.
- The **negation-focus** task was unsuccessful at eliciting informative results

What happens next?

Based on results:

- Suggest that artificial pitch can be heard by normal hearing listeners
- Experiment design/task can be improved
- Create a control condition for Phase 1 using normal speech, spoken by a person rather than artificial stimuli
- How hard it is to hear pitch in our modified SWS vs. How hard the task is to do (even with natural speech)

Thank You!

Any questions or comments?

References:

Andrew Oxenham (2021) “Perception and Neural Coding for Pitch for Speech and Music.” Presented at Acoustical Society of America webinar series, July 15th, 2021., https://acousticalsociety.org/asa-webinar-series/

Fan-Gang Zeng (2021) “Celebrating the 1 Millionth Cochlear Implant.” Presented at Acoustical Society of America webinar series, August 18th, 2021., https://acousticalsociety.org/asa-webinar-series/

Remez & Rubin (1984), “On the perception of intonation from sinusoidal sentences." Perception & Psychophysics 35 (5), 429-440.

Rooth, Mats (1996). Focus. In Shalom Lappin (ed), The handbook of contemporary semantic theory, 271-298. Oxford, Blackwel

Tamati TN, Ray C, Vasil KJ, Pisoni DB, Moberly AC. High- and Low-Performing Adult Cochlear Implant Users on High-Variability Sentence Recognition: Differences in Auditory Spectral Resolution and Neurocognitive Functioning. J Am Acad Audiol. 2020 May;31(5):324-335. doi: 10.3766/jaaa.18106. Epub 2020 Jun 9. PMID: 31580802; PMCID: PMC7103548.