

<https://doi.org/10.1038/s42003-025-07544-8>

The timing of speech-to-speech synchronization is governed by the P-center

Check for updates

Tamara Rathcke

ARISING FROM C. Mares et al. *Communications Biology* <https://doi.org/10.1038/s42003-023-04976-y> (2023)

The ability to synchronize a motor response to an auditory signal is central to human activities such as dancing, joint music making, or conversing with others. Examining the temporal alignment between hands or articulators as motor effectors and speech syllables or music tones as auditory prompts, Mares et al.¹ concluded that the sensorimotor synchronization (SMS) ability varies greatly in the general population, with a group of people (called “low synchronizers”) being particularly disrupted when asked to synchronize to auditory sequences containing variable units. However, the well-foundedness of the conclusion is limited by a methodological oversight: the stimuli of the study do not consider the P-center effect that is central to the perception of temporal structure in speech and other acoustically complex sounds, thus making it difficult to draw meaningful comparisons between SMS to sequences containing identical vs. variable prompts and undermining the conclusions of the study.

Mares et al.¹ replicate the results of previous studies^{2,3} showing that the task of synchronizing articulatory gestures of the syllable “tah” with sequences of isochronous but varied syllables leads to a split of the general population into two groups: the “high synchronizers” who are able to repeat the syllable “tah” at a constant rate resembling the rate of the auditory prompt, and the “low synchronizers” who cannot maintain a steady rate of the “tah” syllable production when listening to sequences of varied syllables. Mares et al.¹ confirm that the difficulties of low synchronizers persist during the synchronization with sequences of variable tones and when the motor effector changes from articulators to hands. They further add to existing evidence that these difficulties disappear when the low synchronizers are asked to synchronize with auditory prompts containing sequences of identical units (either syllables or tones). Moreover, the authors demonstrate that sensorimotor priming with a sequence of identical tones can temporarily restore the low synchronizers’ ability to maintain a steady train of motor gestures during subsequent exposure to sequences of variable tones or syllables.

Unfortunately, the study suffers from a methodological oversight that limits direct comparison of SMS with identical vs. varied sequences. The phenomenon being overlooked in the study is the well-documented P-center effect. The P-center (the “perceptual center”) of a sound refers to the subjective moment of occurrence and signifies that the acoustic and the perceptual onset of a sound do not co-occur⁴. The P-center tends to be located after the acoustic onset of the corresponding sound, though its exact

location has been a matter of debate⁵ and differs across languages and possibly individuals^{6–9}. Studies broadly agree that the P-center approximates (and possibly anticipates)⁸ vowel onsets^{7,10,11}. The P-center has been attested in multifarious speech materials and by means of different tasks^{6,8}, with emerging evidence for its role in neural speech tracking¹². Moreover, it is not unique to speech, it has also been documented in musical sounds such as tones^{5,9,13,14} and will therefore apply to the tonal stimuli of the Mares et al. study¹ in a similar way.

Overall, previous research has convincingly documented that evenly concatenated syllables—i.e., sequences like the ones used in the experiment by Mares et al.¹—sound irregular to listeners^{4,10,15–17}. Similarly, when asked to synchronize the production of varied syllables with a metronome, speakers do not align syllable onsets in time with the metronome beat^{10,18–22}. To illustrate this, Fig. 1 compares the timing of concatenated syllables used in the experiment by Mares et al.¹ to the timing of the same syllables produced by a male speaker in time with a metronome set at a comparable rate (here, 250 ms). As can be seen in Fig. 1, stimuli of the speech-to-speech synchronization task (panel 1-A) display higher variability of inter-vocalic ($m_{(ISI)} = 227.67$ ms, $s_{(ISI)} = 19.29$ ms) than inter-syllabic ($m_{(ISI)} = 232.5$ ms, $s_{(ISI)} = 9.32$ ms) intervals. In contrast, the timing of syllables produced with the metronome (panel 1-B) shows the P-center effect, as indicated by a lower regularity of inter-syllabic intervals ($m_{(ISI)} = 248.89$ ms, $s_{(ISI)} = 18.86$ ms) and a higher regularity of inter-vocalic intervals ($m_{(ISI)} = 249.78$, $s_{(ISI)} = 6.10$ ms), with the latter approaching the metronome rate^{20–22}.

The methodological oversight is problematic because SMS requires auditory prompts to have temporal regularity and predictability^{23,24}. Given that the perceived temporal regularity in varied spoken and tonal units is a matter of the P-center timing, all sequences of the Mares et al. study¹ containing varied prompts may have sounded irregular to all participants. These irregularities (see Fig. 1A) resemble stimuli of previous experiments that used finger taps as motor effectors and examined responses to temporal perturbations of local inter-onset intervals in isochronous metronome sequences^{24,25}. Such phase perturbations have been shown to elicit error correction responses reflective of perceptual monitoring for temporal reference-frames within an incoming auditory stimulus^{24–26}. For example, when the onset timing of a local event is slightly shifted to deviate from isochrony of the remaining events in the sequence, participants shift their synchronization, even if explicitly instructed to ignore occasional

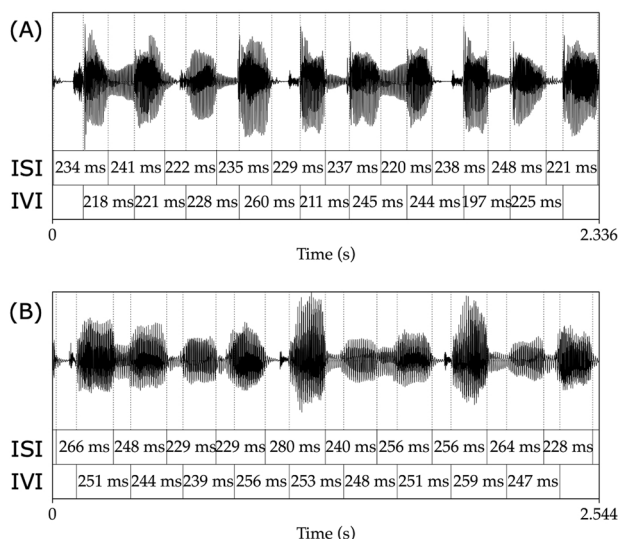


Fig. 1 | Timing of inter-syllabic and inter-vocalic intervals in a synthesised vs. naturally spoken train of syllables. Temporal analyses of inter-syllabic (ISI) and inter-vocalic (IVI) intervals in the materials of Mares et al.¹ (A, taken from the onset of a stimulus steadily paced at 4.3 units/second)⁴⁶ as compared to the production of a male speaker articulating the same materials (here, syllables) in time with a metronome paced at the rate of 250 ms (B). Annotations of syllable and vowel onsets were conducted manually by the author.

perturbations²⁷. The process of phase correction is therefore considered automatic and different from deliberate period correction elicited in response to global tempo changes within a sequence²⁷.

Given these properties of SMS, the task of the Mares et al. study¹ could only be performed if synchronization with temporally jittered prompts was not attempted at all. This means that “high synchronizers” performed well by ignoring the precise acoustic timing of local synchronization attractors and kept producing “tah” or clapping their hands at a rate broadly commensurate with the auditory prompt (listeners excel at establishing the distal rate of spoken input^{28,29}). “Low synchronizers”, on the other hand, may have conscientiously followed the prompt trying to synchronize with the jittered P-centers of concatenated syllables, repeatedly deploying phase correction and failing to establish synchronization. The measure of synchronization used in the study – the phase-locking value – captures exactly this SMS property by calculating distal phase covariance between amplitude envelopes of the perceived and produced sequences, disregarding the actual synchronization accuracy²³.

In a sense, then, “low synchronizers” were actually better at synchronizing with the external auditory prompt than “high synchronizers”. Since the grouping of participants into “high” and “low synchronizers” could no longer be maintained when the task involved acoustic prompts containing repetitions of the same unit (syllable or tone), it is very likely that the bipartite grouping of participants¹ does not arise from individual differences in synchronization in its classic definition^{23,24}. Indeed, it has been well established that individuals can vary in their general synchronization ability with different kinds of prompts^{30,31} and in their ability to adapt synchronization to tempo-changing³² or temporally perturbed³³ prompts – but so far, without a strong indication that synchronization may be non-unimodally distributed in the non-clinical population. One piece of evidence currently missing is an experiment with varied syllables concatenated such as to establish equal spacing between successive P-centers (rather than concatenating jittered syllables, see Fig. 1). This will help to illuminate the role of the P-center timing in the task¹.

Even though the speech-to-speech production studied by Mares et al.¹ is unlikely to test auditory-motor synchronization proper, the consistency with which it divides the general population into two groups^{2,34} is

remarkable and worth further consideration. In this context, the grouping can be hypothesized to arise from—hitherto poorly understood—individual differences in the interplay of feedback and feedforward control mechanisms during speech production^{35,36}. According to the neurocomputational DIVA model, for example, speech production can best be understood to emerge from the relations between brain activity, speech motor commands and their sensory output, and to be governed by two control mechanisms^{35,36}. Feedback control operates by identifying discrepancies between anticipated and actual outcomes of articulatory actions and adjusting motor commands in response. If feedback control detects auditory or somatosensory errors, corrections start to apply to feedforward processes. Feedforward control constitutes an internal motor program of speech sounds and syllables. During the production of a syllable like “tah”, the two mechanisms are assumed to interact, starting with the activation of the sensorimotor representations of the consonant and vowel gestures whose execution is monitored by feedback control. The model has found extensive support in auditory perturbation experiments^{37–40} – a paradigm that resembles in some ways the task of the study by Mares et al.¹ Within this framework, “low synchronizers” may be primarily recruiting feedback control for adjusting the timing of the articulatory gestures to align with the P-centers of the input syllables while “high synchronizers” may be exclusively relying on feedforward commands to perform the task⁴¹. The task likely involves somatosensory (rather than auditory) feedback mechanism, since the grouping of “high” vs. “low” synchronizers persists across effectors¹ while loudness adjustments do not affect the performance on this task. Open questions remain about how perceptual and motor abilities of the speaker but also the auditory stimulus itself influence the moment-to-moment balance of feedforward representations and feedback information (for relevant discussion, see refs. 38,42,43).

Other explanations (e.g., the presence of subjective rhythmization at a fast input tempo⁴⁴ or socio-psychological factors⁴⁵) are, however, also conceivable and would warrant careful examination.

Data availability

The stimuli displayed in Fig. 1A and used in the study by Mares et al.¹ are available from an open depository. The stimuli displayed in Fig. 1B are available from the author upon request.

Received: 16 May 2024; Accepted: 10 January 2025;

Published online: 22 January 2025

References

- Mares, C., Echavarría Solana, R. & Assaneo, M. F. Auditory-motor synchronization varies among individuals and is critically shaped by acoustic features. *Commun. Biol.* **6**, 658 (2023).
- Assaneo, M. F. et al. Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nat. Neurosci.* **22**, 627–632 (2019).
- Assaneo, M. F., Rimmele, J. M., Sanz Perl, Y. & Poeppel, D. Speaking rhythmically can shape hearing. *Nat. Hum. Behav.* **5**, 71–82 (2020).
- Morton, J., Marcus, S. & Frankish, C. Perceptual centers (P-centers). *Psychol. Rev.* **83**, 405–408 (1976).
- Villing, R. C., Repp, B. H., Ward, T. E. & Timoney, J. M. Measuring perceptual centers using the phase correction response. *Atten. Percept. Psychophys.* **73**, 1614–1629 (2011).
- Rathcke, T., Smit, E. A., Lin, C.-Y. & Kubozono, H. Testing an acoustic model of the P-center in English and Japanese. *J. Acoust. Soc. Am.* **155**, 2698–2706 (2024).
- Hoequist, C. E. The perceptual center and rhythm categories. *Lang. Speech* **26**, 367–376 (1983).
- Rathcke, T. The P-center effect and the domain of beat perception in speech. in L. Meyer and A. Strauß (eds.) *Rhythms of Speech and Language: Culture, Cognition, and the Brain* (Cambridge University Press, forthcoming) (2025).

9. Danielsen, A. et al. Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds. *J. Exp. Psychol. Hum. Percept. Perform.* **45**, 402–418 (2019).
10. Marcus, S. M. Acoustic determinants of perceptual center (P-center) location. *Percept. Psychophys.* **30**, 247–256 (1981).
11. Franich, K. Tonal and morphophonological effects on the location of perceptual centers (p-centers): evidence from a Bantu language. *J. Phon.* **67**, 21–33 (2018).
12. Oganian, Y. & Chang, E. F. A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Sci. Adv.* **5**, eaay6279 (2019).
13. London, J. et al. A comparison of methods for investigating the perceptual center of musical sounds. *Atten. Percept. Psychophys.* **81**, 2088–2101 (2019).
14. Vos, J. & Rasch, R. The perceptual onset of musical tones. *Percept. Psychophys.* **29**, 323–335 (1981).
15. Cooper, A. M., Whalen, D. H. & Fowler, C. A. The syllable's rhyme affects its P-center as a unit. *J. Phon.* **16**, 231–241 (1988).
16. Pompino-Marschall, B. On the psychoacoustic nature of the P-center phenomenon. *J. Phon.* **17**, 175–192 (1989).
17. Scott, S. K. The point of P-centres. *Psychol. Res.* **61**, 4–11 (1998).
18. Chow, I., Belyk, M., Tran, V. & Brown, S. Syllable synchronization and the P-center in Cantonese. *J. Phon.* **49**, 55–66 (2015).
19. Fowler, C. A. "Perceptual centers" in speech production and perception. *Percept. Psychophys.* **25**, 375–388 (1979).
20. Fox, R. A. & Lehiste, I. The effect of vowel quality variations on stress-beat location. *J. Phon.* **15**, 1–13 (1987).
21. Šturm, P. & Volín, J. P-centres in natural disyllabic Czech words in a large-scale speech-metronome synchronization experiment. *J. Phon.* **55**, 38–52 (2016).
22. Tuller, B. & Fowler, C. A. Some articulatory correlates of perceptual isochrony. *Percept. Psychophys.* **27**, 277–283 (1980).
23. Repp, B. H. & Su, Y.-H. Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bull. Rev.* **20**, 403–452 (2013).
24. Repp, B. H. Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* **12**, 969–992 (2005).
25. Repp, B. H. Multiple temporal references in sensorimotor synchronization with metrical auditory sequences. *Psychol. Res.* **72**, 79–98 (2007).
26. Vorberg, D. & Schulze, H.-H. Linear phase-correction in synchronization: predictions, parameter estimation, and simulations. *J. Math. Psychol.* **46**, 56–87 (2002).
27. Repp, B. H. Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization. *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 410–430 (2002).
28. Baese-Berk, M. M., Dilley, L. C., Henry, M. J., Vinke, L. & Banzina, E. Not just a function of function words: distal speech rate influences perception of prosodically weak syllables. *Atten. Percept. Psychophys.* **81**, 571–589 (2019).
29. Dilley, L. C. & McAuley, J. D. Distal prosodic context affects word segmentation and lexical processing. *J. Mem. Lang.* **59**, 294–311 (2008).
30. Dalla Bella, S. et al. BAASTA: battery for the assessment of auditory sensorimotor and timing abilities. *Behav. Res. Methods* **49**, 1128–1145 (2017).
31. Rathcke, T., Lin, C.-Y., Falk, S. & Bella, S. D. Tapping into linguistic rhythm. *Lab. Phonol.* **12**, 1–32 (2021).
32. Pecenka, N. & Keller, P. E. The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Exp. Brain Res.* **211**, 505–515 (2011).
33. Colley, I. D., Keller, P. E. & Halpern, A. R. Working memory and auditory imagery predict sensorimotor synchronisation with expressively timed music. *Q. J. Exp. Psychol.* **71**, 1781–1796 (2018).
34. Sjul, G. S., Vulchanova, M. D. & Assaneo, M. F. Replication of population-level differences in auditory-motor synchronization ability in a Norwegian-speaking population. *Commun. Psychol.* **1**, 47 (2023).
35. Guenther, F. H. A neural network model of speech acquisition and motor equivalent speech production. *Biol. Cybern.* **72**, 43–53 (1994).
36. Perkell, J. S. Movement goals and feedback and feedforward control mechanisms in speech production. *J. Neurolinguistics* **25**, 382–407 (2012).
37. Oschkinat, M. & Hoole, P. Compensation to real-time temporal auditory feedback perturbation depends on syllable position. *J. Acoust. Soc. Am.* **148**, 1478–1495 (2020).
38. Oschkinat, M., Hoole, P., Falk, S. & Dalla Bella, S. Temporal malleability to auditory feedback perturbation is modulated by rhythmic abilities and auditory acuity. *Front. Hum. Neurosci.* **16**, 885074 (2022).
39. Karlin, R. & Parrell, B. Speakers monitor auditory feedback for temporal alignment and linguistically relevant duration. *J. Acoust. Soc. Am.* **152**, 3142–3154 (2022).
40. Purcell, D. W. & Munhall, K. G. Compensation following real-time manipulation of formants in isolated vowels. *J. Acoust. Soc. Am.* **119**, 2288–2297 (2006).
41. Jones, J. A. & Keough, D. Auditory-motor mapping for pitch control in singers and nonsingers. *Exp. Brain Res.* **190**, 279–287 (2008).
42. Keough, D. & Jones, J. A. The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *J. Acoust. Soc. Am.* **126**, 837–846 (2009).
43. Subramaniam, K., Kothare, H., Mizuiri, D., Nagarajan, S. S. & Houde, J. F. Reality monitoring and feedback control of speech production are related through self-agency. *Front. Hum. Neurosci.* **12**, 82 (2018).
44. Parncutt, R. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Percept.* **11**, 409–464 (1994).
45. Cohn, M., Keaton, A., Beskow, J. & Zellou, G. Vocal accommodation to technology: the role of physical form. *Lang. Sci.* **99**, 101567 (2023).
46. Lizcano-Cortés, F. et al. Speech-to-Speech Synchronization protocol to classify human participants as high or low auditory-motor synchronizers. *STAR Protoc.* **3**, 101248 (2022).

Acknowledgements

The author would like to thank Phil Hoole, Eniko Ladanyi, and Antje Strauß for a discussion of the issues raised in the commentary.

Competing interests

The author declares no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Tamara Rathcke.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025