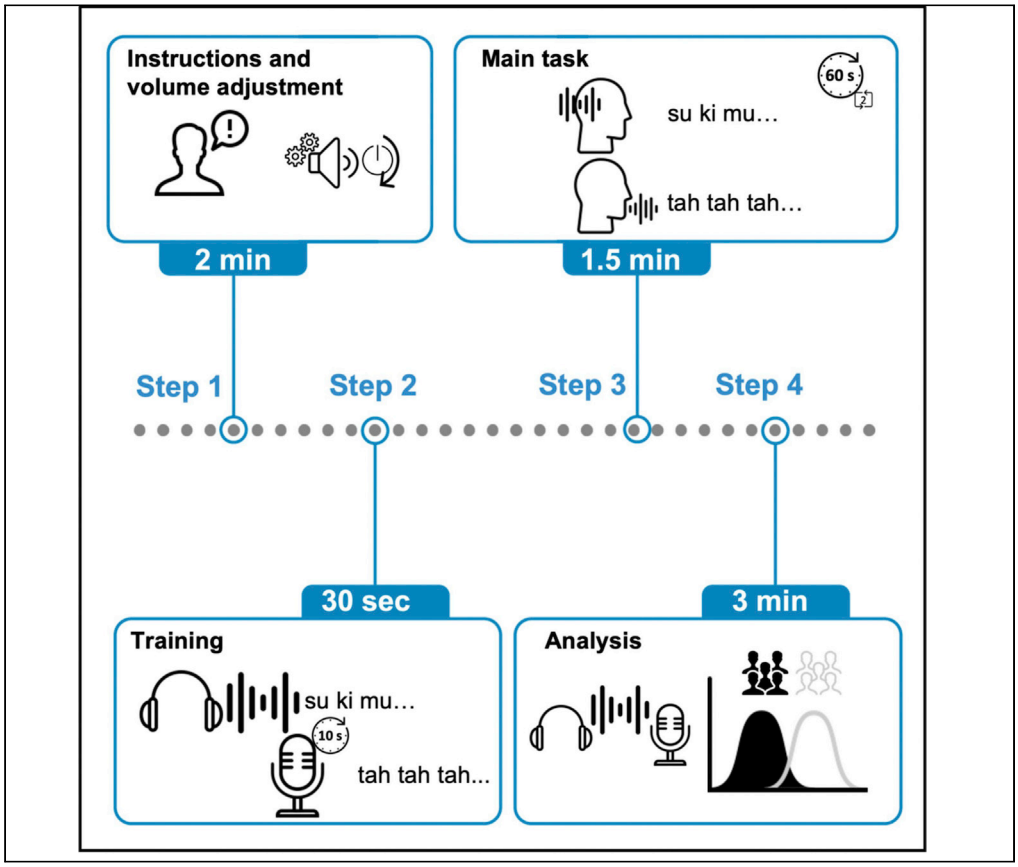
STAR Protocols

CellPress
OPEN ACCESS

Protocol

Speech-to-Speech Synchronization protocol to classify human participants as high or low auditory-motor synchronizers



The ability to synchronize a motor action to a rhythmic auditory stimulus is often considered an innate human skill. However, some individuals lack the ability to synchronize speech to a perceived syllabic rate. Here, we describe a simple and fast protocol to classify a single native English speaker as being or not being a speech synchronizer. This protocol consists of four parts: the pretest instructions and volume adjustment, the training procedure, the execution of the main task, and data analysis.

Fernando Lizcano-Cortés, Ireri Gómez-Varela, Cecilia Mares, ..., David Poeppel, Pablo Ripollés, M. Florencia Assaneo

ceciliap.maresr@gmail.com (C.M.)
fassaneo@inb.unam.mx (M.F.A.)

Highlights

Behavioral protocol to assess individuals' degree of speech auditory motor synchrony

Individuals can be labeled as high or low speech synchronizers

When assessed over time, classifications were stable

Differences between groups have been shown at brain and cognitive levels

Lizcano-Cortés et al., STAR Protocols 3, 101248
June 17, 2022
https://doi.org/10.1016/j.xpro.2022.101248

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Speech-to-Speech Synchronization protocol to classify human participants as high or low auditory-motor synchronizers

Fernando Lizcano-Cortés,[1,7] Ireri Gómez-Varela,[1,7] Cecilia Mares,[1,7,8,*] Pascal Wallisch,[2] Joan Orpella,[2] David Poeppel,[2,3,4,5] Pablo Ripollés,[2,4,5,6] and M. Florencia Assaneo[1,8,9,*]

[1]Institute of Neurobiology, UNAM, Querétaro 76230, México

[2]Department of Psychology, New York University, New York, NY 10003, USA

[3]Ernst Struengmann Institute for Neuroscience, 60528 Frankfurt, Germany

[4]Center for Language, Music and Emotion (CLaME), New York University, New York, NY, USA

[5]Max Plank Institute for Empirical Aesthetics, 60322 Frankfurt, Germany

[6]Music and Audio Research Laboratory (MARL), New York University, New York, NY 11201, USA

[7]These authors contributed equally

[8]Technical contact

[9]Lead contact

*Correspondence: ceciliap.maresr@gmail.com (C.M.), fassaneo@inb.unam.mx (M.F.A.)
https://doi.org/10.1016/j.xpro.2022.101248

SUMMARY

The ability to synchronize a motor action to a rhythmic auditory stimulus is often considered an innate human skill. However, some individuals lack the ability to synchronize speech to a perceived syllabic rate. Here, we describe a simple and fast protocol to classify a single native English speaker as being or not being a speech synchronizer. This protocol consists of four parts: the pretest instructions and volume adjustment, the training procedure, the execution of the main task, and data analysis.

For complete details on the use and execution of this protocol, please refer to Assaneo et al. (2019a).

BEFORE YOU BEGIN

A recent study introduced the Speech-to-Speech Synchronization test, a behavioral protocol showing that the general population can be separated into two groups according to individual differences in the degree of speech auditory-motor synchronization (Assaneo et al., 2019a). Such a binary classification has been evidenced in the study by the bimodal nature of the distribution of the obtained synchronization measurements. Specifically, this study showed that when individuals listen to a rhythmic train of syllables and - concurrently and continuously - whisper the syllable "tah", some speakers spontaneously align their produced syllabic rate to the perceived one (high synchrony group), while others do not (low synchrony group). Importantly, synchronization measurement remains stable while assessed in two sessions separated by a month ($n = 34$, Spearman correlation between sessions, $r = 0.78$, $P < 0.001$; see Assaneo et al., 2019a). This result implies that group belonging represents a stable individual feature. This work, along with a set of follow up studies, showed that group membership (i.e., being a high or a low synchronizer) is predictive of performance in a set of cognitive tasks (e.g., word learning) as well as of brain structural and functional features (Assaneo et al., 2019a, 2019b; Kern et al., 2021). Furthermore, the inclusion of the different synchrony groups in the analysis of experimental outcomes has resulted in the emergence of relevant results that would have been masked by pooling together all individuals in the sample (Assaneo

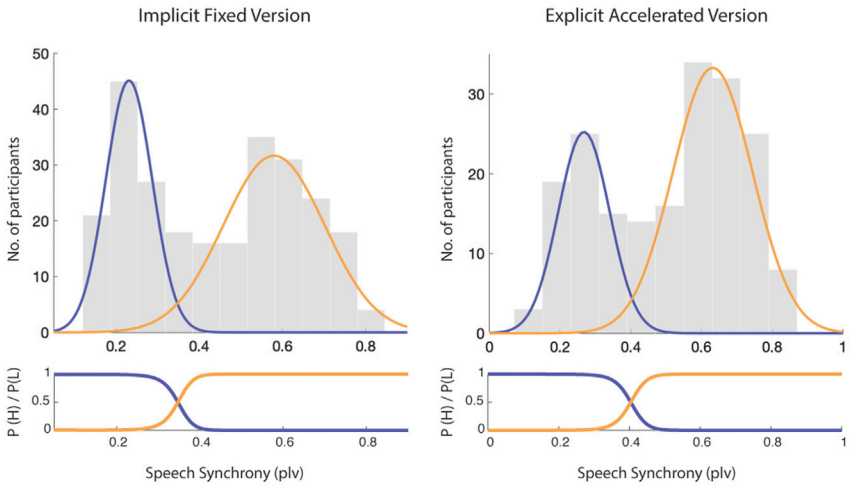


Figure 1. Bimodal distributions produced by the two versions of the Speech-to-Speech Synchronization test

Upper panels: Histograms for the synchronization measurements (Phase Locking Values) obtained by using both versions of the test. Implicit Fixed version on the left (N=255). Explicit Accelerated version on the right (N=190). The colored traces represent the two normal distributions obtained by adjusting a two component gaussian mixture model on the data (Implicit Fixed: Component 1, High Synchronizers; mixing proportion: 0.60, mean: 0.58. Component 2, Low Synchronizers; mixing proportion: 0.40, mean: 0.23. Explicit Accelerated: Component 1, High Synchronizers; mixing proportion: 0.67, mean: 0.63. Component 2, Low Synchronizers; mixing proportion: 0.33, mean: 0.27). Lower panels: Probability of belonging to one of the two groups as a function of the participant's degree of synchrony. Probability curves are derived from the distributions obtained from the gaussian mixture models adjusted to the datasets. In all panels, orange and blue represent the high and low synchronizers, respectively.

et al., 2020, 2021). In that vein, the Speech-to-Speech Synchronization test appears as a robust and easy to administer behavioral tool to enable the inclusion of relevant individual differences into experimental protocols.

Currently there are two different versions of the test, which have been employed in different studies. First, the original *Implicit Fixed Version*, in which the external (i.e., auditorily presented) syllabic rate remains stable at 4.5 Hz, and participants are not explicitly instructed to synchronize their vocalizations to the auditory stimulus. Second, the *Explicit Accelerated Version*, in which the external syllabic rate starts at 4.3 Hz and increases in steps of 0.1 Hz every 10 s until it reaches 4.7 Hz, and the participants are explicitly instructed to synchronize their speech output to the perceived speech rate. In this accelerated case, the spontaneous nature of the synchrony relies on the fact that, although participants cannot detect the 0.1 Hz increments in the external syllabic rate, high synchronizers still automatically adjust their spoken pace to the subtly accelerating speech input.

Both versions result in a bimodal distribution of the synchrony measurements (Figure 1) and have been repeatedly used in previous research. Although distinctive brain features between synchrony groups (structural MRI, MEG) were only assessed using the original version (Assaneo et al., 2019b, 2020), differences in behavior have been reported with both test versions, with the *Implicit Fixed* (Assaneo et al., 2019a, 2020, 2021) as well as with the *Explicit Accelerated* versions (Kern et al., 2021).The protocol outlined here describes the specific steps required to classify a given native English speaker as a low or a high synchronizer by means of the two existing versions of the Speech-to-Speech Synchronization test. The selection of one of the two alternative test versions is left to the researcher and to considerations about the potential effects of steady rhythmic syllabic trains on the rest of the tasks in a researcher's protocol.

All protocols were reviewed and approved by the local institutional review board, New York University's Committee on Activities Involving Human Subjects.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Auditory stimuli, wav files	This paper	https://zenodo.org/badge/latestdoi/407612860
MATLAB code for running both versions of the test and analyzing its outcome	This paper	https://zenodo.org/badge/latestdoi/407612860
Python code analyzing the outcome	This paper	https://doi.org/10.5281/zenodo.6148008
Gorilla open materials to run both versions of the test remotely	This paper	https://app.gorilla.sc/openmaterials/290032
Experimental models: Organisms/strains		
255 Human participants (112 males; mean age, 30 years; age range, 19–55 years; native English speakers)	(Assaneo et al., 2019b)	N/A
190 Human participants (81 males; mean age, 25 years; age range, 19–45 years; native English speakers)	This paper	N/A
Software and algorithms		
MATLAB; Version: 9.10.0.1739362 (R2021a)	MathWorks	https://www.mathworks.com
Psychtoolbox v3.0.17	(Assaneo et al. (2019a))	http://psychtoolbox.org
Gorilla Experiment Builder	(Anwyl-Irvine et al., 2020)	https://gorilla.sc
Praat	(Boersma and Weenink, 2001)	https://praat.en.softonic.com/

STEP-BY-STEP METHOD DETAILS

We focus on two versions of the Speech-to-Speech Synchronization test: the *Implicit Fixed* and the *Explicit Accelerated*. Below we detail the steps necessary to carry out both versions. When neither version is mentioned, the particular step does not differ between them. Both versions can be conducted in-lab or remotely/online. The in-lab experimental scripts, as well as the wav files described below are available at https://zenodo.org/badge/latestdoi/407612860 (MATLAB version) whereas the online scripts to conduct the experiment can be found at https://app.gorilla.sc/openmaterials/290032.

For the general setup of the experiment, participants should sit in front of a computer close to a microphone and wear headphones. Wearing headphones represents a crucial aspect of the design, for two main reasons: (i) recording should not be contaminated by the auditory stimulus and (ii) participants own voice feedback should be masked by the external stimulus (i.e., participants should not be able to listen to their own vocal production). The microphone can be externally connected to the computer; the computer's internal microphone can also be used. Participants are instructed to maintain a short distance (below 30 cm) between their mouth and the microphone throughout the test. When the test is conducted in-lab, instructions are given verbally at the start of the experiment and appear written on the computer screen before the beginning of each step.

Part one

Volume adjustment

Timing: approximately 2 min (15 s of volume adjusting + 1.5 min for instructing the participant)

Note: In case of applying the test remotely/online, two additional steps are added to ensure that the participant is wearing headphones and that the microphone is working properly (Woods et al., 2017).

1. Have participants listen to a train of synthesized syllables played backwards (i.e., the same audio wav used as stimulus in the main task, made of 16 syllables randomly concatenated, but reversed in time) while asking them to whisper the syllable "tah".

2. Ask participants to gradually increase the volume until they cannot hear their own whisper while still being at a comfortable level.
3. Once they select the volume level, instruct them not to change the volume throughout the task.

⚠ CRITICAL: The maximal volume reached by the used device and stimuli in our case was 100 dB, stated by the WHO as a safe sound level if listened for 15 min each day (World Health Organization, 2015).

⚠ CRITICAL: Crucially, all included participants should report not hearing their own voice.

Note: The volume selected by the participants does not distinguish between high and low synchronizers (Mann–Whitney–Wilcoxon test, two-sided P = 0.69; highs: n = 15, meanVol = 95.16 dB, s.d. = 3.84 dB, and lows: n = 16 meanVol = 95.67 dB, s.d. = 4.07 dB).

Part two

Training

Timing: approximately 30 s

4. Have participants passively listen to a 10-s rhythmic train of syllables.
5. Syllables are presented at 4.5 Hz for the Implicit Fixed Version and 4.3 Hz for the Accelerated Explicit Version (example_45 Hz.wav and example_43 Hz.wav, respectively).
6. Once the rhythmic train ends, ask participants to whisper the syllable "tah" at the same pace for 10 s.

⚠ CRITICAL: In the implicit version tell participants that this step is for them to get an idea of how they are supposed to be whispering continuously during the main task. More precisely, give the following instructions: "First, you will be presented with an example audio of how to continuously and repeatedly whisper the syllable 'tah.' Pay attention to it and once it ends it will be your turn to practice the whispering."

Part three

Main task

Timing: approximately 1.5 min (60 s for the main task run + 30 s for the two-alternative forced choice questions only in the Implicit Fixed Version)

7. For the *Implicit Fixed* version, ask participants to pay attention to the perceived syllables while continuously whispering the syllable 'tah'. Explain to them that after the presentation, they will be required to identify a subset of the presented syllables and that the point of the continuous 'tah' whispering is to make more challenging the syllables recall. Do not disclose *that the objective of the whispering is to measure their speech-to-speech synchrony*. For the Explicit Accelerated Version, ask participants to synchronize the repeated whispered syllable to the rate of the auditory stimulus.
8. Have participants listen to a 60-s audio comprising a rhythmic train of syllables (*stimulus_fix.wav* for the Implicit Fixed version, *stimuls_acc.wav* for the Explicit Accelerated).
9. Have participants continuously whisper the syllable "tah" while looking at a fixation cross in the center of the screen during the whole audio presentation and record participants' vocalizations.
10. Only for the Implicit Fixed Version, once the audio presentation ends. Ask participants to answer four two-alternative forced choice questions about whether a particular syllable was presented or not (e.g., "Did you hear the syllable /bah/?"). Have them respond with the keyboard by pressing Y for yes or N for no.

Note: The purpose of this assessment is to direct the participant's attention to the syllable detection task and to avoid having them intentionally synchronizing their whisper to the auditory stimulus. There is no useful information in the participants' responses, it has been shown that lows and high synchronizers have equal poor performance on this task (Assaneo et al., 2019a).

Note: The test consists of two runs, therefore Part two and three are repeated.

Part four
Analysis

Timing: approximately 3 min

11. Visualize and listen to the recorded audio signals. We suggest using the software *praat* (Boersma and Weenink, 2001). If none of the exclusion criteria are reached (see below for details regarding exclusions), each participant is labeled as a high or low synchronizer, following the analysis described below.

EXPECTED OUTCOMES

For a single participant, the outcome comprises two audio files of 60 s each. Importantly, the recorded acoustic signal should not contain background noise loud enough to mask the reconstruction of the produced speech signal (see Figure 2). Also, researchers should listen to and visualize the audio signal to control for all other exclusion criteria detailed in the following section. If a large sample study (N>60) is conducted, a bimodal distribution is expected for the synchronization measures (see Figure 1 for the expected outcome: see troubleshooting, problem 1 for a possible solution if the bimodal is not clear).

QUANTIFICATION AND STATISTICAL ANALYSIS

Before estimating the individual's degree of speech-to-speech synchrony the audio signal should be evaluated according to the following exclusion criteria. The data should be excluded if: (i) the participant speaks aloud (i.e., activating the vocal cords) instead of whispering, (ii) the recording shows high environmental noise, masking the reconstruction of the speech envelope (see Figure 2), (iii) silence gaps between "tah" utterances are longer than 3 s , (iv) the spoken rate is equal to or lower than 2 Hz (i.e., the participant produces 2 "tahs" or less per second) and (v) the recorded audio is contaminated with the leaking stimulus (troubleshooting, problem 2, for a possible solution).

If none of the listed issues are present in the recordings, the phase locking value (PLV) between the envelope of the produced and perceived acoustic signal is computed to classify the participant as a high or a low synchronizer. The PLV for the first and second runs of the test are estimated following the steps described in Assaneo et al. (2019a). For the datasets presented here (Figure 1), we fitted a linear regression with the PLV obtained during the second run as the dependent variable and the PLV of the first run as the independent variable. The results ($y = mx + b$) showed a strong correlation between runs in both versions, with coefficients (with 95% confidence bounds): (i) Implicit Fixed; m= 0.90 (0.83, 0.97), b= 0.07 (0.04, 0.10); and (ii) Explicit Accelerated; m = 0.99 (0.94, 1.05), b = 0.01 (-0.02, 0.04). If a given pair of (PLV run1, PLV run2) is outside the 95% confidence bounds of the fitted lines, the participant will be labeled as inconsistent and excluded. Otherwise, the PLV, averaged across runs, will be assigned as the individual's degree of synchrony. This value will be used to estimate the participant's probability of being a high (or a low) synchronizer. For this purpose, we fit a gaussian mixture model with two components to the distribution of the synchrony measurements obtained with each version of the test (N=255 for the Implicit Fixed; N=190 for the Explicit Accelerated). The obtained distributions for each component, which represent the low and high synchronizers, have been used to compute a probability function for the individual degrees of synchrony

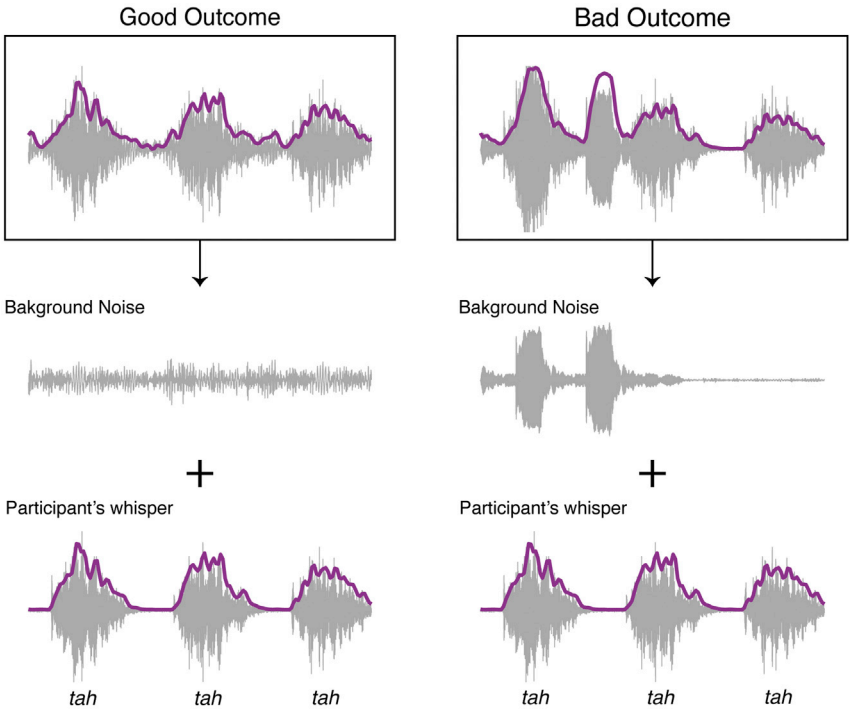


Figure 2. Examples of a bad and a good audio recording

The upper panel show two schematic outcomes, which are composed of the acoustic signals represented in the middle and bottom panels. In both cases, the participant produced the same train of "tahs" (two bottom rows) and it is the background noise that differs between the right and the left examples. In the example on the left, the audio was recorded with a stable and relatively low background noise, which did not alter the whisper's envelope. In the example on the right, the background noise shows abrupt increments in amplitude (which could represent different naturalistic sounds such as other voices, a dog's barking, or a telephone ringing). In this case, as shown in the upper panel, the envelope of the recording does not recover the one of the participant's whispers. In all panels, the acoustic signal is depicted in gray while the corresponding envelope is highlighted in purple.

to belong to one or other group (see Figure 1). Given the features of the sample used to compute the probability distribution (i.e., large n and unrestricted participation requirements other than being between 18 and 50 years old) it represents a good approximation of the general adult American population synchronization properties.

A Matlab script implementing the complete analysis pipeline can be downloaded from https://zenodo.org/badge/latestdoi/407612860 (or from https://doi.org/10.5281/zenodo.6148008 implemented in python). In this script, the researcher should specify the name of the recorded wav files and the version of the test employed. Specifically, the script will perform the following steps: (1) extract the envelope of the produced speech signal and filter it around the stimulus syllabic rate; (2) compute the PLV between the produced and perceived filtered envelopes in windows of 5 s length with an overlap of 2 s; (3) average the PLVs for each audio file (i.e., for run1 and run2); (4) control for consistency between runs; (5) give the probability for the participant being a high or a low synchronizer, according to the distributions presented in this manuscript (see Figure 1 and previous paragraph). For more detail about steps (1), (2) and (3) refer to (Assaneo et al., 2019a).

LIMITATIONS

Although both versions of the test have repeatedly produced bimodal distributions when assessing English (Assaneo et al., 2019a) and German (Assaneo et al., 2021) speakers, the protocol still needs to be validated for other languages. While differences in behavior have been reported applying the

Explicit Accelerated version, brain differences between groups, as defined by this version, remain unexplored. Further research is required to assess which of the two existing versions splits the population into groups with sharper distinctions.

In addition, the remote/online application of the test yields a high attrition rate: approximately 30%–40% of participants are excluded (in-lab: < 10%). Data are discarded for different reasons, in a hierarchical order from the most to the less frequent they are: (i) recordings collected in very noisy environments or while the participant is not wearing headphones, (ii) participants speak loudly instead of whispering, due to the absence of the researcher to correct them during the training, (iii) technical issues related to the participant's equipment, most notably the microphone,

TROUBLESHOOTING

The Speech-to-Speech Synchronization test is short and straightforward to complete, therefore, when any of the exclusion criteria are found, a possible solution is to repeat the assessment. It is worth noting that, since being a high or a low synchronizer is a stable individual feature, a repetition of the test will not modify the outcome. Stability can be inferred from the fact that: (i) synchronization measurements are stable across different sessions and (ii) participants do not show an improvement in the second run of the test while compared against the first one (see the linear regression fitted to the data in section quantification and statistical analysis).

Problem 1

For studies intended to compare brain or behavioral features between high and low synchronizers it could be advantageous to get a large sample with a similar number of participants from each group (Keppel and Wickens, 2004; Rusticus and Lovato, 2019). While there are statistical methods specifically designed to deal with unbalanced samples (i.e., (Parra-Frutos, 2013)), even for single case studies (Crawford and Garthwaite, 2002), it is convenient to get balanced samples when possible. Specifically, if researchers want to use some of the commonly chosen approaches (e.g., ANOVA, linear mixed models or decoding strategies) and want to maximize the statistical power. For this reason, in this section we expose why getting balanced samples could be problematic and a plausible strategy to overcome this issue, if required.

Since the proportion of high and low synchronizers in the general population is unbalanced (see the gaussian mixture models fitted to the data described in Figure 1); the number of highs will generally exceed the number of lows. Furthermore, having musical training increases the odds of being a high synchronizer (Assaneo et al., 2019a), and the standard participants are students who, in general, have some musical training. For these reasons, it can be sometimes complicated to recruit enough low synchronizers (step 11).

Potential solution

Debriefing the participants about their music experience before including them into the study using, for example, the Gold-MSI (Müllensiefen et al., 2014). Assessing only participants with no musical training increases the chances of getting a balanced number of high and low synchronizers.

Problem 2

It is typical to find participants who set the volume too loud, which produces headphones' sound leakage contaminating the recordings (step 2). In these cases, when the researcher listens to the recordings (step 11), they will hear not only the participant's whisper but also the stimulus. The combination of the perceived and produced acoustic signals into the recording leads to an incorrect estimation of the participant's synchrony.

Potential solution

If the experiment is conducted in-lab, we recommend using insert earplugs. Specifically, we found that the ER 1 etymotic earplugs (http://www.etymotic.com) show no sound leakage even for very

loud volume settings. When collecting data remotely, we suggest using the stimulus filtered with a bandpass between 0 to 3 kHz (filtered audio files are included in both versions shared in the Gorilla platform). Typically, the whisper's spectral content exceeds 3 kHz. Therefore, applying a stopband filter between 0 and 3 kHz to the recording allows for the removal of the stimulus contamination without altering the envelope of the train of "tahs". Summarizing, to solve this problem, use the filtered stimulus and apply a stopband filter [0 3000] Hz to the recorded audio signals before running the analysis.

Problem 3

It is important to notice that the test has only been validated for "whispering participants". It is not clear whether the bimodal will still be evidenced if participants speak aloud. For those reasons, if the recordings show voiced periods (i.e., periods where the vocal folds are active) longer than 3 s this participant should be removed (step 11). Thus, researchers should avoid participants speaking aloud and we noticed that for some individuals, the difference between speaking aloud and whispering is not clear.

Potential solution

When the test is performed in the lab, the experimenter should check that the participant understands how to whisper before starting the experiment (step 6). More specifically, the researcher should ask the participant to try and ensure that they are really whispering. If the participant is not doing it correctly, a plausible strategy is to instruct them to place one hand on the throat to feel the difference between activating or not the vocal cords. Another is to imagine telling a secret to a friend in a very quiet place while surrounded by many people.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, M. Florencia Assaneo (fassaneo@inb.unam.mx).

Materials availability

Ethical restrictions apply to the original data set comprising individuals' voices, and it cannot be shared on a public server. All relevant measurements extracted from the original data to construct Figure 1 are available from the lead contact author upon request.

Data and code availability

The scripts to run both versions of the Speech-to-Speech Synchronization remotely are available at https://app.gorilla.sc/openmaterials/290032. The scripts to run the tests in-lab and to analyze the data can be found at https://zenodo.org/badge/latestdoi/407612860 (MATLAB version) and at https://doi.org/10.5281/zenodo.6148008 (python analysis version).

ACKNOWLEDGMENTS

This work was supported by UNAM-DGAPA-PAPIIT IA202921 (M.F.A.), IBRO Return Home Fellowship (M.F.A.), and by the National Science Foundation under grant 2043717 (D.P., P.R., and M.F.A). F.L.-C., I.G.-V., and C.M. received CONACYT funding (CVU: 779254, 1045881 & 1086447, respectively) from the Mexican government.

AUTHOR CONTRIBUTIONS

M.F.A., P.R., and D.P. conceived and supervised the project. P.W., M.F.A., and J.O. conducted the experiments. M.F.A., F.L.-C., I.G.-V., and C.M. analyzed the data. M.F.A. and F.L.-C. wrote the manuscript. All authors read and approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Assaneo, M.F., Ripollés, P., Orpella, J., Lin, W.M., de Diego-Balaguer, R., and Poeppel, D. (2019a). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nat. Neurosci. *22*, 627–632.

Assaneo, M.F., Rimmele, J.M., Orpella, J., Ripollés, P., de Diego-Balaguer, R., and Poeppel, D. (2019b). The lateralization of speech-brain coupling is differentially modulated by intrinsic auditory and top-down mechanisms. Front. Integr. Neurosci. *13*, 28.

Anwyl-Irvine, A.L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J.K. (2020). Gorilla in our midst: An online behavioral experiment builder. Behav. Res. *52*, 388–407.

Assaneo, M.F., Orpella, J., Ripollés, P., Noejovich, L., López-Barroso, D., Diego-Balaguer, R. de, and Poeppel, D. (2020). Population-level differences in the neural substrates supporting statistical learning. Preprint at bioRxiv, 2020.07.03.187260.

Assaneo, M.F., Rimmele, J.M., Sanz Perl, Y., and Poeppel, D. (2021). Speaking rhythmically can shape hearing. Nat. Hum. Behav. *5*, 71–82.

Boersma, P., and Weenink, D. (2001). PRAAT, a system for doing phonetics by computer. Glot Int. *5*, 341–345.

Brainard, D.H. (1997). The Psychophysics Toolbox. Spat. Vis. *10*, 433–436.

Crawford, J.R., and Garthwaite, P.H. (2002). Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences. Neuropsychologia *40*, 1196–1208.

Keppel, G., and Wickens, T.D. (2004). Design and Analysis: A Researcher's Handbook (Pearson College Div).

Kern, P., Assaneo, M.F., Endres, D., Poeppel, D., and Rimmele, J.M. (2021). Preferred auditory temporal processing regimes and auditory-motor synchronization. Psychon. Bull. Rev. *28*, 1860–1873.

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. PLoS ONE *9*, e89642.

Parra-Frutos, I. (2013). Testing homogeneity of variances with unequal sample sizes. Comput. Stat. *28*, 1269–1297.

Rusticus, S., and Lovato, C. (2019). Impact of sample size and variability on the power and type I error rates of equivalence tests: a simulation study. Pract. Assess. Res. Eval. *19*, 11.

Woods, K.J.P., Siegel, M.H., Traer, J., and McDermott, J.H. (2017). Headphone screening to facilitate web-based auditory experiments. Atten. Percept. Psychophys. *79*, 2064–2072.

World Health Organization (2015). Make Listening Safe (No. WHO/NMH/NVI/15.2).