Abstract 5aSC19: Where is the Beat?

Lauryn Zipse,[1] Olivia Murton,[2,3] Katie Dunn,[2] & Stefanie Shattuck-Hufnagel[2]

[1]MGH Institute of Health Professions, School of Health and Rehabilitation Sciences

[2]Massachusetts Institute of Technology, Speech Communication Group, Research Laboratory for Electronics

[3]Harvard University, Program in Speech and Hearing Bioscience and Technology

INTRODUCTION

Most people can easily entrain their movements to a beat, whether dancing, tapping to music, or singing.

- Entrainment has primarily been examined through experiments in which people are asked to tap along to a metronome.
- A consistent finding: the Negative Mean Asynchrony (NMA)
  - Taps precede the stimuli in the sequence by tens of milliseconds.[1]
  - NMA may be due to neural transmission times, e.g. ear to brain is shorter than brain to hand.
  - NMA is larger for foot tapping than hand tapping.[2]

Less is known about speech entrainment

- An important area: entrainment has been used to improve speech fluency for people with stuttering, apraxia of speech, dysarthria, and aphasia.[3,4]
- Evidence that the moment of occurrence of a syllable (P-center), is near the onset of voicing for the vowel when a person hears speech.[5]

Given the clinical importance of entrainment as a means to facilitating speech, it is important to ask whether the beat location depends on whether a speaker is ***listening to*** or ***producing*** a syllable.

RESEARCH QUESTIONS:

Q1: Where in the syllable do people perceive the beat when ***listening to*** an isochronous sequence of CV syllables?

Prediction: A few tens of milliseconds before the onset of voicing. ***Voicing onset is the most acoustically salient point***; NMA means tap precedes beat location.

Q2: Where in the syllable do people perceive the beat when ***producing*** an isochronous sequence of CV syllables?

Prediction: On or slightly before the consonant burst. ***The burst is the most somatosensorially salient point***, and any NMA will be small since the effectors for speech are close to the brain, resulting in a short neural transmission time.

METHODS

Participants

Thirty-one neurotypical college students completed the task. Data from 11 participants were not reliably scored by the automated algorithms used. As a result, the final data set contained data from 20 individuals:

- Mean age = 19.3 years (SD = 1.4 years)
- 5 males and 15 females
- All with native-speaker-level proficiency in English

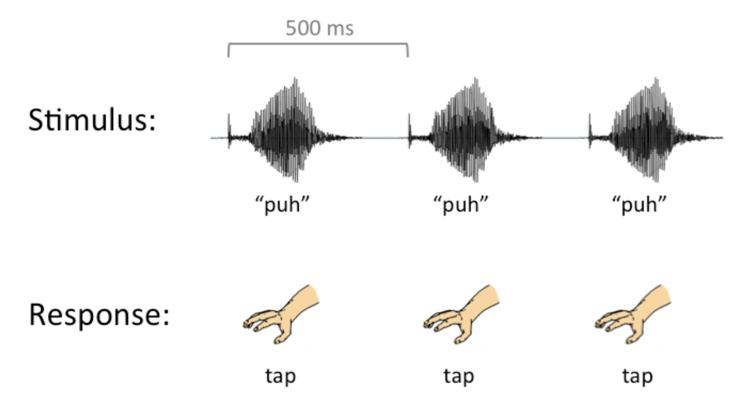
Procedure

Participants sat in a sound-proof booth with headphones through which stimuli were delivered. A microphone for recording spoken responses was in a stand in front of the participant. The participant held a second microphone in one hand and tapped on it with the other hand.

- Two conditions of interest presented pseudorandomly along with a third condition (not reported here), with 24 trials of each condition
- Each trial consisted of an orienting tone, a short instruction (i.e., "tap" or "puh"), and 20 repetitions of the auditory stimulus.

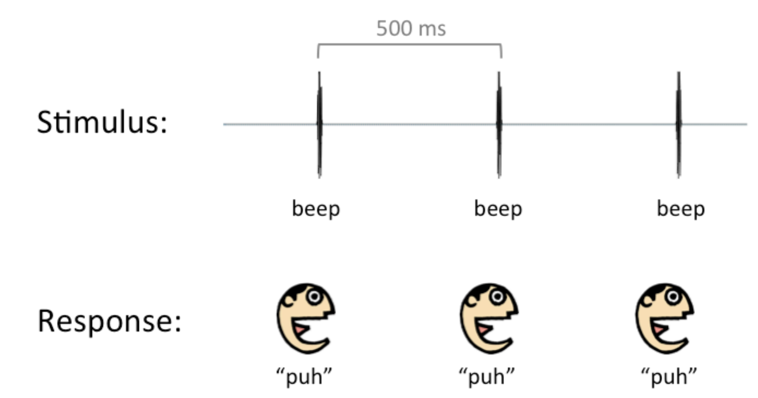
Condition 1, Tapping to Speech

Participants tapped to an isochronous sequence of the syllable "puh" (mean 182 Hz, 27 ms duration)



Condition 2, Speaking to Beeps

Participants said "puh" to an isochronous sequence of beeps (442 Hz, 15 ms duration)



Labeling and Analysis

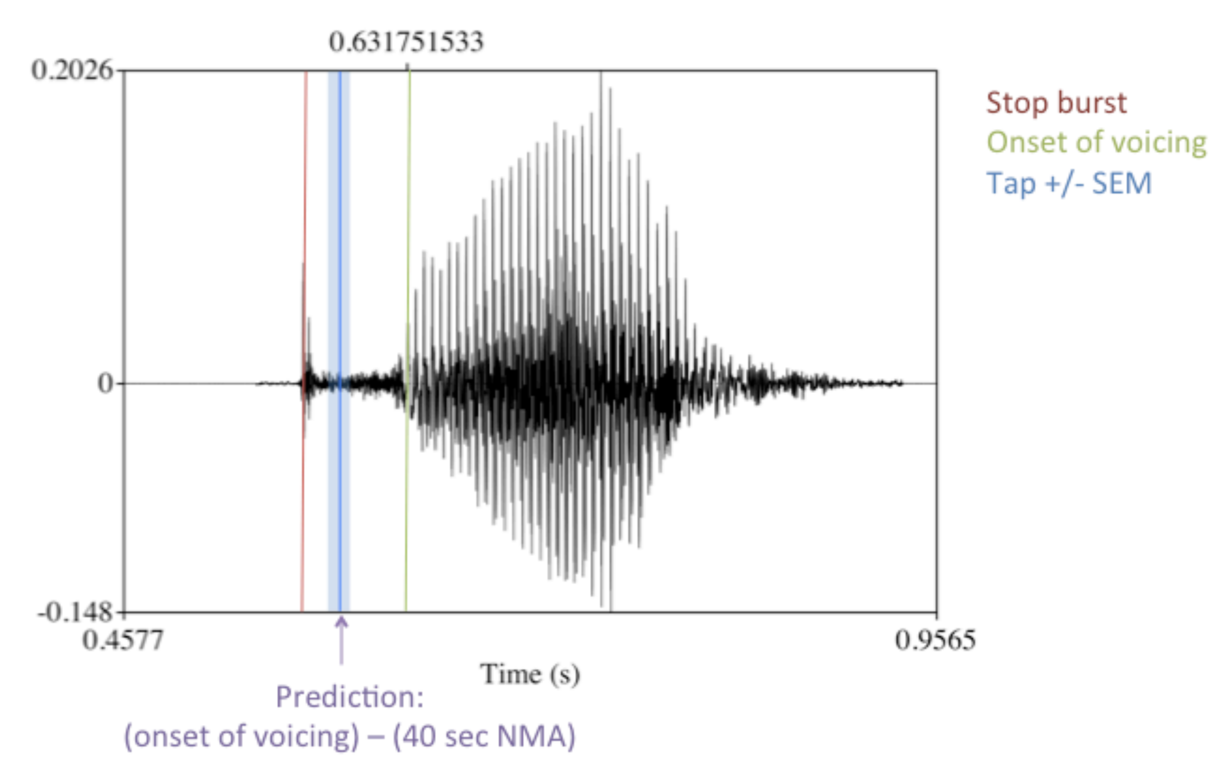
- First trial of each type and first 4 tokens of each trial omitted from analysis
- Automated algorithms in Praat[6] used to detect taps, stop consonant bursts, and onsets of voicing in the audio files; all were reviewed and hand-corrected by trained research assistants
- Time stamps were extracted and intervals between the various stimulus and response events were compared statistically

RESULTS

Condition 1, Tapping to Speech: Beat perceived at voicing onset

Consistent with our prediction, the taps occurred, on average, 41 ms prior to the onset of voicing.

- In a third condition (not reported here) participants both tapped and said "puh" to beeps. The NMA for the taps relative to the beeps was 40 ms.
- Therefore, participants appear to be perceiving the beat at the onset of voicing, and anticipating this by ~40 ms, consistent with an NMA.

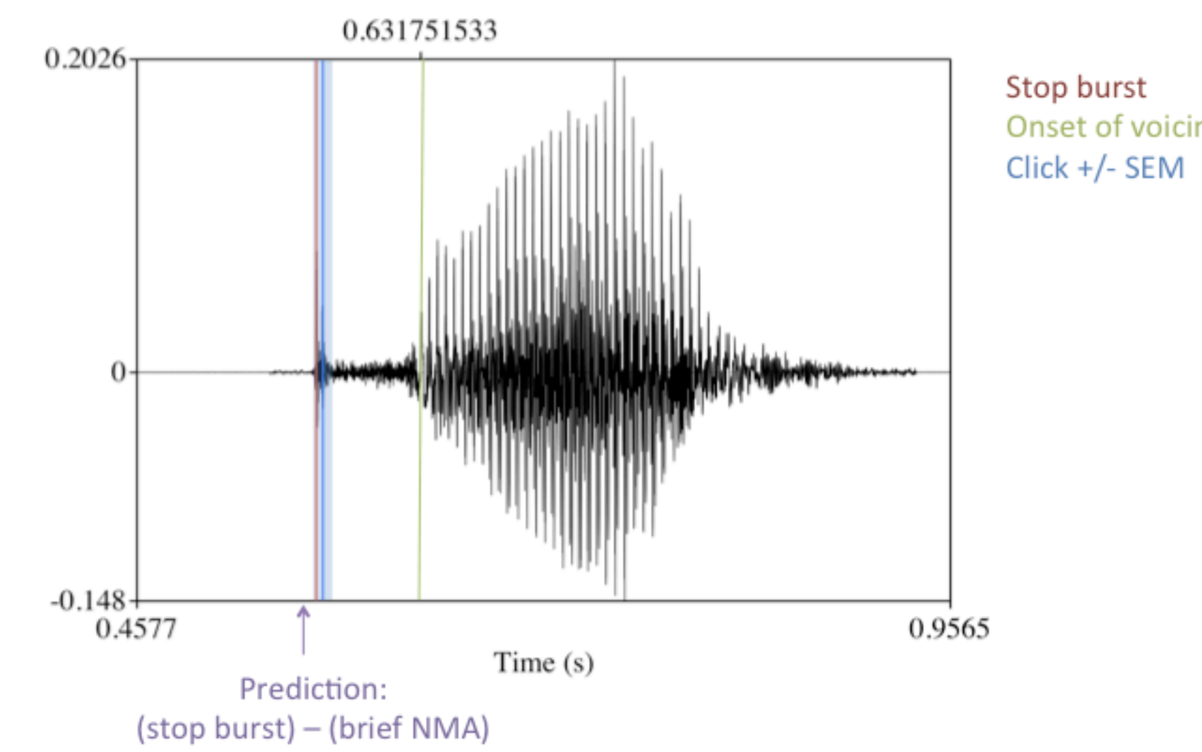


On average, taps fell 19 ms after the burst and 41 ms before voicing, SEM = 7 ms

Condition 2, Speaking to Beeps: Beat perceived at release burst of consonant

Consistent with our prediction, the burst was produced, on average, 5 ms before the stimulus beep.

- This is consistent with participants perceiving the beat at the consonant burst and perhaps producing a very slight NMA (although 5 ms is within measurement error, which we estimate to be approximately +/-10 ms).



On average, bursts occurred 5 ms before the stimulus beep and voicing onsets occurred 59 ms after the stimulus beep, SEM = 5.5 ms

DISCUSSION

Most of the literature examining where the beat occurs in speech tasks (i.e., the P-center) has used listening tasks, in which listeners adjust the intervals between speech stimuli or adjust click tracks to align with speech stimuli in some way.

Findings from the present study indicate that where the beat occurs in a syllable is task-dependent.

- For a person *hearing* the syllable, the vowel onset (i.e. the most acoustically salient point) may be where the beat is perceived.
- For a person *producing* the syllable, the burst of the onset stop consonant, (i.e. the most somatosensorially salient point) may be where the beat is perceived.

While this study was limited in scope (only looked at one syllable), the findings are suggestive that beat location in speech tasks is task-dependent. This could have important implications for using speech entrainment in clinical contexts.

References

1. Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic Bulletin & Review, 12*(6), 969-992.
2. Aschersleben, G., & Prinz, W. (1995). Synchronizing actions with events: the role of sensory information. *Perception & Psychophysics, 57,* 305-317.
3. Fujii, S., & Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: the SEP hypothesis. *Frontiers in Human Neuroscience.*
4. Ziegler, W., Aichert, I., & Staiger, A. (2010). Syllable- and rhythm-based approaches in the treatment of apraxia of speech. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders, 20,* 59-66.
5. Morton, J., Marcus, S., & Frankish C. (1976). Perceptual centers (P-centers). *Psychological Review, 83*(5), 405-408.
6. Boersma, P. & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.22