**A Programmatic Stage 1 Registered Report of global song-speech relationships replicating and extending Ozaki et al. (2024) and Savage et al. (2025)**

|  |
| --- |
| **Please note:** This is a preprint that has not yet completed peer review. We welcome questions, comments, citation, and constructive criticism, bearing in mind that this is a non-peer-reviewed draft subject to revision. Please direct correspondence to [patrick.savage@auckland.ac.nz](mailto:patrick.savage@auckland.ac.nz)    **Recommended citation:** Savage, P. E., Jia, Z., Ozaki, Y., Pavlovich, D., Purdy, S., Ampiah-Bonney, A., Arabadjiev, A., Arnese, F., Bamford, J., Barbosa, B. S., Beck, A.-K., Cabildo, A., Chiba, G., Dabaghi Varnosfaderani, S., Echim, S., Fujii, S., Gabriel, S., Grassi, M., Guiotto Nai Fovino, L., Hajič jr., J., Hartmann, M., Hansen, N. C., He, Y., Kolios, S., Krzyżanowski, W., Kuikuro, U., Kurdova, D., Liu, F., Loui, P., Mikova, Z., Moya, D., Natsitsabui, R., Niiranen, M., Nguqu, N., Nuska, P., Nweke, F. E., Opondo, P., Parkinson, H., Parselelo, M. L., Perry, G., Pfordresher, P. Q., Podlipniak, P., Popescu, T., Ross, R. M., Shi, Z., Silva-Zurita, J., Soto-Silva, I., Štěpánková, B., Thompson, W. F., Vaida, S., Vanden Bosch der Nederlanden, C. (2025). A Programmatic Stage 1 Registered Report of global song-speech relationships replicating and extending Ozaki et al. (2024) and Savage et al. (2025). *PsyArXiv* preprint: <https://doi.org/10.31234/osf.io/c2dba> |

**Authors**

*NB: All authors besides the first five are listed in alphabetical order. For planned authorship for the different Stage 2 manuscripts resulting from this Stage 1 protocol, see Table 2.*

Patrick E. Savage[[1]](#endnote-1),[[2]](#endnote-2)\*, Zixuan Jia[[3]](#endnote-3), Yuto Ozaki[[4]](#endnote-4), Danya Pavlovich[[5]](#endnote-5), Suzanne Purdy[[6]](#endnote-6), Adwoa Ampiah-Bonney[[7]](#endnote-7), Aleksandar Arabadjiev[[8]](#endnote-8), Flavia Arnese[[9]](#endnote-9), Joshua S. Bamford[[10]](#endnote-10),[[11]](#endnote-11), Brenda Suyanne Barbosa[[12]](#endnote-12), Ann-Kathrin Beck[[13]](#endnote-13), Anne Cabildo[[14]](#endnote-14), Gakuto Chiba[[15]](#endnote-15), Shahaboddin Dabaghi Varnosfaderani[[16]](#endnote-16), Simion Echim[[17]](#endnote-17), Shinya Fujii[[18]](#endnote-18), Shira Gabriel[[19]](#endnote-19), Massimo Grassi[[20]](#endnote-20), Lucrezia Guiotto Nai Fovino27, Jan Hajič jr.[[21]](#endnote-21), Kai Yi Han[[22]](#endnote-22), Niels Chr. Hansen[[23]](#endnote-23),[[24]](#endnote-24),[[25]](#endnote-25), Martin Hartmann[[26]](#endnote-26), Ye He[[27]](#endnote-27), Sotirios Kolios[[28]](#endnote-28), Wojtek Krzyżanowski[[29]](#endnote-29), Urise Kuikuro[[30]](#endnote-30), Dilyana Kurdova[[31]](#endnote-31),[[32]](#endnote-32), Fang Liu[[33]](#endnote-33), Psyche Loui[[34]](#endnote-34), Zoya Mikova[[35]](#endnote-35), Dayna Moya[[36]](#endnote-36), Rogerdison Natsitsabui[[37]](#endnote-37), Minja Niiranen[[38]](#endnote-38), Nozuko Nguqu[[39]](#endnote-39), Petr Nuska[[40]](#endnote-40), Florence Ewomazino Nweke[[41]](#endnote-41), Patricia Opondo[[42]](#endnote-42), Hineatua Parkinson[[43]](#endnote-43), Mark Lenini Parselelo[[44]](#endnote-44),[[45]](#endnote-45), Gemma Perry[[46]](#endnote-46), Peter Q. Pfordresher[[47]](#endnote-47), Piotr Podlipniak[[48]](#endnote-48), Tudor Popescu[[49]](#endnote-49), Robert M. Ross[[50]](#endnote-50), Zhuozhuang (Nelson) Shi[[51]](#endnote-51), Javier Silva-Zurita[[52]](#endnote-52),[[53]](#endnote-53), Ignacio Soto-Silva[[54]](#endnote-54),[[55]](#endnote-55), Barbora Štěpánková[[56]](#endnote-56), William Forde Thompson[[57]](#endnote-57), Sebastian Vaida[[58]](#endnote-58), Christina Vanden Bosch der Nederlanden[[59]](#endnote-59)

\*Corresponding author: Patrick E. Savage ([patrick.savage@auckland.ac.nz](mailto:patrick.savage@auckland.ac.nz))

**ABSTRACT: How cross-culturally consistent and general are relationships between song and speech? Ozaki et al. (2024)[1](https://www.zotero.org/google-docs/?jR7VMl) analysed singing and speaking from 75 individuals speaking 55 languages, concluding that songs are “slower and higher and use more stable pitches than speech”. However, the degree to which their findings would generalise to other speakers of their languages remains unclear. We will test the replicability of Ozaki et al.'s (2024) findings using audio recordings of matched singing and conversational speech from 15-30 individuals per language in 26 sites around the world, collected as part of a broader study of global song/speech relationships (Savage et al., 2025 [In Principle Accepted][2](https://www.zotero.org/google-docs/?640MuD)). For each site, we will replicate Ozaki et al.'s (2024) analyses for their three key features hypothesised to differ between song and speech: 1) pitch height (*f0*); 2) temporal rate (inter-onset interval of acoustic units [e.g., syllables/moras/notes]); 3) pitch stability (−|Δ*f0*|). This Programmatic Registered Report will allow us to test the replicability and generalizability of Ozaki et al.'s (2024) original findings in a rigorous and equitable manner by enabling each local research team to lead its own separate first-authored Stage 2 Registered Report based on a single shared Stage 1 protocol. This will allow us to test the replicability of Ozaki et al.'s (2024) original findings across diverse languages while also giving each team the flexibility to pursue their own exploratory analyses in directions most relevant to their local context. Regardless of results, our multi-site large-scale replications will enhance our understanding of cross-cultural relationships between music and language and provide a template for equitable global collaboration.**



**INTRODUCTION**

Music and language are two human cultural universals found in all known societies: separately (e.g., instrumental music, speech), but also together in the form of songs with words[3–10](https://www.zotero.org/google-docs/?9aCdfA). Previous research in fields including musicology, linguistics, psychology, anthropology, and neuroscience has identified neural, acoustic, and behavioural relationships between song and speech[11,8,12,13,9,14](https://www.zotero.org/google-docs/?iFGuG4). However, most previous research has been limited to speakers of English and other European languages, limiting the generality of conclusions that can be drawn[8,15,16](https://www.zotero.org/google-docs/?fNyVt1).

A key question is what, if anything, consistently distinguishes song from speech across languages? Steven Pinker famously dismissed music as an evolutionarily “useless” byproduct of adaptative traits such as language[17](https://www.zotero.org/google-docs/?8twDbs), while others have argued that the regular pitches and rhythms of music facilitate adaptative functions such as bonding individuals together or signaling group membership beyond the capacities of language[9,18,19](https://www.zotero.org/google-docs/?xZa00h). However, these debates have mostly been conducted in the absence of direct cross-cultural comparisons of actual singing and speaking[20–23](https://www.zotero.org/google-docs/?9YKtxG).

Recently, Ozaki et al. (2024) compared audio recordings of singing and speaking from 75 individuals speaking 55 languages, concluding that “*Globally, songs and instrumental melodies are slower and higher and use more stable pitches than speech*” and speculating that “*the slower and more stable pitches may facilitate synchronization, harmonization, and ultimately bonding between multiple individuals*”[1](https://www.zotero.org/google-docs/?Afw8Ye). However, these individuals were coauthors who were mostly researchers and professional musicians and were not representative of general speakers/singers of their languages, so the degree to which their findings would generalise to other speakers of their languages remains unclear[24](https://www.zotero.org/google-docs/?HgX5Mp). While this limitation was mitigated by comparisons of singing and speaking recordings from separate cross-linguistic databases[1,25–27](https://www.zotero.org/google-docs/?piD1y7), these databases did not include annotated segmentations into acoustic units (e.g., syllables/notes), meaning it was not possible to directly replicate Ozaki et al.'s (2024) analyses. And because Ozaki et al. (2024) only included one or a few speakers of each language and averaged their results across many languages, it is possible that some of their results may display different effects within different languages. For example, tonal languages such as Mandarin or Yoruba could conceivably use more stable spoken pitches, while “mora-timed” languages such as Japanese could be faster[28](https://www.zotero.org/google-docs/?d6BCcY).

Another key limitation of previous datasets is that they included only solo singing/speaking, whereas most singing and speaking throughout the world tends to be done in groups[10,29,30](https://www.zotero.org/google-docs/?nLGMMk). To overcome this, we have designed a new study in collaboration with over 80 researchers aiming to collect data on group singing and speaking in diverse languages from 1,800 participants across 60 different sites around the world[2](https://www.zotero.org/google-docs/?FzUTIH). However, there remains the challenge of annotating all this data in an efficient and equitable way.

**Equitable coauthorship in global collaboration**

One factor underlying the annotation issue is the broader challenge in big team science of ensuring equitable credit and authorship for all collaborators at all locations, rather than only having them listed as middle authors in a large coauthored publications (or not listed as coauthors at all)[31–35](https://www.zotero.org/google-docs/?d3mQTI). High-quality segmentation of cross-cultural audio recording corpuses requires many different researchers who are speakers of diverse languages to spend substantial time manually annotating audio recordings[1](https://www.zotero.org/google-docs/?MApctv). While one might hope that automated segmentation technology might reduce or eliminate this barrier, Ozaki et al.'s (2024) analyses found that automated segmentation tools are not yet sufficiently reliable for our goals, and in fact could have led to incorrect conclusions:

*While automatic segmentation can be effective for segmenting some musical instruments and animal songs (e.g., percussion instruments and bird song notes separated by microbreaths), we found that they did not provide satisfactory segmentation results compared to human manual annotation for the required task of segmenting continuous song/speech into discrete acoustic units such as notes or syllables…. For example, Mertens’ automated segmentation algorithm used by Hilton et al. mis-segmented two of the first three words “by a lonely” from the English song used in our pilot analyses (“The Fields of Athenry”), oversegmenting “by” into “b-y,” and undersegmenting “lonely” by failing to divide it into “lone-ly”...if we had used this automated method, then we would have mistakenly concluded that there is no meaningful difference in IOI [Inter-Onset Interval] rates of singing and speech…collaboration with native/heritage speakers who recorded and annotated their own speaking/singing relying on their own Indigenous/local knowledge of their language and culture allowed us to achieve annotations faithful to their perception of vocal/instrumental sound production that we could not have achieved using automated algorithms…This highlights that equitable collaboration is not only an issue of social justice but also an issue of scientific quality*”[1](https://www.zotero.org/google-docs/?rKdv20)).

Using the Programmatic Registered Report model[36,37](https://www.zotero.org/google-docs/?vkAWx5), we aim to overcome these challenges by enabling each local team to publish a first-authored article based on their segmentation and analysis of data from their own language/culture . We propose to create up to 27 Stage 2 Registered Reports (Table 1 & Fig. 1) that all follow the basic protocol of this Stage 1 Registered Report. These teams represent a subset of all 26 interested teams from the 60 global teams that have agreed to collect singing/speaking data from 15-30 participants each as part of a broader study on the behavioural effects of singing/speaking on social bonding[2](https://www.zotero.org/google-docs/?f6GuSH).

**Table 1. Planned languages/communities, authors and journals for Stage 2 manuscripts resulting from this Programmatic Stage 1 Registered Report protocol.**

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Language (/community)** | **Authors**  **(\*corresponding)** | **Preferred journal** |
| 1 | Mapudungun | Moya, Silva-Zurita, \*Savage, \*Soto-Silva | *Royal Society Open Science* |
| 2 | isiZulu | Nguqu, \*Savage, \*Opondo | *Peer Community Journal* |
| 3 | Yoruba | Nweke, [1-2 unnamed students], \*Savage | *Peer Community Journal* |
| 4 | te reo Māori | Pavlovich, Parkinson, Purdy, \*Savage | *Royal Society Open Science* |
| 5 | Danish | [1-2 unnamed students], \*Savage, \*Hansen | *Royal Society Open Science* |
| 6 | English (Auckland) | Shi, Purdy, \*Savage | *Peer Community Journal* |
| 7 | English (Gold Coast) | Perry, Ross, \*Savage, \*Thompson | *Royal Society Open Science* |
| 8 | English (Boston) | Han, \*Savage, \*Loui | *Royal Society Open Science* |
| 9 | English (NY) | Ampiah-Bonney, Gabriel, \*Savage, \*Pfordresher | *Peer Community Journal* |
| 10 | English (Toronto) | Cabildo, \*Savage, \*Vanden Bosch der Nederlanden | *Peer Community Journal* |
| 11 | Greek | Kolios, \*Savage | *Peer Community Journal* |
| 12 | Farsi | Dabaghi Varnosfaderani, \*Savage, \*Beck | *Peer Community Journal* |
| 13 | Italian | Guiotto Nai Fovino, \*Savage, \*Grassi | *Peer Community Journal* |
| 14 | Romanian | Vaida, Echim, \*Savage, \*Popescu | *Royal Society Open Science* |
| 15 | Spanish (Santiago) | Moya, Soto-Silva, \*Savage, \*Silva-Zurita | *Peer Community Journal* |
| 16 | Bulgarian | Kurdova, \*Savage, \*Mikova | *Peer Community Journal* |
| 17 | Czech | Nuska, Štěpánková, \*Savage, \*Hajič jr. | *Royal Society Open Science* |
| 18 | Macedonian | Arabadjiev, \*Savage | *Royal Society Open Science* |
| 19 | Polish | Krzyzanowski, \*Savage, \*Podlipniak | *Royal Society Open Science* |
| 20 | Japanese | Chiba, Ozaki, Fujii, \*Savage | *Peer Community Journal* |
| 21 | DhoLuo (Luo) | \*Opondo, \*Savage | *Royal Society Open Science* |
| 22 | Maasai | Parselelo, \*Savage | *Peer Community Journal* |
| 23 | Mandarin (Auckland) | Jia, Purdy, \*Savage | *Peer Community Journal* |
| 24 | Mandarin (Reading) | He, \*Savage, \*Liu | *Peer Community Journal* |
| 25 | Guarani | Barbosa, Kuikuro, Natsitsabui, \*Savage | *Peer Community Journal* |
| 26 | Finnish | Bamford, Niiranen, Hartmann, \*Savage | *Royal Society Open Science* |
| 27 | All 26 languages | Jia, Ozaki, Pavlovich, all other authors, Purdy, \*Savage | *Advances in Methods and Practices in Psychological Science\*\** |

*\*\*as “*[*Registered Replication Report” that “involve coordination between at least three (but preferably more) independent teams of researchers”*](https://rr.peercommunityin.org/about/pci_rr_friendly_journals#h_9155735686741652439066888)

By unifying these Stage 2 Registered Reports around a small shared set of three hypotheses for confirmatory testing, this should allow for coherence across different teams using shared methods, while also giving each team the flexibility to add additional exploratory analyses according to their own interests. For example, some sites are based in ethnomusicology departments and may add qualitative ethnographic analyses; others are based in psychology departments and may add extra analyses of demographic data; others are based in computer science departments and may add extra acoustic analyses. However, all teams will collect, analyse and report the same basic confirmatory hypothesis testing replicating Ozaki et al.'s (2024) original acoustic comparison of song and speech[1](https://www.zotero.org/google-docs/?KwWtSB).

**Hypotheses**

We hypothesize that Ozaki et al.'s (2024) findings of three key differences between singing and speaking will replicate in all languages and all sites tested. Specifically:

1. Singing uses higher pitch than speech
2. Singing is slower than speech
3. Singing uses more stable pitches than speech

For each of the 26 sites/Stage 2 reports, we will conclude whether or not each of Ozaki et al.'s (2024) three key findings (regarding tempo, pitch height, and pitch stability) generalise to their given language/location. We will also include a meta-analysis comparing all sites with Ozaki et al.'s (2024) original results to conclude whether their findings generalise across all studied languages/locations. For each of the three features in each language, we will conclude whether songs are significantly higher/faster/more stable than speech (replicating Ozaki et al.), significantly equivalent[38](https://www.zotero.org/google-docs/?EYs7j2) (contradicting Ozaki et al.), or inconclusive (if neither null hypothesis testing nor equivalence testing are statistically significant; see Table 2).

Since this is a Programmatic Registered Report where one Stage 1 protocol will result in multiple Stage 2 outputs, it is possible that different Stage 2 outputs will produce different results for different languages. This will allow us to evaluate criticisms that global analyses of cross-cultural trends fail to address the importance of internal diversity (“*How many exceptions are researchers willing to ignore?*”[28](https://www.zotero.org/google-docs/?gbnbnM)).

**2** **Registered Report Design Planner.** Includes three hypotheses (H1-H3).

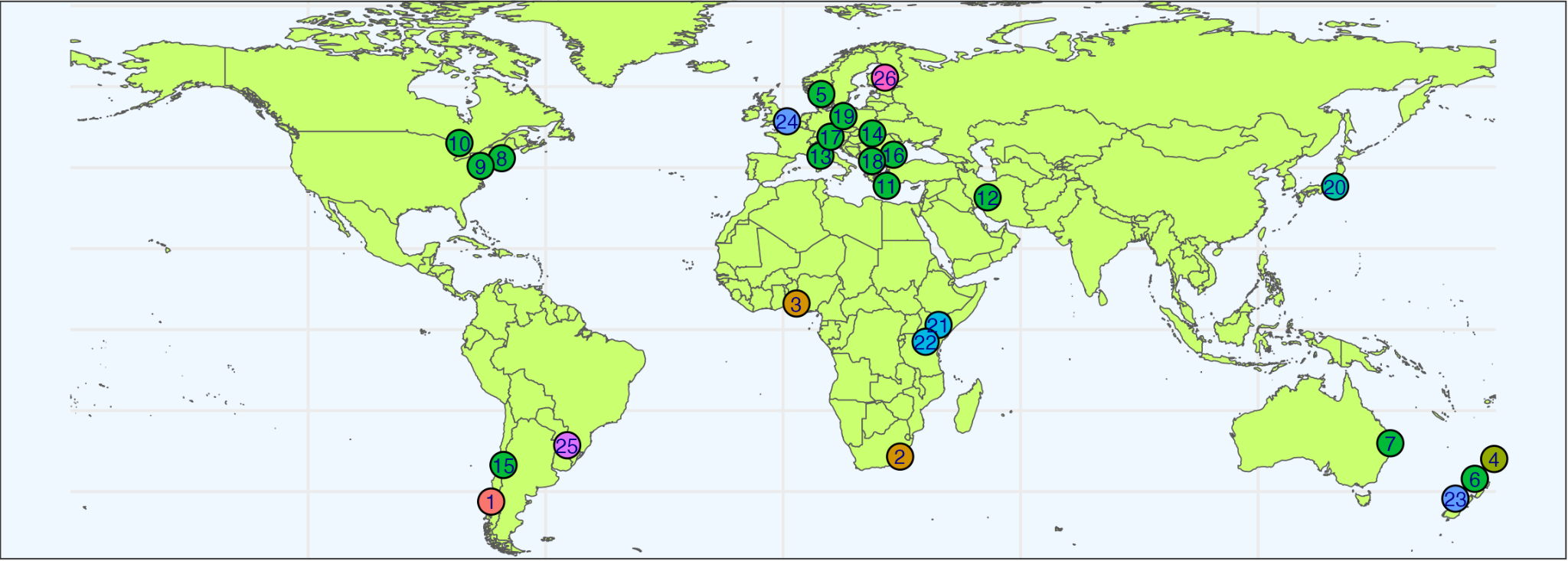
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Sampling plan** | **Analysis plan** | **Rationale for deciding the test sensitivity** | **Interpretation given different outcomes** | **Theory that could be shown wrong by the outcomes** | **Actual outcome** |
| Are any acoustic features reliably **different** between song and speech across cultures? | 1) Singing uses **higher pitch** than speech | **Use sample from ref.** [2](https://www.zotero.org/google-docs/?qqyatv)(n=15-30 pairs of audio recordings per language for up to 26 languages; **Fig. 1**). Each pair sung/ spoken by the same person. Sample size is determined by recruitment logistics. | **Replicate analysis from ref.** [**1**](https://www.zotero.org/google-docs/?eLMOu5)**:**  Meta-analysis framework **(Fig. 5)** calculates a paired effect size for **pitch height (*f0;* Fig. 4)** foreach song/ speech pair | Power analysis estimate (**β > .999**) given fixed samples from ref. [1](https://www.zotero.org/google-docs/?dSJn94) (n=15-30 per site) and effect size estimate fromref. [2](https://www.zotero.org/google-docs/?mMT22d) (*d* = 1.6). Bonferroni correction to account for multiple hypotheses tested (α = .05/3 =.017) | The null hypothesis of no difference in ***f*0** between sung and spoken pitch height is rejected if the population effect size is significantly larger than 0 **( *pre* > 0.5)**. If we fail to reject the null hypothesis, we will perform equivalence testing (using equivalence boundaries of **.39 < *pre* < 0.61** following ref. [2](https://www.zotero.org/google-docs/?LFtxGc)). Otherwise, we neither reject nor accept the hypothesis. | Ozaki et al.’s (2024) title claim that songs are “*slower and higher and use more stable pitches than speech*” | [TBD in Stage 2] |
| 2) Singing is **slower** than speech | Same as H1, but for **temporal rate (*inter-onset interval (IOI) rate*)** instead of **pitch height (*f*0)** [onsets are based on acoustic units corresponding to syllables or notes in English; see Fig. 4] | | | | |
| 3) Singing uses **more stable pitches** than speech | Same as H1, but for **pitch stability** (-|**Δ*f0***|) instead of **pitch height.** Lower estimated power of **β = .64** (for n=15) **to .94** (for n=30)given smaller effect size estimate (*d* = 0.7) | | | | |

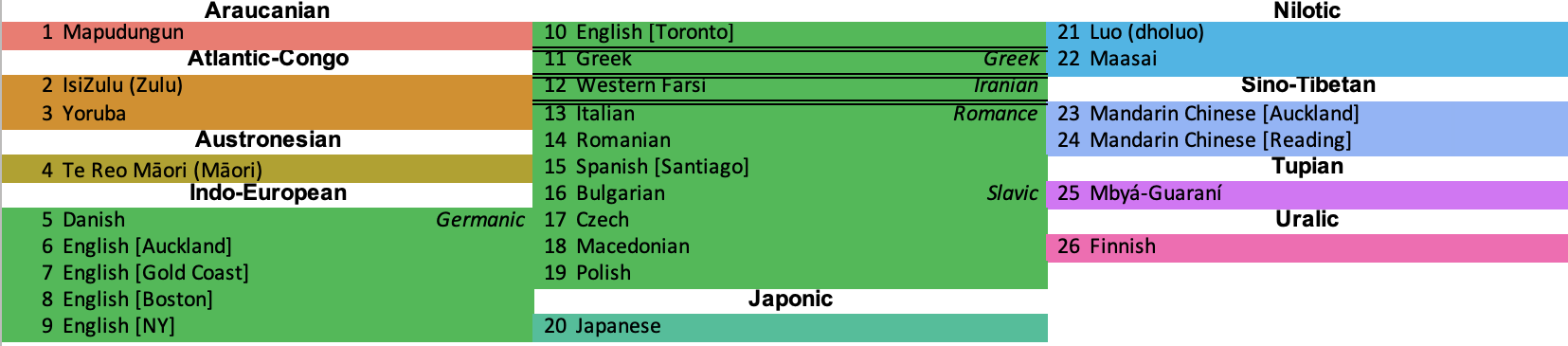
**METHODS**

**Study design**

Essentially, this Programmatic Registered Report combines two previous Registered Report protocols[1,2](https://www.zotero.org/google-docs/?PgVpn1) to replicate and extend Ozaki et al.'s (2024) acoustic analysis of 75 individuals singing/speaking solo[1](https://www.zotero.org/google-docs/?fVruB0) to Savage et al.’s larger sample of singing/speaking in many different groups[2](https://www.zotero.org/google-docs/?B7THWX).

Of the 60 sites that have agreed to collect singing/speaking data for Savage et al.’s global experiment, 26 are interested in using this Programmatic option to replicate/extend Ozaki et al.'s (2024) analyses for their own language/culture (Fig. 1 & Table 1). We will transcribe the monophonic singing and speaking audio from all 15-30 participants per site and segment the recordings into acoustic units (e.g., syllables/notes[1](https://www.zotero.org/google-docs/?Xp2L1s); see Fig. 4 below and videos in Appendix S1 for details) and use these segmented recordings to replicate Ozaki et al.'s (2024) key analyses.

****

**Figure 1. Map of the 26 planned languages/locations of the Stage 2 reports in Table 2 (subset of the 60 sites in ref. [2](https://www.zotero.org/google-docs/?ljxR4O)).**

The selection of these 26 sites was determined before any Stage 2 data collection began for Savage et al.[2](https://www.zotero.org/google-docs/?Xl1GXg). All 60 research sites were invited to participate. Inclusion in this Programmatic Registered Report depended only on the interest and availability of researchers at each site. In particular, they had to be willing to wait to begin data collection until this Programmatic Registered Report also receives In Principle Acceptance to ensure maximum bias control (Level 6[39](https://www.zotero.org/google-docs/?EfLDzA)).

#27 will replicate Ozaki et al.'s (2024) cross-linguistic meta-analysis approach to analyse average trends across all languages, which can be compared with the results of each individual Stage 2 reports #1-26 to achieve a much broader evaluation of the cross-linguistic replicability and generalisability of Ozaki et al.'s (2024) original results.

They are welcome and encouraged to explore additional unregistered exploratory analyses, but these must be “justified, methodologically sound, and informative”[39](https://www.zotero.org/google-docs/?PJBV5I).

Full details of study design can be found in the original publications[1,2](https://www.zotero.org/google-docs/?2o4F22) - here we summarize and quote key aspects relevant to this Programmatic protocol, highlighting any differences from the original protocols.

***Sample size and singing/speaking conditions***

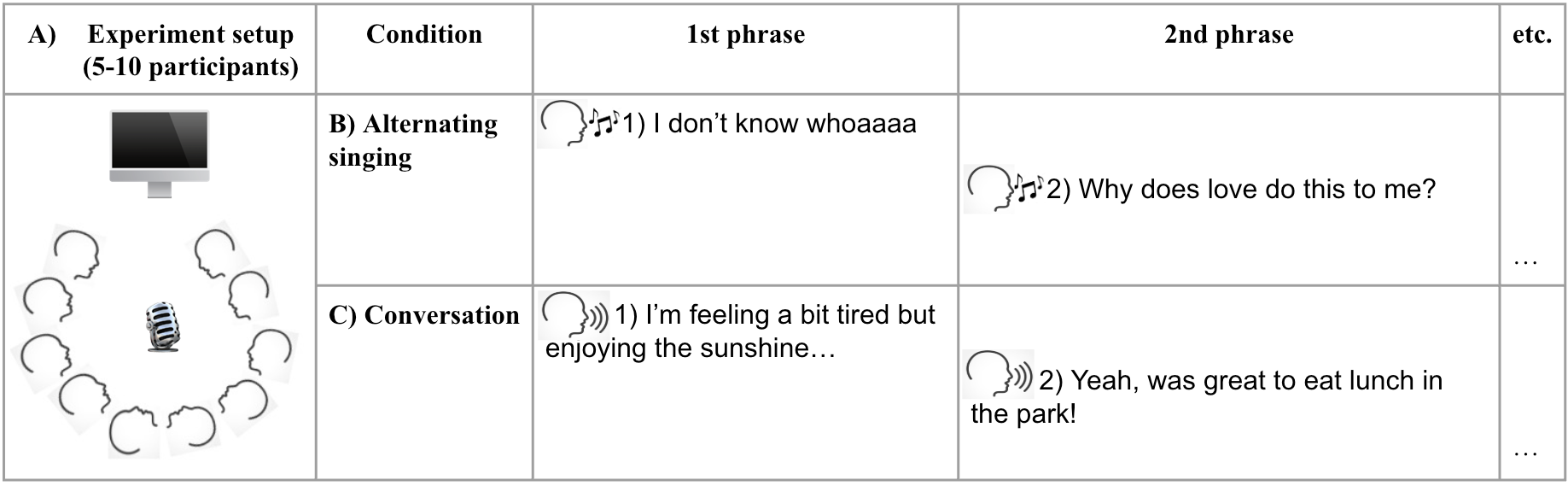
Each site will collect up to30 adults singing/speaking in three groups of 5-10 people at a time. The rationale for this sampling size was:

*Our multi-site design means that our sample size estimation is primarily constrained by the maximum feasible sample size. Through initial consultation with potential collaborators, we determined the optimal sample size that would allow us to maximise diversity across many sites while allowing experimenters to feasibly recruit relatively large groups of participants was up to 30 participants per site (max 10 per condition across three conditions) for each of each of the sites shown in Fig. 3. Pilot experiments suggested that getting all participants to show up at the agreed location on time was a major unavoidable logistical issue, and that groups of 4 or fewer may not be large enough to test the predictions of the social bonding hypothesis (since singing in small groups “might only produce a small increase… relative to conversation”*[*18*](https://www.zotero.org/google-docs/?lJ9NgQ)*). We thus decided to allow for experiments to run if at least 5 participants assembled on time for a given group.[2](https://www.zotero.org/google-docs/?wHKdfz)*

Note that this optimal sample size was determined by combining a quantitative survey asking each potential collaborator to list the maximum number of participants they could feasibly recruit with qualitative, holistic discussions via a series of Zoom meetings with collaborators.

Each of the three groups will engage in four conditions: group singing, alternating singing, conversation, lyric recitation. In contrast, Ozaki et al.’s (2024) confirmatory analyses compared 75 individuals singing a traditional song solo and then describing the meaning of the lyrics (also solo), comparing the first 20s of singing/speaking for each individual. (Ozaki et al.[1](https://www.zotero.org/google-docs/?nBZIwQ) also included a solo lyric recitation condition and an instrumental melody condition, although these were not included in their confirmatory analyses).

In order to maximize comparability with Ozaki et al.’s (2024) analyses, we will focus our confirmatory analyses on only the conversation and alternating singing conditions (Fig. 2), as these are the ones with monophonic recordings of individual singing/speaking voices to enable comparison of singing vs. speaking for each individual. (Any comparisons with the unison singing and lyric recitations conditions will be reserved for exploratory analyses.)



**Figure 2. Schematic overview demonstrating an example of the two conditions analysed in confirmatory analyses. Here, only the first two participants are shown singing or speaking sequentially, but the total number of participants will be between 5-10 per experiment (A). Text columns #1 and #2 represent the first and second phrases of alternating singing (B) / speaking (C) conditions. This example shows lyrics for “Why Does Love Do This To Me?”, the song chosen for participants using New Zealand English, and hypothetical conversation based on the ice-breaker prompt “How is your week going?”, but note that the actual song and conversation prompt will be different (and generally in a different language) at each site. (See Savage et al.[2](https://www.zotero.org/google-docs/?sxZUMr)’s Fig. 1 for an illustration of the lyric recitation and synchronised singing conditions not included in the current confirmatory analyses.)**

Note that, while Savage et al. compare social bonding effects of these different conditions using a between-participant design, our acoustic analysis proposed here instead compares singing vs. speaking for each individual participant in a *within*-participant design (i.e., comparing the same person’s singing voice with their speaking voice, following Ozaki et al.'s (2024) original acoustic analyses).

Audio recordings from all conditions will be made publicly available for analysis/replication. Participants will also be monitored by video. Video will not be published, but will be used to ensure experiment instruction compliance and to match singing/speaking recordings for individuals within the group.

Note that the original studies this Programmatic Registered Reports replicates and extends had very different minimum sample size requirements: Ozaki et al. (2024) specified a minimum sample size of 60 participants, while Savage et al. (2025) required a minimum sample size of 450 participants total (minimum of 30 sites, each with a minimum of 15 participants). For this Programmatic report, Stage 2 reports #1-26 will rely on acoustic data from the subset of sites from Savage et al., and thus also have a minimum of 15 participants each. However, it is possible that in some sites the number of participants with analyseable singing and speaking audio recordings may be fewer than the number of participants (e.g., if a participant does not have a chance to speak during the conversation condition).

We will specify a minimum of 10 participants per site. For the meta-analysis (#27), it is likely that some of the 26 proposed sites will not be able to complete their Stage 2 Reports within. However, a meta-analysis of even a small number of sites would still be valuable, meeting criteria such as *Advances in Methods and Practices in Psychological Science*’s “Registered Replication Reports (RRRs)”, which require “direct (i.e. close) replications in any area of psychology that involve coordination between at least three (but preferably more) independent teams of researchers” (<https://rr.peercommunityin.org/about/pci_rr_friendly_journals#h_9155735686741652439066888>). For consistency with this and with Ozaki et al.'s (2024) original minimum sample size of 60 participants for cross-linguistic meta-analysis, we will plan to continue the meta-analysis Stage 2 Report (#27 in Table 2) even with as few as 60 participants worth of data from as few as 3 sites are collected, analysed, and published as Stage 2 reports (i.e., minimum sample size for the meta-analysis of 60 participants from a minimum of 3 languages).

**Blinding**: Following ref.[40](https://www.zotero.org/google-docs/?TaIltX), Savage et al.[2](https://www.zotero.org/google-docs/?U38PtH) classify their experiment as “blinded” because the experimenters will be “not present during the manipulation and measurement of outcome variables”. Participants are unlikely to be aware of our three hypotheses tested here (particularly because the primary goal of Savage et al.’s experiment is to test social bonding, not acoustic features). Because the first authors of each article will do the annotation/segmentation of the recordings from their own language/culture, they cannot be blinded to hypotheses or condition. However, the likelihood that bias in the annotation process could influence the results is low, particularly since Ozaki et al. (2024) found high levels of inter-rater reliability in these annotations (intraclass correlation coefficients >.99).

**Randomisation:** At each site, the 15-30 participants will be randomly assigned into one of three groups. Each group completes the same four conditions (conversation, monophonic singing, unison singing, lyric recitation) but in different orders. When the (unaccompanied) monophonic singing condition follows the unison singing accompanied by karaoke-style accompaniment, participants may be influenced by having just heard and sung at the key and tempo matching this accompaniment. Likewise, it is possible that people may sing/speak differently depending on whether they have a conversation before or after singing. For these reasons, Savage et al. counter-balanced the order of conditions in the three participant groups as follows, enabling exploratory analyses of potential order effects:

***-Group 1:*** *1) conversation, 2) monophonic (alternating) singing, 3) unison singing, 4) lyric recitation*

***-Group 2:*** *1) unison singing, 2) lyric recitation, 3) monophonic (alternating) singing, 4) conversation,*

***-Group 3:*** *1) lyric recitation, 2) unison singing, 3) conversation, 4) monophonic (alternating) singing*

**Song/speech selection and participant inclusion criteria:**

Each site in Savage et al.[2](https://www.zotero.org/google-docs/?HQZNxf)will recruit 15-30 participants and choose its own song (cf. Table S1 from ref. [2](https://www.zotero.org/google-docs/?5Tp4It)) and conversation prompt using the following criteria. Note that the need to recruit participants to sing together in groups means it is not feasible to allow each participant to choose their own song as Ozaki et al. (2024) did:

***Participant inclusion criteria:***

*Each site will recruit participants who meet the following inclusion criteria:*

*-Age 18 or over*

*-Able to sing the song chosen for that site (with lyrics provided)*

*-Able to converse in the same language its lyrics are written in*

*-Have access to a phone or other device that can scan QR codes*

*-Willing to have their singing/speaking voice recorded and shared publicly (without being identified by name)*

***Song selection criteria***

*Each site has chosen a song that would be appropriate for their language/culture. The criteria for choosing a song were:*

***-lyrics*** *are mostly in the* ***same language*** *that participants will use for their group conversation (some lyrics in other languages or meaningless vocables like “la la” are acceptable, but should not make up the majority of the song)*

*-should be* ***easy for most potential participants from that society to sing together in synchrony*** *(e.g., unison, homophony) with karaoke-style pre-recorded instrumental accompaniment without needing to practise ahead of time. If possible, this should be in the form of a karaoke-style video with plain background and lyrics that appear in real-time to help the participants to sing at the right time, with no guide melody (e.g.,* [*https://youtu.be/OhRUYf\_yn\_s?si=eL4mt\_-utRwqrFMj&t=10*](https://youtu.be/OhRUYf_yn_s?si=eL4mt_-utRwqrFMj&t=10)*). If pre-recorded instrumental accompaniment would not be appropriate for a given site/society, an a cappella (unaccompanied) song may be chosen instead.*

*-should be the kind of song that would be appropriate to sing by young adults who don’t already know each other as a short “ice-breaker” exercise. As such,* ***songs that might easily become awkward, embarrassing, or offensive should be avoided*** *(e.g., children’s songs, songs with polarising content or associations such as national anthems or religious songs). However, these factors may vary from site to site (e.g., for some communities a national anthem or religious song might be the best choice, while in others it might be the worst). The experimenters from each site should interpret this on the basis of their own local knowledge.*

*-the song should take* ***between 2-3 minutes*** *to sing (you are welcome to modify the number of verses/choruses (including repeating the song) to make this happen*

*-if the song has* ***instrumental interludes/introductions/outros,*** *these should* ***not be longer than 1 minute*** *total and there* ***should still be 2-3 minutes of singing time*** *not including these instrumental sections.*

***Conversation ice-breaker question criteria:***

*Each team will choose their own unique ice-breaker question for the conversation condition (this can be taken directly from one of the following lists, adapted from them, or newly created themselves, but teams should all choose different questions):*[*https://www.mural.co/blog/icebreaker-questions*](https://www.mural.co/blog/icebreaker-questions)

[*https://museumhack.com/list-icebreakers-questions/*](https://museumhack.com/list-icebreakers-questions/)

[*https://www.parabol.co/resources/icebreaker-questions/*](https://www.parabol.co/resources/icebreaker-questions/)***Criteria for questions:***

*-Should not be about music/singing*

*-Should not use words/concepts that will be rated to create our dependent variable (i.e., “team”, “similar”, “trust”, “close”, “ties”, “common”).*

*-Should not ask sensitive/personally identifiable information (e.g., name, address, birthday, religion, sexuality, etc.)*

*-Should be capable of short answers (5-15 seconds per person)*

**Outcome-neutral criteria** (“*designed prior to knowledge of the results and …independent of the main study hypotheses*”[39](https://www.zotero.org/google-docs/?ZvOyut)**)**:

Savage et al.’s experiment will employ the following outcome-neutral exclusion criteria[2](https://www.zotero.org/google-docs/?6Js4sd)

*-Participants who fail to show up on time at the agreed location*

*-Participants who fail to complete the experiment and submit the Qualtrics survey*

*-Participants who are unable to complete the singing/speaking task in the specified language*

*-Participants who fail the attention check*

*-Participants with any confirmatory dependent variable’s data missing or corrupted due to technical glitches*

*-Participants with mean baseline social bonding scores of >80/100 (to avoid ceiling effects)*

*-Duplicate submissions by the same participant*

*-All participants from groups where “Instruction compliance” for the main experimental task (first condition) is judged unacceptable by the experimenter (<25 out of 100)*

*-Sites where useable data are only collected from fewer than 15 participants across all 3 groups*

Savage et al.’s criteria are focused on their confirmatory analysis goals of comparing social bonding rating data, rather than the acoustic recordings. Thus, it is possible that some participants from that study will pass those outcome-neutral criteria, but fail to provide reliable audio data (e.g., if the audio fails to record due to a technical glitch). It is also possible that some participants could fail their outcome-neutral controls (e.g., failing to submit the Qualtrics survey) but still provide useable audio recordings for this Programmatic protocol. Therefore, while our new protocol relies on audio recordings collected by Savage et al., these audio recordings will be subject to the following separate set of outcome-neutral inclusion criteria to ensure the recordings are of sufficient quality, duration, and reliability that they can be reliably used for our confirmatory hypothesis testing comparing acoustic features of singing vs. speaking.

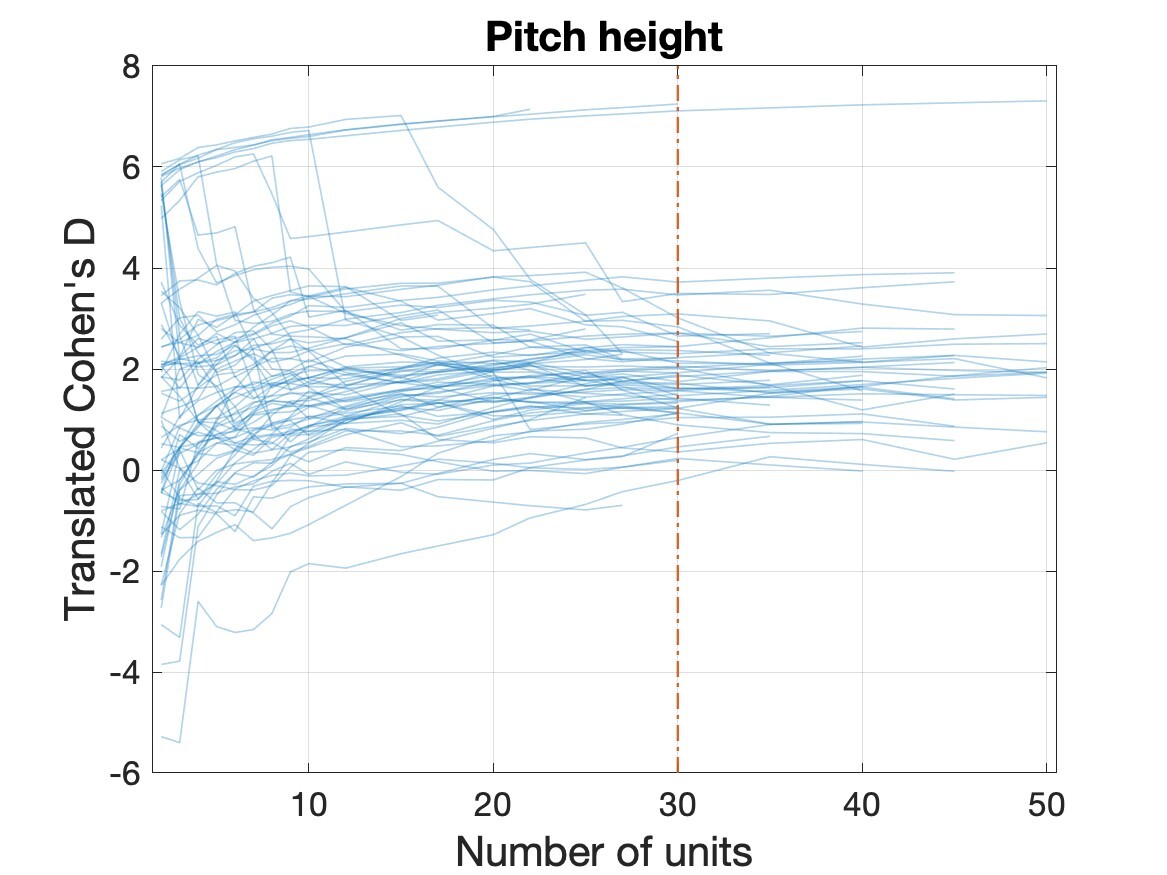
***Audio quality:***

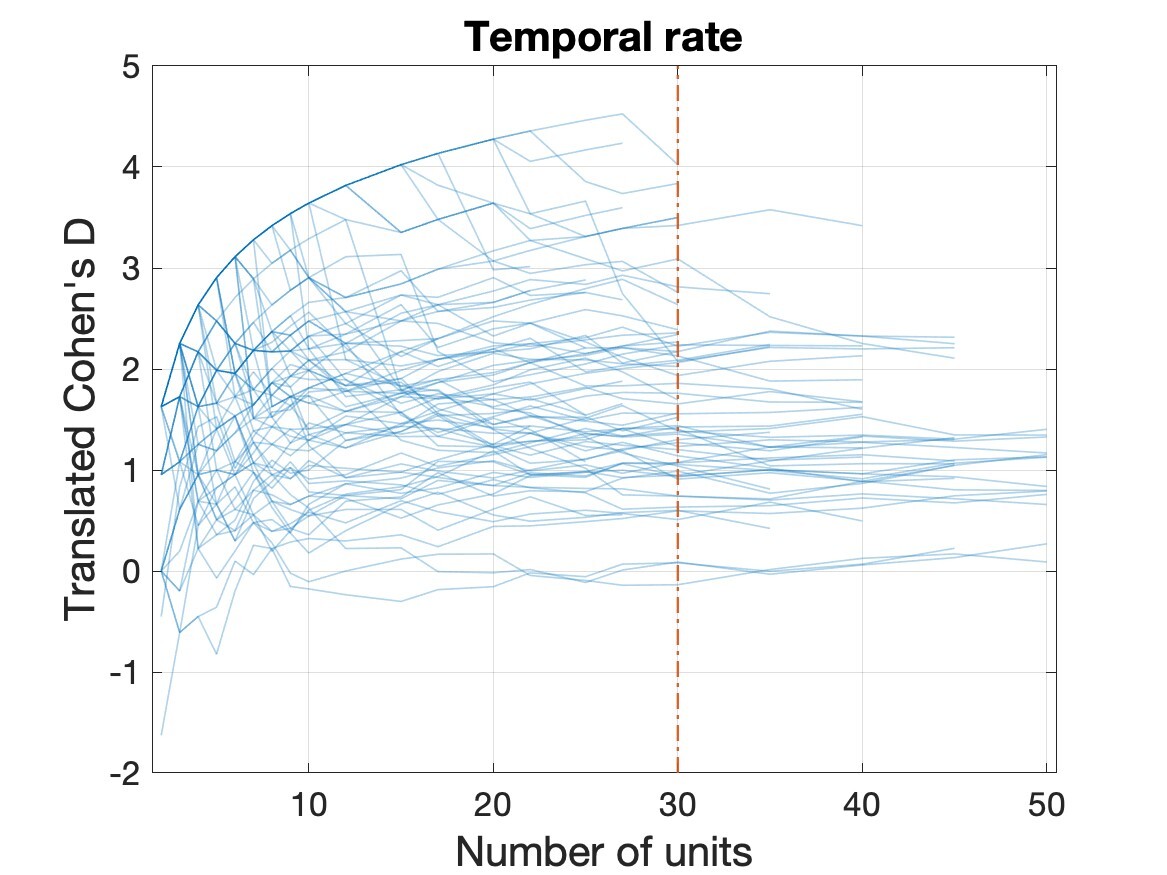
To ensure high-quality recordings, the equipment should include either a laptop running Audacity or a high-quality digital recorder (such as Zoom H4, H5, or H6) with an external microphone at a sampling rate of 44.1 kHz or 48 kHz. The recording device should be placed centrally among the participants to maintain an equal distance from each speaker.To be analysable, audio recordings must meet minimum standards of quality, such that our three confirmatory dependent variables (pitch height, temporal rate, and pitch stability) can be reliably measured (i.e., at least 10 units of matched singing/speaking whose fundamental frequency can be extracted; see Inclusion/Exclusion Criteria and Fig. 2 simulation below). This means they need to be recorded accurately with low enough noise and high enough quality that fundamental pitch can be automatically extracted using the pYIN algorithm[41](https://www.zotero.org/google-docs/?1EpaLO), and the units (syllables or notes) can be clearly determined.

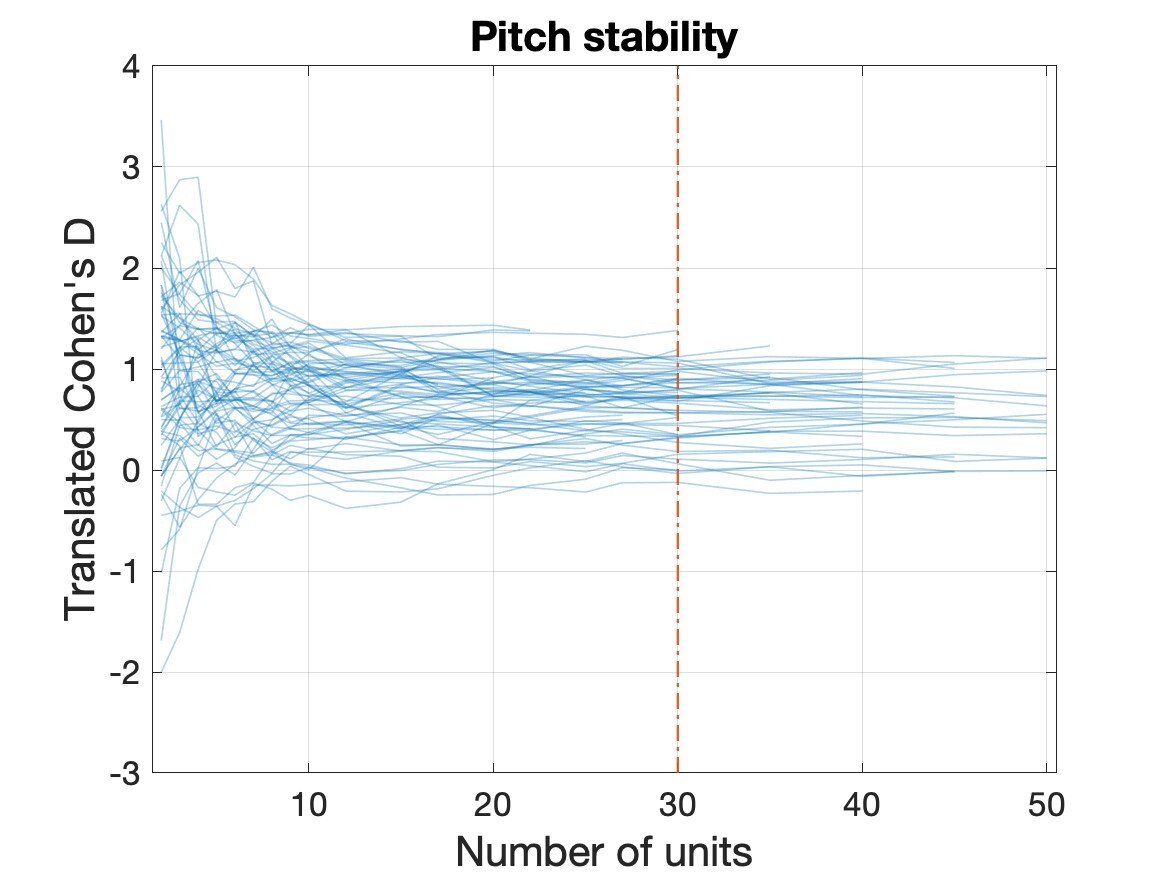
***Duration of singing/speaking***

One methodological challenge is that we cannot guarantee that all participants will speak for at least 20s in the group conversation condition as Ozaki et al. (2024) did for their solo spoken description condition. However, this limitation is mitigated by the fact that 20s was chosen by Ozaki et al. (2024) as the minimum length needed to reliably measure phrase contour for comparing pitch declination (which requires longer units separated by breaths), whereas the three features we propose to test here do not include pitch declination but only features requiring shorter acoustic units (e.g., syllables/notes).

While Ozaki et al. (2024) annotated the first 20s of both singing and speaking, this resulted in different numbers of acoustic units, ranging from as low as 18 sung notes (Thai) to as high as 113 spoken syllables (Greek and Hebrew). Since the number of acoustic units is the limiting factor for our three proposed features, we simulated the effects of using different numbers of acoustic units (ranging from 2-50 syllables/moras/notes) from Ozaki et al.'s (2024) data in order to optimise the amount of annotation needed for reliable results (Fig. 3).







**Figure 3. Simulation using Ozaki et al.'s (2024) singing/speaking data from 75 individuals speaking 55 languages showing effects of varying number of spoken/sung units analysed on estimated effect sizes for the proposed three confirmatory dependent variables (pitch height, temporal rate, and pitch stability). Estimates are unreliable below 10 acoustic units, but stabilise and do not change substantially beyond our proposed cutoff of 30 acoustic units (dashed vertical line).**

Simulation analysis of Ozaki et al.'s (2024) data (Fig. 3) suggests that effect size estimates from fewer than 10 acoustic units each of singing/speaking are not reliable, but that using more than 30 acoustic units each does not substantially increase reliability. Participants will be singing/speaking for an average of around 15-30 seconds each (5-10 participants speaking/singing for 2-3 minutes each), but this will vary particularly in the conversation condition as some individuals tend to speak more/less than others and there is no forced alternation of speaking in the experimental condition (unlike the alternating singing condition which will result in less variation in singing time).

From a logistics perspective, the most time-consuming part of analysis is the manual annotation of acoustic units. On average, it takes approximately one minute to annotate two acoustic units (e.g., syllables/notes). Thus we estimate it should take the lead author from each site approximately 4-8 hours to annotate 30 acoustic units each of speaking and singing recordings (following the instructions in Appendix S1) from their 15-30 participants (corresponding to approximately 8 seconds of spoken conversation and 15 seconds of singing per participant).

Given the simulation results and experimental logistics, we propose analysing the first 30 acoustic units for all singers/speakers, but including all singers/speakers who produce at least 10 useable acoustic units of both singing and speaking. Other participants (e.g., those who don’t speak at all or produce less than 10 syllables of speaking during the conversation condition) will be excluded from these acoustic analyses. Therefore, we will analyse the first 30 acoustic units of singing and of speaking, and exclude any participants who produce fewer than 10 analyseable acoustic units of either singing or speaking.

**Inter-rater reliability:** Before annotating audio, each coder will watch the training tutorial video (Appendix S1). We will measure inter-rater reliability (IRR) following Ozaki et al. (2024) by having author Jia independently re-annotate onsets of singing and speaking from one randomly selected participant from each Stage 2 report. Like Ozaki et al., Jia will be blind to the specific onset timings annotated by the original coder, but will have access to their segmented texts (since otherwise Jia will not know the correct way to segment acoustic units such as syllables/notes spoken/sung in languages she does not speak). For reference, Ozaki et al. (2024) found “strong intraclass correlations (>0.99)” when using this method to compare 10s excerpts of singing vs speaking from 8 individuals randomly selected from the full sample of 75 individuals. Any sites with intraclass correlations of less than 0.75 (a typical threshold for “good”[42](https://www.zotero.org/google-docs/?ikqMiv) or “excellent”[43,44](https://www.zotero.org/google-docs/?0TuBtj) reliability) will be independently re-checked by Savage for another randomly selected song. If this is also less than 0.75, then all songs from that site will be checked and re-annotated until they achieve coefficients of at least 0.75.

**Inclusion/exclusion criteria summary:**

***Inclusion criteria:***

-The first 30 useable acoustic units (e.g., syllables/notes) of monophonic singing/speaking audio collected from each of the 15-30 adult participants from each site collected by Savage et al.[2](https://www.zotero.org/google-docs/?0Mrx4k) will be used in confirmatory analyses

-Minimum sample size for Stage 2 reports #1-26: 10 participants

-Minimum sample size for Stage 2 report #27: 60 participants speaking at least 3 different languages

-For the meta-analysis (#27) all useable data from Stage 2 reports #1-26 collected and analysed within 18 months after In Principle Acceptance will be included

***Exclusion criteria:***

-Sections with simultaneous singing/speaking and vocalizations that are not singing or speaking (e.g., laughing, groaning)

-Audio where the fundamental pitch cannot be automatically extracted

-Participants with fewer than 10 analyseable acoustic units for both singing and speaking (e.g., 10 spoken syllables of conversation and 10 notes of monophonic singing for English)

-Audio whose annotations do not achieve “good” levels of inter-rater reliability (intraclass correlations at least 0.6)

-All participants from groups where “Instruction compliance” for the conversation or alternating singing conditions is rated unacceptable by the experimenter (<25 out of 100). (These ratings will be made immediately after the experiment before viewing or analysing the data, to avoid confirmation bias)

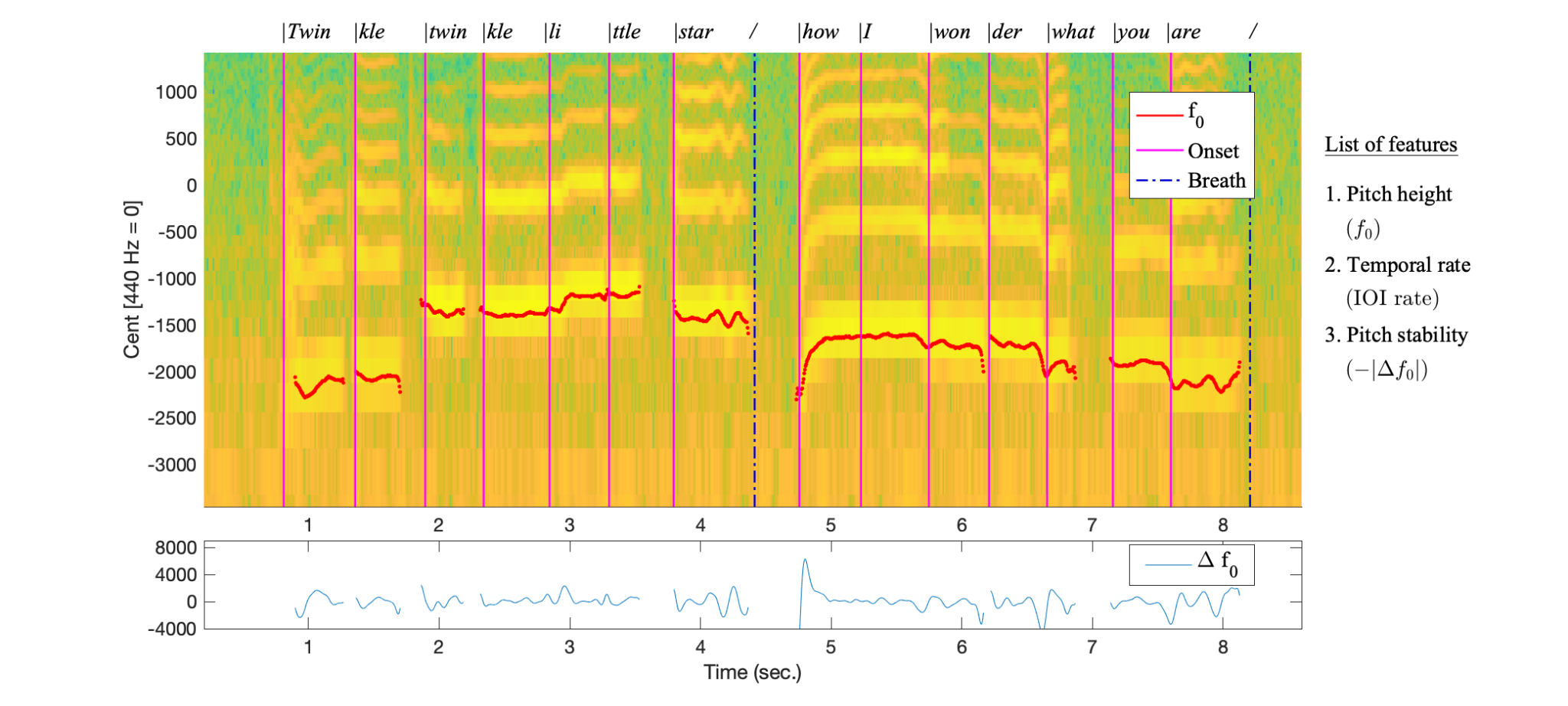
-If any sites choose to withdraw (i.e., not to publish a Stage 2 report), their data will also not be included in the meta-analysis confirmatory analyses in #27. In such cases, the meta-analysis will report the reasons for withdrawal (e.g., lack of time to analyse data or write up analyses; researcher graduating/changing jobs; data not meeting inclusion criteria standards) and describe how much, if any, of the data were collected/analysed before withdrawal, summarising any preliminary results if they exist. Note that we cannot commit to analysing all data if sites withdraw because our proposed acoustic analyses require time-consuming manual annotation by researchers with knowledge of the local language/music. However, we commit to not making decisions about whether or not to withdraw based on how these affect our conclusions.

## ***Analysis Plan***

**Independent variable: Vocal modality** (alternating singing vs conversational speaking)

**Dependent variables** (all identical to the first three from Ozaki et al.[1](https://www.zotero.org/google-docs/?IRsaWp); see Fig. 4):

1. **Pitch height** (fundamental frequency [f0] in hertz)
2. **Temporal rate** (interonset interval [IOI] rate in hertz)
3. **Pitch stability** (−|Δf0| in cents per second)

****

**Figure 4. Schematic illustration of the three features analysed for confirmatory analyses, using a recording of author Savage singing the first two phrases of “Twinkle Twinkle Little Star” as an example. This figure is identical to Figure 3 in Ozaki et al., but only shows the three features proposed to test here of pitch height, temporal rate, and pitch stability. Onset annotations used to calculate Inter-Onset Intervals (IOIs) are based on the segmented texts displayed at the top of the spectrogram (breaths are excluded from IOI calculations). For this English song, these onsets correspond to syllables, which also correspond to sung notes, as Twinkle Twinkle uses one note per syllable. However, the choice of acoustic units can vary depending on the language and song**[45](https://www.zotero.org/google-docs/?BneTYy)**. For example, in Japanese it typically corresponds to a “mora” (e.g., みんな = mi|n|na) and songs often use multiple notes per syllable**[46](https://www.zotero.org/google-docs/?fV5AvU)**. Following Ozaki et al.**[1](https://www.zotero.org/google-docs/?LoVCTm)**, the appropriate segmentation unit for each language is chosen by the lead researcher analysing that language (who is also a speaker of the language). In most cases, these units will be syllables or moras for speaking and notes for singing. Note that our analyses are intended to address speaking/singing rate, so higher- or lower-level units such as stressed syllables, metric downbeats, or phonemes**[8,47](https://www.zotero.org/google-docs/?yUtgwJ) **are not the focus of our confirmatory analyses.**

For f0 extraction, we will use only the fully automated probabilistic YIN (PYIN)[41](https://www.zotero.org/google-docs/?hDsJX5) method used by Ozaki et al. (2024) without also employing their custom semi-automated tool, since their analyses found “negligible differences” between results using the fully automated and semi-automated methods (see Ozaki et al.'s (2024) Fig. 6 and “Exploring recording representativeness and automated scalability” section for full details). Note this automated f0 extraction method is only relevant to the measurement of pitch height (H1) and pitch stability (H3). Ozaki et al.'s (2024) analysis of temporal rate (H2) showed that automated methods of syllable/note segmentation are not reliable enough to replace manual annotation (see quote in the Introduction above).

### 

### ***Statistical analysis:***

We will follow essentially the same analysis methods as Ozaki et al. (2024) using a meta-analysis framework to compare effect sizes from each within-participant singing vs. speaking comparison across many different participants. The main differences are:

1. we are only testing three hypotheses (pitch height, temporal rate, and pitch stability) rather than Ozaki et al.'s (2024) six
2. Each site will test whether the hypotheses replicates for its own language/society, rather than comparing across many different languages simultaneously as Ozaki et al. (2024) did (though we will also run the cross-linguistic comparison for the final meta-analysis of all 26 languages/cultures)

The full analysis plan is adapted from Ozaki et al. (2024) as follows:

*We use null hypothesis testing to test whether the effect size of the difference between song and speech for a given feature is null. There are various ways to quantify the statistical difference or similarity (e.g., Kullbak-Leibler divergence, Jensen-Shannon divergence, Earth mover’s distance, energy distance, Ln norm, Kolmogorov-Smirnov statistic). Here we focus on effect sizes to facilitate interpretation of the magnitudes of differences.*

*Since our main interest lies in the identification of whether three features - pitch height, pitch stability, and temporal rate - demonstrate differences between song and speech, we perform the within-participant comparison of these features between the pairs of singing and speech, using the alternating singing and conversation conditions as proxies for singing and speech, respectively (comparisons with synchronised singing and synchronised recitation are reserved for exploratory analyses). Terms in the computed difference scores are arranged so that for our predicted differences (H1-H3), a positive value indicates a difference in the predicted direction [cf. Fig. 5].*

*Evaluation of difference in the magnitude of each feature is performed with nonparametric relative effects[48](https://www.zotero.org/google-docs/?Uk0W2j) which is also known as stochastic superiority[49](https://www.zotero.org/google-docs/?VJLl3I) or probability-based measure of effect size[50](https://www.zotero.org/google-docs/?Z8IPQA). This measure is a nonparametric two-sample statistics and allows us to investigate the statistical properties of a wide variety of data in a unified way.*

*We apply the meta-analysis framework to synthesize the effect size across recordings to make statistical inference for each hypothesis [see Fig. 8 in Ozaki et al. (2024) for graphic overview]. In this case, the study sample size corresponds to the number of data points of the feature in a recording and the number of studies corresponds to the number of participants. We use Gaussian random-effects models[51,52](https://www.zotero.org/google-docs/?ogzTt1), and we frame our hypotheses as the inference of the mean parameter of Gaussian random-effects models, which indicates the population effect size, as follows:*

*The Gaussian random-effects model used in meta-analysis is[51,52](https://www.zotero.org/google-docs/?ipGZqt):*

**

* is the effect size (or summary statistics) from th study,  is the study-specific population effect size,  is the variance of th effect size estimate (e.g. standard error of estimate) which is also called the within-study variance,  is the population effect size,  is the between-study variance, and  is the number of studies. In our study,  is the relative effect and  is its variance estimator[48](https://www.zotero.org/google-docs/?f9Yh8u). In addition, the term “studies” usually used in meta-analysis corresponds to recording sets. This model can also be written as*

******

*Our null hypotheses for the features predicted showing difference is that the true effect size is zero (i.e. relative effects of 0.5). We test three features, and thus test three null hypotheses.*

*Since we test multiple hypotheses, we will use the false discovery rate method with the Benjamini-Hochberg step-up procedure[53](https://www.zotero.org/google-docs/?cr4K3n) to decide on the rejection of the null hypotheses. We define the alpha level as 0.05. We test whether the endpoints of the confidence interval of the mean parameter of the Gaussian random-effects model are larger than 0.5. We use the exact confidence interval proposed by Liu et al.[52](https://www.zotero.org/google-docs/?c6sFZw) and Wang and Tian[54](https://www.zotero.org/google-docs/?w7vuWs) to construct the confidence interval.*

*For the equivalence testing, we first estimate the mean parameter (i.e., overall treatment effect) with the exact confidence interval (98, 100) and the between-study variance with the DerSimonian-Laird estimator[55](https://www.zotero.org/google-docs/?FBTHoW). Since Gaussian random- effects models can be considered Gaussian mixture models having the same mean parameter, the overall variance parameter can be obtained by averaging the sum of the estimated between-study variance and the within-study variance. Then, we plug the mean parameter and overall variance into Romano’s[56](https://www.zotero.org/google-docs/?4TDfD8) shrinking alternative parameter space method to test whether the population mean is within the equivalence region as stated in Table 1 (i.e., relative effects of 0.39 and 0.61).*

## ***Power analysis***

Our sample size of 15-30 per site is fixed by multi-site collaboration logistics from ref. [2](https://www.zotero.org/google-docs/?3gDnUm). Therefore, rather than determining the sample size needed to find a given effect size with a given power, our power analysis estimates what power we can expect to achieve *a priori* given our fixed sample size and effect size estimates from the study to be replicated[1](https://www.zotero.org/google-docs/?uukNPw). Note that, while using published effect sizes as estimates is often discouraged due to publication bias and the need to power for the Smallest Effect Size Of Interest, we believe it is an appropriate strategy given the unique sampling design of this report and the fact that the published effect sizes were themselves measured using a Registered Report (and therefore are not affected by publication bias).

Thus, we estimate the following power to reject false negatives (β**)** to test our three hypotheses given a fixed sample size (**n = 15-30** participants per site), fixed site-wide false positive rate (**α = .05/3** after Bonferroni correction for testing 3 hypotheses), and varying effect size estimates:

**H1 (pitch height):** **β > .999** (effect size: ***d* = 1.6**)

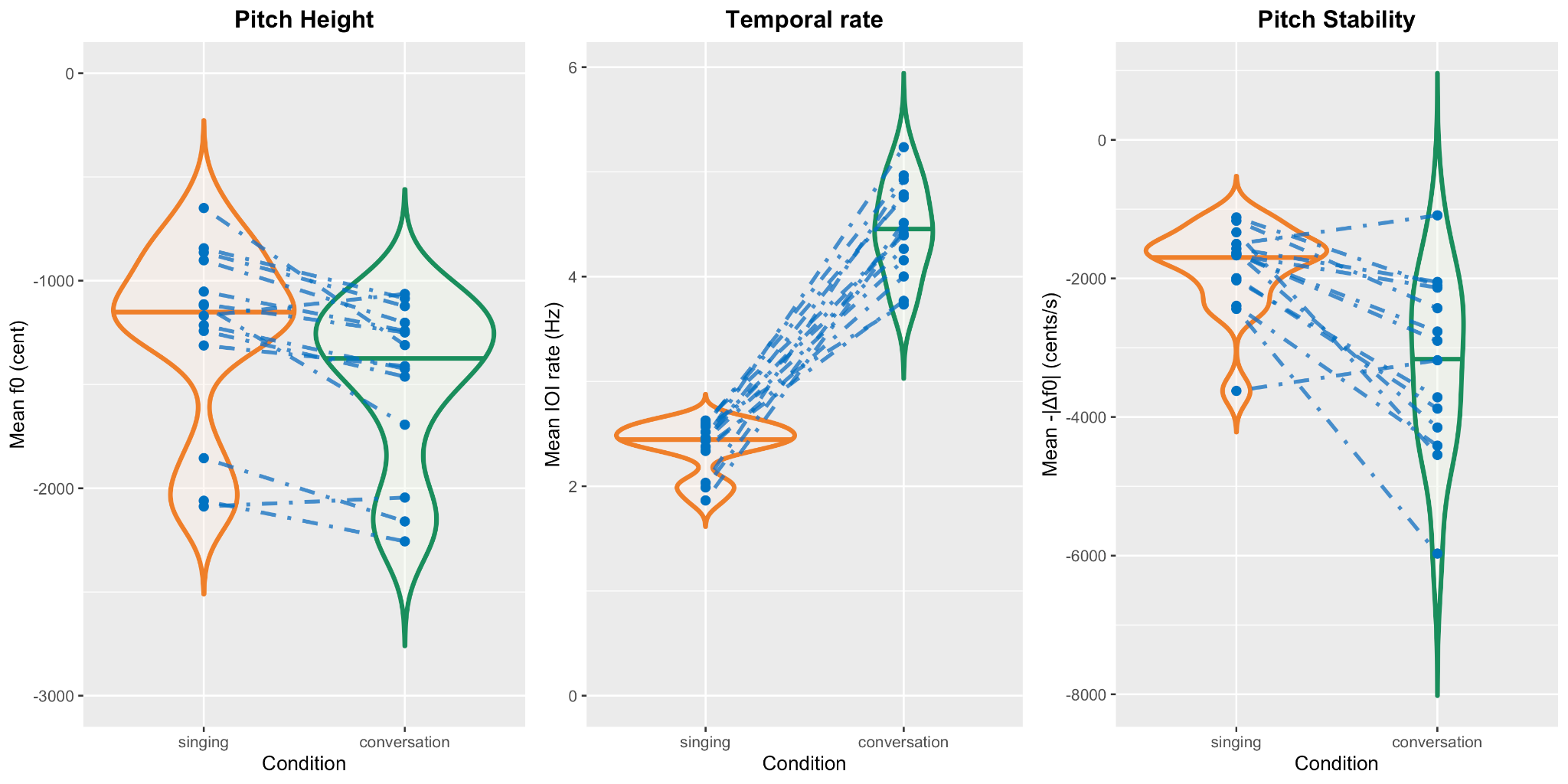
**H2 (temporal rate):** **β > .999** (effect size: ***d* = 1.6**)

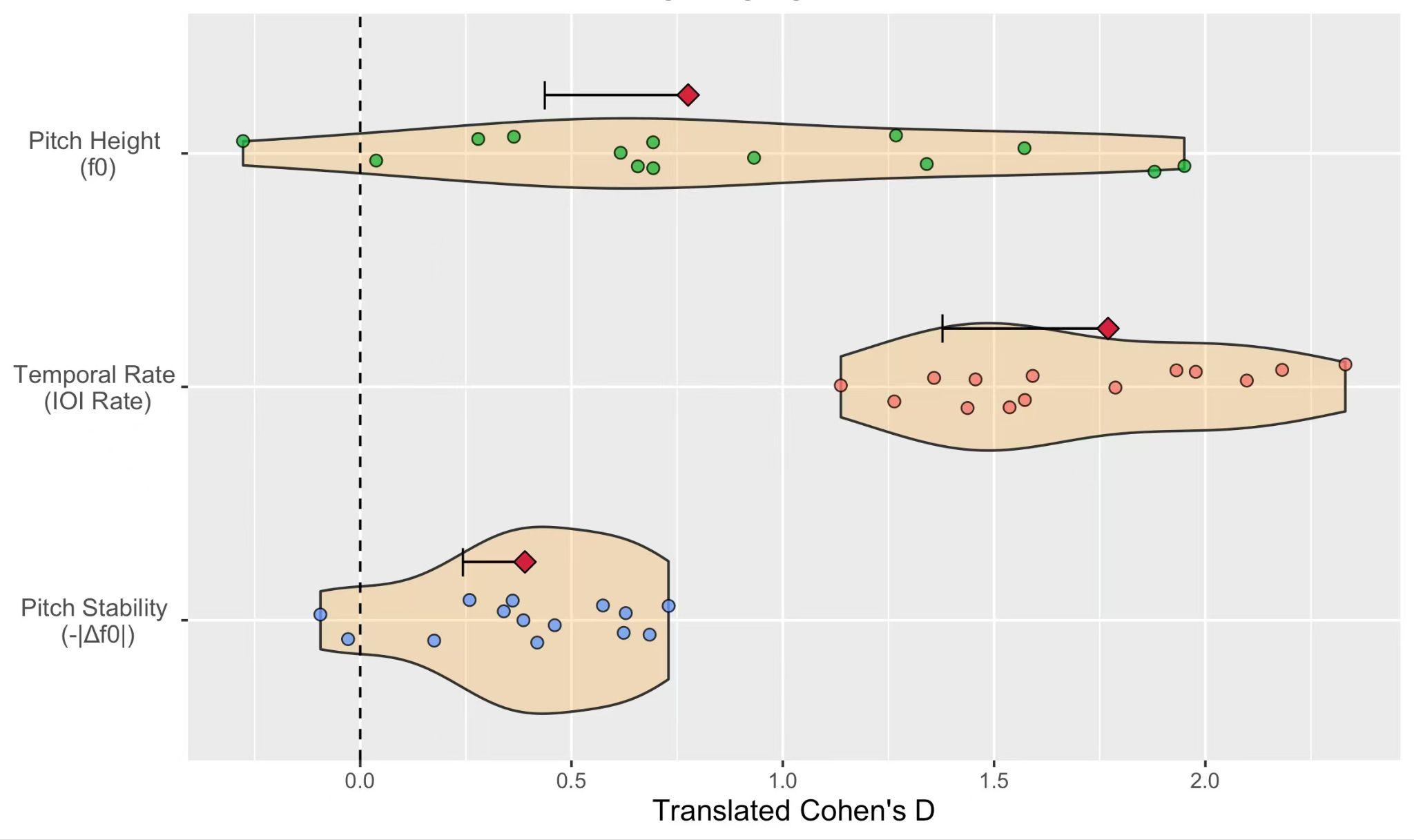
**H3 (pitch stability):** **β =** **.64** [for n=15] to **.94** [for n=30] (effect size: ***d* = 0.7**)

## ***Pilot data***

Figure 5 shows our proposed analyses (replicating Figs. 4-5 from Ozaki et al.) for the same pilot experiment shown in Savage et al.’s Figure 4[2](https://www.zotero.org/google-docs/?WyhqGw) (i.e., n=14 participants from Auckland, New Zealand singing/speaking in English in three groups of 4-5 people at a time in June/July 2024). Note that these pilot experiments will not be used for confirmatory analyses in Stage 2 reports resulting from either ref. [2](https://www.zotero.org/google-docs/?EUU7vu) or the current protocol - they are only used for pilot analyses.

Notably, all of the 14 participants produced enough analyseable data of high enough quality that no participants had to be excluded. This suggests that the various changes made from Ozaki et al.'s (2024) design (e.g., a single microphone recording multiple individuals engaged in alternating singing/free conversation) did not compromise our ability to accurately replicate their analyses. The general trends (i.e., song again appearing higher, slower, and more stable than speech) also indicate that any changes in experimental design are unlikely to result in major changes to results.





**Figure 5. Pilot analyses replicating analyses of pitch height, temporal rate, and pitch stability (cf. Ozaki et al. (2024) Figs. 4-5) for 14 pilot participants singing/conversing in English in Auckland, New Zealand (cf. Savage et al. Fig. 4).**

Although we do not want to over-interpret minor differences in pilot data, we do note that, while the mean effect size for temporal rate (translated Cohen’s *D*=1.8) is quite similar to Ozaki et al.'s (2024) (*D*=1.6), the effect sizes for pitch height (*D*=0.8) and pitch stability (*D*=0.4) appear somewhat smaller than the original analysis (*D*=1.6 and *D*=0.7, respectively; see Table 2). While any such differences that may arise in our Stage 2 data could reflect normal variation, they could also reflect systematic differences such as language-specific effects, differences in musical training of participants, or the fact that unlike Ozaki et al.'s (2024) design, singing in a group context (even when each singer sings one solo line at a time) usually results in all singers matching their singing to a single shared key (tonal centre), which may not have been the one best suited to the natural range of all singers’ voices. These possibilities (and others, such as potential order effects described in the “Randomisation” section above) may be worth addressing in the Discussion section and Exploratory Analysis sections of the resulting Stage 2 reports.

**Table 2. Comparison of effect sizes (translated Cohen’s *D*) for the three proposed features from Ozaki et al.'s (2024) original study (n=75 participants[1](https://www.zotero.org/google-docs/?aXDIzX)) with pilot data from the current protocol (n=14 participants from Savage et al.[2](https://www.zotero.org/google-docs/?uAu7EE))**

|  |  |  |
| --- | --- | --- |
| **Feature** | **Original effect size** | **Pilot effect size** |
| Pitch height | 1.6 | 0.8 |
| Temporal rate | 1.6 | 1.8 |
| Pitch stability | 0.7 | 0.4 |

## **Acknowledgments**: We thank all the previous coauthors of refs.[1,2](https://www.zotero.org/google-docs/?SaC4LA) for discussion of ideas leading to this protocol. We also thank Chris Chambers and Rachael Shaw for discussion regarding *Peer Community In Registered Reports*’ Programmatic format. We thank all 12 *Peer Community In Registered Reports* reviewers and recommenders who helped improve this protocol (Corlina Logan, Nai Ding, Fermin Moscoso del Prado Martin, and Makiko Sadakata) and the two previous protocols that it combines (Chris Chambers, Nai Ding, and Bob Slevc for ref.[1](https://www.zotero.org/google-docs/?CG1XAn), Katherine Moore, Melissa Brandon, Erin Hannon, Manuela Maria Marin and Anja Göritz for ref.[2](https://www.zotero.org/google-docs/?AkHHpr)). Their reviews and editorial recommendations are publicly available at [57,58](https://www.zotero.org/google-docs/?JKCxmt).

**Funding:** This work is supported by funding from the New Zealand government as follows:

-Marsden Fast-Start Grant from the Royal Society Te Apārangi (MFP-UOA2236 to Savage, Purdy, and Opondo)

-Rutherford Discovery Fellowship from the Royal Society Te Apārangi (RDF-UOA2202 to Savage)

## The funders have/had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## ***Ethics information***

As this protocol involves secondary analysis of data collected using a previously approved protocol[2](https://www.zotero.org/google-docs/?PWdUr1), no new ethics approvals are required.

**Data availability:** Pilot data are available at <https://osf.io/e4pqv/>, where full data will also be uploaded after Stage 2 data collection.

**Code availability:** Analysis code is available at <https://github.com/comp-music-lab/manyvoices3>

**Author contributions**:

Conceptualisation: Savage, Ozaki, Jia, Purdy, Opondo, Fujii, Pfordresher

Investigation [participated in pilot experiments]: Pavlovich, Jia, Savage, Purdy, Chiba, Nweke, Nguqu, Opondo

Analysis: Jia, Ozaki, Savage, Purdy

Data collection: [See Table 2 for details of which coauthors will collect data from which sites during Stage 2. The first author from each group will lead data collection and will segment their recordings into acoustic units (e.g., syllables/notes)]

Writing: Savage, Jia, Ozaki

Project administration: Savage

Funding acquisition: Savage, Purdy, Opondo

**Conflicts of interest.** The authors declare no financial conflicts of interest. Savage and Bamford are Recommenders at *Peer Community In Registered Reports*.

**REFERENCES**

[1. Ozaki, Y.](https://www.zotero.org/google-docs/?rsUtIV)

[*et al.* Globally, songs and instrumental melodies are slower and higher and use more stable pitches than speech: A Registered Report. *Sci. Adv.* **10**, adm9797 (2024).](https://www.zotero.org/google-docs/?rsUtIV)

[2. Savage, P. E.](https://www.zotero.org/google-docs/?rsUtIV)

[3. *Universals of Language*. (MIT Press, Cambridge, MA, 1963).](https://www.zotero.org/google-docs/?rsUtIV)

[4. Brown, D. E. *Human Universals*. (Temple University Press, Philadelphia, 1991).](https://www.zotero.org/google-docs/?rsUtIV)

[5. Fitch, W. T. Unity and diversity in human language. *Philos. Trans. R. Soc. B Biol. Sci.* **366**, 376–388 (2011).](https://www.zotero.org/google-docs/?rsUtIV)

[6. Brown, S. & Jordania, J. Universals in the world’s musics. *Psychol. Music* **41**, 229–248 (2013).](https://www.zotero.org/google-docs/?rsUtIV)

[7. Mehr, S. A. *et al.* Universality and diversity in human song. *Science* **366**, eaax0868 (2019).](https://www.zotero.org/google-docs/?rsUtIV)

[8. Patel, A. D. *Music, Language and the Brain*. (Oxford University Press, Oxford, 2008). doi:10.1093/acprof:oso/9780195123753.001.0001.](https://www.zotero.org/google-docs/?rsUtIV)

[9. Savage, P. E.](https://www.zotero.org/google-docs/?rsUtIV) *[Comparative Musicology: Evolution, Universals, and the Science of the World’s Music](https://www.zotero.org/google-docs/?rsUtIV)*[. (PsyArXiv preprint, 2025). doi:10.31234/osf.io/b36fm.](https://www.zotero.org/google-docs/?rsUtIV)

[10. Savage, P. E., Brown, S., Sakai, E. & Currie, T. E. Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8987–8992 (2015).](https://www.zotero.org/google-docs/?rsUtIV)

[11. Feld, S. & Fox, A. A. Music and language. *Annu. Rev. Anthropol.* **23**, 25–53 (1994).](https://www.zotero.org/google-docs/?rsUtIV)

[12. Ding, N. *et al.* Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* **81**, 181–187 (2017).](https://www.zotero.org/google-docs/?rsUtIV)

[13. Harris, I., Niven, E. C., Griffin, A. & Scott, S. K. Is song processing distinct and special in the auditory cortex? *Nat. Rev. Neurosci.* **24**, 711–722 (2023).](https://www.zotero.org/google-docs/?rsUtIV)

[14. Sammler, D. *Oxford Handbook of Language and Music [in Press]*. (Oxford University Press, 2025).](https://www.zotero.org/google-docs/?rsUtIV)

[15. Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D. & Majid, A. Over-reliance on English hinders cognitive science. *Trends Cogn. Sci.* **26**, 1153–1170 (2022).](https://www.zotero.org/google-docs/?rsUtIV)

[16. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–135 (2010).](https://www.zotero.org/google-docs/?rsUtIV)

[17. Pinker, S. *How the Mind Works*. (Norton, New York, 1997).](https://www.zotero.org/google-docs/?rsUtIV)

[18. Savage, P. E. *et al.* Music as a coevolved system for social bonding. *Behav. Brain Sci.* **44**, 1–22 (2021).](https://www.zotero.org/google-docs/?rsUtIV)

[19. Mehr, S. A., Krasnow, M. M., Bryant, G. A. & Hagen, E. H. Origins of music in credible signaling. *Behav. Brain Sci.* **44**, 23–39 (2021).](https://www.zotero.org/google-docs/?rsUtIV)

[20. Honing, H. Unravelling the origins of musicality: Beyond music as an epiphenomenon of language. *Behav. Brain Sci.* **44**, 66–68 (2021).](https://www.zotero.org/google-docs/?rsUtIV)

[21. Leivada, E. The origins of music in (musi)language. *Behav. Brain Sci.* **44**, 108–109 (2021).](https://www.zotero.org/google-docs/?rsUtIV)

[22. Benítez-Burraco, A. Evolutionary linguistics can help refine (and test) hypotheses about how music might have evolved. *Behav. Brain Sci.* **44**, 44–45 (2021).](https://www.zotero.org/google-docs/?rsUtIV)

[23. Pinker, S. Sex and drugs and rock and roll. *Behav. Brain Sci.* **44**, 115–116 (2021).](https://www.zotero.org/google-docs/?rsUtIV)

[24. Sammler, D. Signatures of speech and song: “Universal” links despite cultural diversity. *Sci. Adv.* **10**, eadp9620 (2024).](https://www.zotero.org/google-docs/?rsUtIV)

[25. Hilton, C. B. *et al.* Acoustic regularities in infant-directed speech and song across cultures. *Nat. Hum. Behav.* **8**, 1545–1556 (2022).](https://www.zotero.org/google-docs/?rsUtIV)

[26. Albouy, P. *et al.* Spectro-temporal acoustical markers differentiate speech from song across cultures. *Nat. Commun.* **15**, (2024).](https://www.zotero.org/google-docs/?rsUtIV)

[27. Anikin, A., Canessa-Pollard, V., Pisanski, K., Massenet, M. & Reby, D. Beyond speech: Exploring diversity in the human voice. *iScience* **26**, 108204 (2023).](https://www.zotero.org/google-docs/?rsUtIV)

[28. Steingo, G. & Ghazanfar, A. Virtual universals and creativity: A new approach to music cognition. *PsyArXiv Prepr.* (2024) doi:10.31219/osf.io/j7gtz.](https://www.zotero.org/google-docs/?rsUtIV)

[29. Shilton, D., Passmore, S. & Savage, P. E. Group singing is globally dominant and associated with social context. *R. Soc. Open Sci.* **10**, (2023).](https://www.zotero.org/google-docs/?rsUtIV)

[30. Dunbar, R. *Grooming, Gossip, and the Evolution of Language*. (Harvard University Press, Cambridge, MA, 1996).](https://www.zotero.org/google-docs/?rsUtIV)

[31. Urassa, M. *et al.* Cross-cultural research must prioritize equitable collaboration. *Nat. Hum. Behav.* **5**, 668–671 (2021).](https://www.zotero.org/google-docs/?rsUtIV)

[32. Coles, N. A., Hamlin, J. K., Sullivan, L. L., Parker, T. H. & Altschul, D. Build up big-team science. *Nature* **601**, 505–507 (2022).](https://www.zotero.org/google-docs/?rsUtIV)

[33. Forscher, P. S. *et al.* The benefits, barriers, and risks of big-team science. *Perspect. Psychol. Sci.* **18**, 607–623 (2023).](https://www.zotero.org/google-docs/?rsUtIV)

[34. Vaidis, D. C. *et al.* The advantage of big team science: Lessons learned from cognitive science. (2024) doi:10.31219/osf.io/yvm5h.](https://www.zotero.org/google-docs/?rsUtIV)

[35. Savage, P. E. *et al.* Building sustainable global collaborative networks: Recommendations from music studies and the social sciences. in *The science-music borderlands: Reckoning with the past, imagining the future* (eds. Margulis, E. H., Loughridge, D. & Loui, P.) 347–365 (MIT Press, 2023). doi:10.7551/mitpress/14186.003.0032.](https://www.zotero.org/google-docs/?rsUtIV)

[36. Chambers, C. The role of behavioural flexibility in promoting resilience to human environmental impacts [Recommendation of Stage 1 Programmatic Registered Report protocol by Logan et al.]. *Peer Community Regist. Rep.* (2022).](https://www.zotero.org/google-docs/?rsUtIV)

[37. Logan, C. J., Shaw, R., Lukas, D. & McCune, K. B. How to succeed in human modified environments. *Peer Community Regist. Rep.* (2022) doi:10.17605/OSF.IO/WBSN6.](https://www.zotero.org/google-docs/?rsUtIV)

[38. Lakens, D. Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Soc. Psychol. Personal. Sci.* **8**, 355–362 (2017).](https://www.zotero.org/google-docs/?rsUtIV)

[39. Guide for recommenders. *Peer Community In Registered Reports* https://rr.peercommunityin.org/help/guide\_for\_recommenders.](https://www.zotero.org/google-docs/?rsUtIV)

[40. Rennung, M. & Göritz, A. S. Prosocial consequences of interpersonal synchrony: A meta-analysis. *Z. Für Psychol.* **224**, 168–189 (2016).](https://www.zotero.org/google-docs/?rsUtIV)

[41. Mauch, M. & Dixon, S. PYIN: A fundamental frequency estimator using probabilistic threshold distributions. in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 659–663 (IEEE, Florence, Italy, 2014). doi:10.1109/ICASSP.2014.6853678.](https://www.zotero.org/google-docs/?rsUtIV)

[42. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research.](https://www.zotero.org/google-docs/?rsUtIV) *[J. Chiropr. Med.](https://www.zotero.org/google-docs/?rsUtIV)* **[15](https://www.zotero.org/google-docs/?rsUtIV)**[, 155–163 (2016).](https://www.zotero.org/google-docs/?rsUtIV)

[43. Cicchetti, D. V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.](https://www.zotero.org/google-docs/?rsUtIV) *[Psychol. Assess.](https://www.zotero.org/google-docs/?rsUtIV)* **[6](https://www.zotero.org/google-docs/?rsUtIV)**[, 284–2990 (1994).](https://www.zotero.org/google-docs/?rsUtIV)

[44. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data.](https://www.zotero.org/google-docs/?rsUtIV) *[Biometrics](https://www.zotero.org/google-docs/?rsUtIV)* **[33](https://www.zotero.org/google-docs/?rsUtIV)**[, 159–174 (1977).](https://www.zotero.org/google-docs/?rsUtIV)

[45. Tilsen, S. & Tiede, M. Parameters of unit-based measures of speech rate. *Speech Commun.* **150**, 73–97 (2023).](https://www.zotero.org/google-docs/?rsUtIV)

[46. Chiba, G., Ozaki, Y. & Savage, P. E. What is a ‘note’? Agreement and disagreement in transcriptions of Japanese folk songs [in press]. *Anal. Approaches World Music* (2025) doi:10.31234/osf.io/nh9d2.](https://www.zotero.org/google-docs/?rsUtIV)

[47. Nolan, F. & Jeon, H.-S. Speech rhythm: a metaphor? *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130396 (2014).](https://www.zotero.org/google-docs/?rsUtIV)

[48. Brunner, E., Bathke, A. C. & Konietschke, F. *Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs: Using R and SAS*. (Springer, 2018).](https://www.zotero.org/google-docs/?rsUtIV)

[49. Vargha, A. & Delaney, H. D. The Kruskal-Wallis test and stochastic homogeneity. *J. Educ. Behav. Stat.* **23**, 170–192 (1998).](https://www.zotero.org/google-docs/?rsUtIV)

[50. Ruscio, J. A probability-based measure of effect size: Robustness to base rates and other factors.](https://www.zotero.org/google-docs/?rsUtIV) *[Psychol. Methods](https://www.zotero.org/google-docs/?rsUtIV)* **[13](https://www.zotero.org/google-docs/?rsUtIV)**[, 19–30 (2008).](https://www.zotero.org/google-docs/?rsUtIV)

[51. Brockwell, S. E. & Gordon, I. R. A comparison of statistical methods for meta‐analysis.](https://www.zotero.org/google-docs/?rsUtIV) *[Stat. Med.](https://www.zotero.org/google-docs/?rsUtIV)* **[20](https://www.zotero.org/google-docs/?rsUtIV)**[, 825–840 (2001).](https://www.zotero.org/google-docs/?rsUtIV)

[52. Liu, S., Tian, L., Lee, S. & Xie, M. Exact inference on meta-analysis with generalized fixed-effects and random-effects models.](https://www.zotero.org/google-docs/?rsUtIV) *[Biostat. Epidemiol.](https://www.zotero.org/google-docs/?rsUtIV)* **[2](https://www.zotero.org/google-docs/?rsUtIV)**[, 1–22 (2018).](https://www.zotero.org/google-docs/?rsUtIV)

[53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing.](https://www.zotero.org/google-docs/?rsUtIV) *[J. R. Stat. Soc. Ser. B Stat. Methodol.](https://www.zotero.org/google-docs/?rsUtIV)* **[57](https://www.zotero.org/google-docs/?rsUtIV)**[, 289–300 (1995).](https://www.zotero.org/google-docs/?rsUtIV)

[54. Wang, Y. & Tian, L. An efficient numerical algorithm for exact inference in meta analysis.](https://www.zotero.org/google-docs/?rsUtIV) *[J. Stat. Comput. Simul.](https://www.zotero.org/google-docs/?rsUtIV)* **[88](https://www.zotero.org/google-docs/?rsUtIV)**[, 646–656 (2018).](https://www.zotero.org/google-docs/?rsUtIV)

[55. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials.](https://www.zotero.org/google-docs/?rsUtIV) *[Control. Clin. Trials](https://www.zotero.org/google-docs/?rsUtIV)* **[7](https://www.zotero.org/google-docs/?rsUtIV)**[, 177–188 (1986).](https://www.zotero.org/google-docs/?rsUtIV)

[56. Romano, J. P. Optimal testing of equivalence hypotheses.](https://www.zotero.org/google-docs/?rsUtIV) *[Ann. Stat.](https://www.zotero.org/google-docs/?rsUtIV)* **[33](https://www.zotero.org/google-docs/?rsUtIV)**[, 1036–1047 (2005).](https://www.zotero.org/google-docs/?rsUtIV)

[57. Chambers, C. Exploring cross-cultural variation in speech and song [based on reviews by Bob Slevc and Nai Ding].](https://www.zotero.org/google-docs/?rsUtIV) *[Peer Community Regist. Rep.](https://www.zotero.org/google-docs/?rsUtIV)* [(2023).](https://www.zotero.org/google-docs/?rsUtIV)

[58. Moore, K. Do humans bond more when singing together or speaking together? A global investigation  [based on reviews by Melissa Brandon, Erin Hannon, Manuela Maria Marin and Anja Göritz].](https://www.zotero.org/google-docs/?rsUtIV) *[Peer Community Regist. Rep.](https://www.zotero.org/google-docs/?rsUtIV)* [(2025).](https://www.zotero.org/google-docs/?rsUtIV)

[59. Boersma, P. & Weenink, D. Praat: Doing phonetics by computer [Computer program]. (2025).](https://www.zotero.org/google-docs/?rsUtIV)

**APPENDIX S1: Video tutorials**

-The video used by Ozaki et al. (2024) showing how to align onsets based on perceptual centers (“P-centers”) is available here: <https://drive.google.com/file/d/1YOiobvoxaM4txdAJDVeLjc--oNLiBb5n/view>

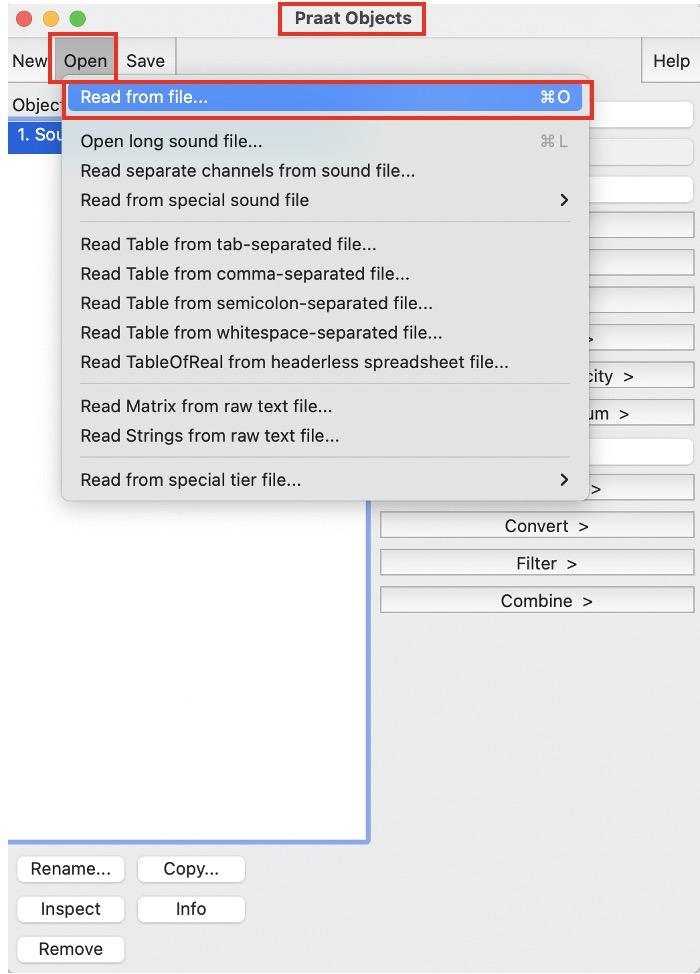
-A video tutorial showing how to use the free software Praat[59](https://www.zotero.org/google-docs/?1b0hIs) to segment acoustic units (e.g., syllables/notes) from a pilot experiment recording containing multiple participants singing/conversing is available here: <https://drive.google.com/file/d/1Nz4h-JSk1d3Z_NNiXN1UEpv3TVTBefdx/view?usp=sharing>.

-A text version of the Praat tutorial follows:

**A Tutorial for Doing Annotation in Praat**

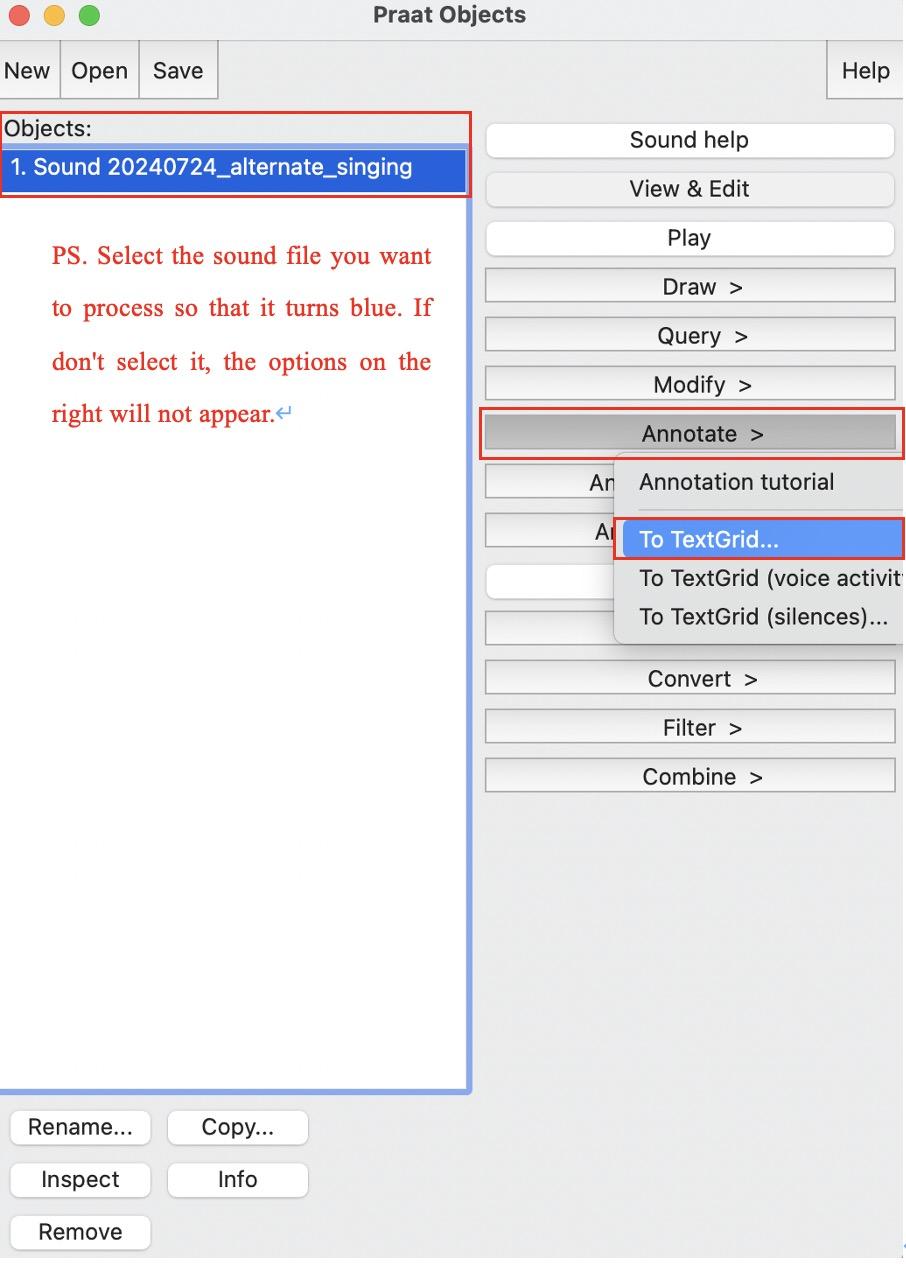
1. Download and install Praat if you haven’t already. <https://www.fon.hum.uva.nl/praat/>
2. **Prepare your audio recordings in a folder**. Make sure your audio files are in **.wav** format.
3. **Open the audio recording you want to process**:

In Praat, go to ***Praat Object*** > ***Open*** > ***Read from file…*** , then select the file you want to work with.



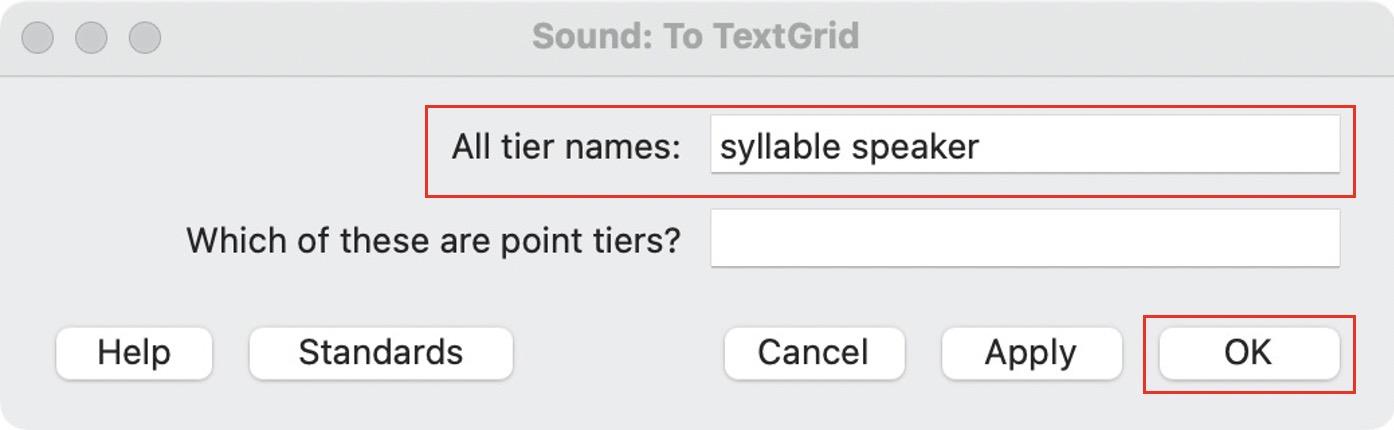
1. **Create a TextGrid for the corresponding audio file to make annotations**:

In the **Objects** window, select the audio file, then go to ***Annotate > To TextGrid****…*

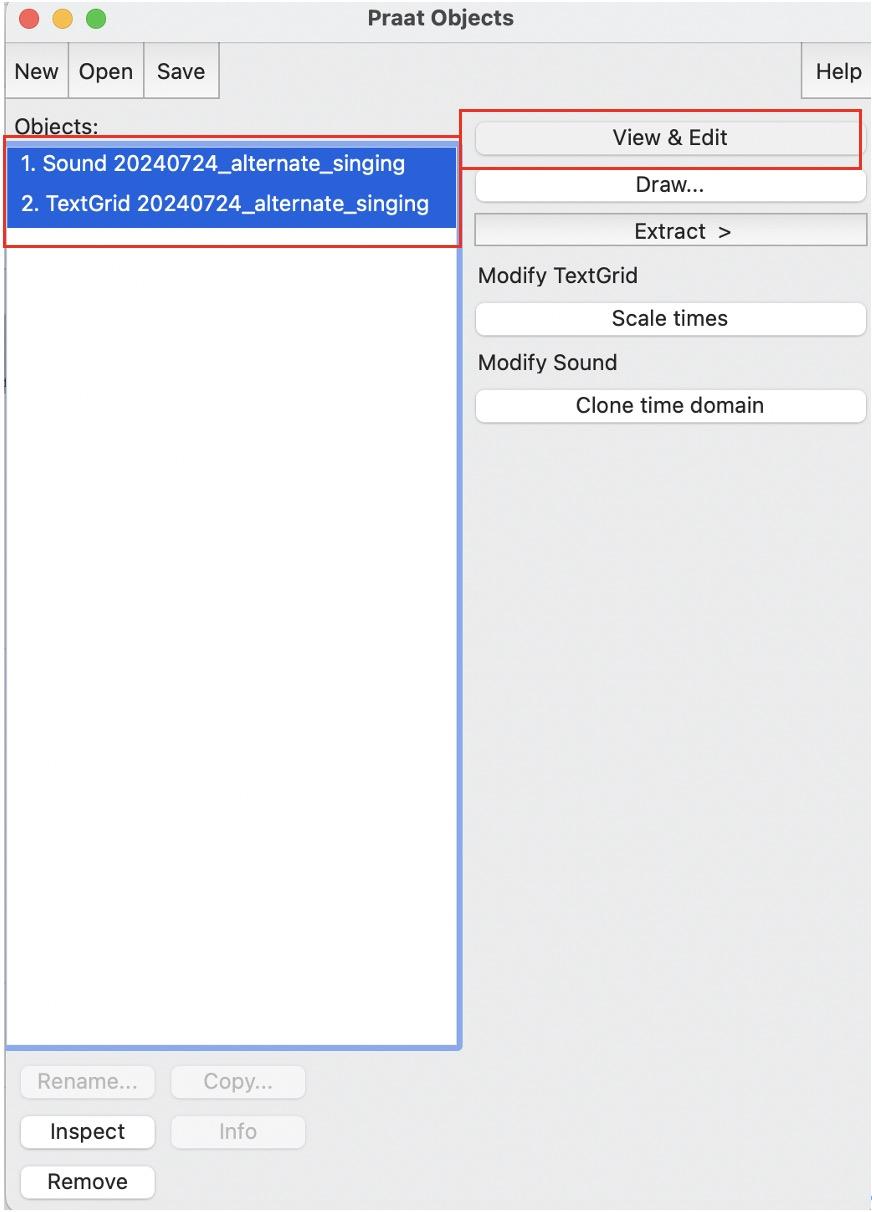


1. Then a small window called **Sound: To TextGrid** will appear. Clear all the default settings.In the **All tier names:** field, enter **syllable speaker** (if it is a song file, then enter **note speaker**), then click **OK**.

(A detailed explanation: We need to annotate two tiers: the first tier for syllables/notes, and the second tier for speakers. The logic for **All tiers names** is as follows: from left to right, list the first tier name, the second tier name, and then the third tier name…. separated by spaces. You are free to choose different tier names if you prefer; the tier names themselves do not affect data processing. They mainly serve as a guide to remind annotators what type of information should be labeled in each tier. Since we do not need to annotate any points, **leave the point tiers empty**.）

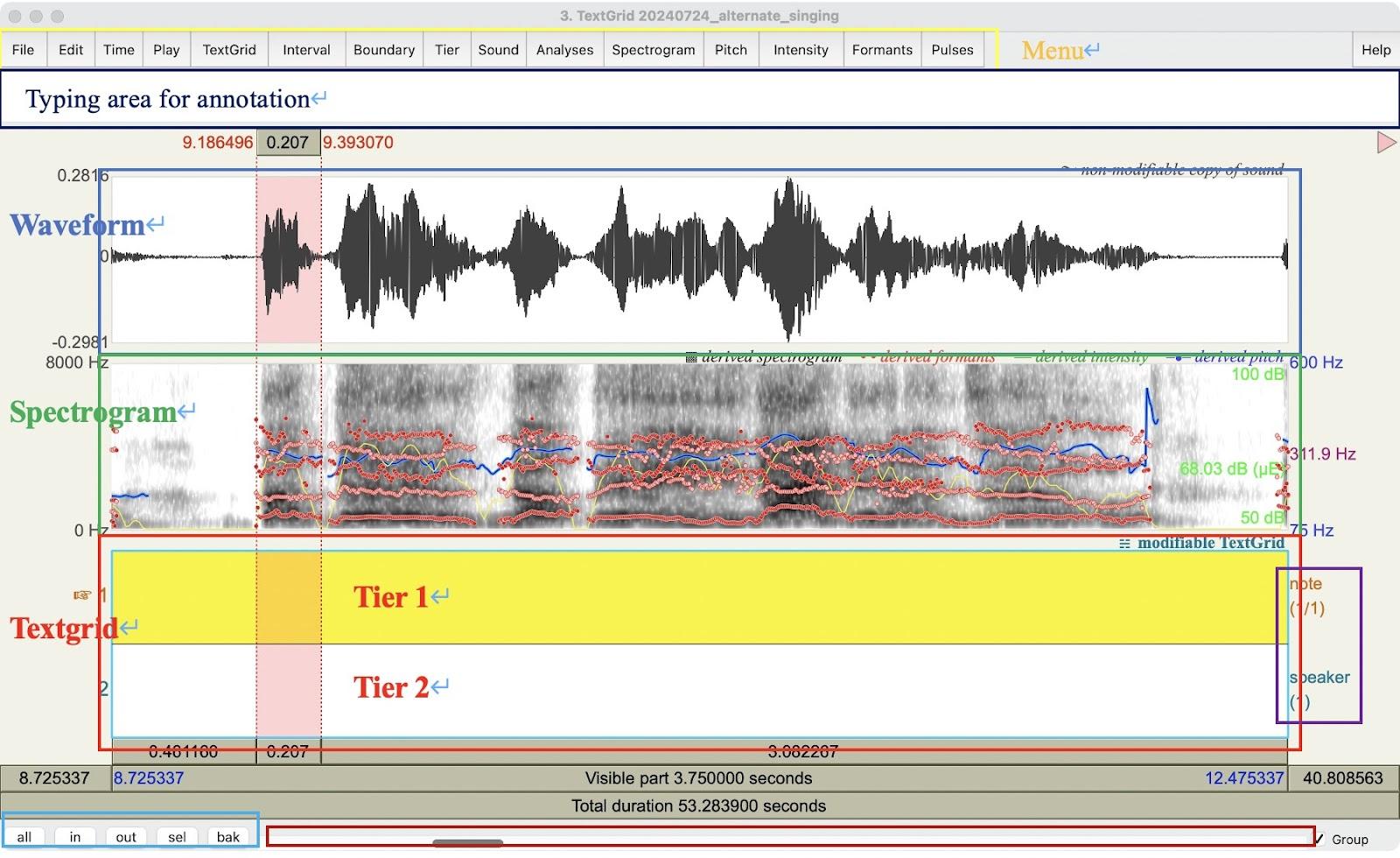


1. Now you have got a Sound file and a corresponding TextGrid file. **Select both** files, and click ***View & Edit***.



1. Then the visualisation of combined sound and textgrid files appear. From top to bottom, you will see: the menu bar, typing area, waveform, spectrogram, TextGrid (tier 1 + tier 2), and time information.

(In the **blue box** at the bottom left corner below, **"all"** means to display the entire audio file within the window, **"in"** means to zoom in, **"out"** means to zoom out, and **"sel"** means to display only the selected regions. You can use these buttons together with the scrollbar on the right to adjust the view and select the appropriate region for annotation.)



1. **Add a boundary**: If you want to add a boundary, click on **the waveform or spectrogram area** at the position where you want the boundary to be. *(****Do not*** *click on the TextGrid area; clicking there is for adding or editing text labels.)*

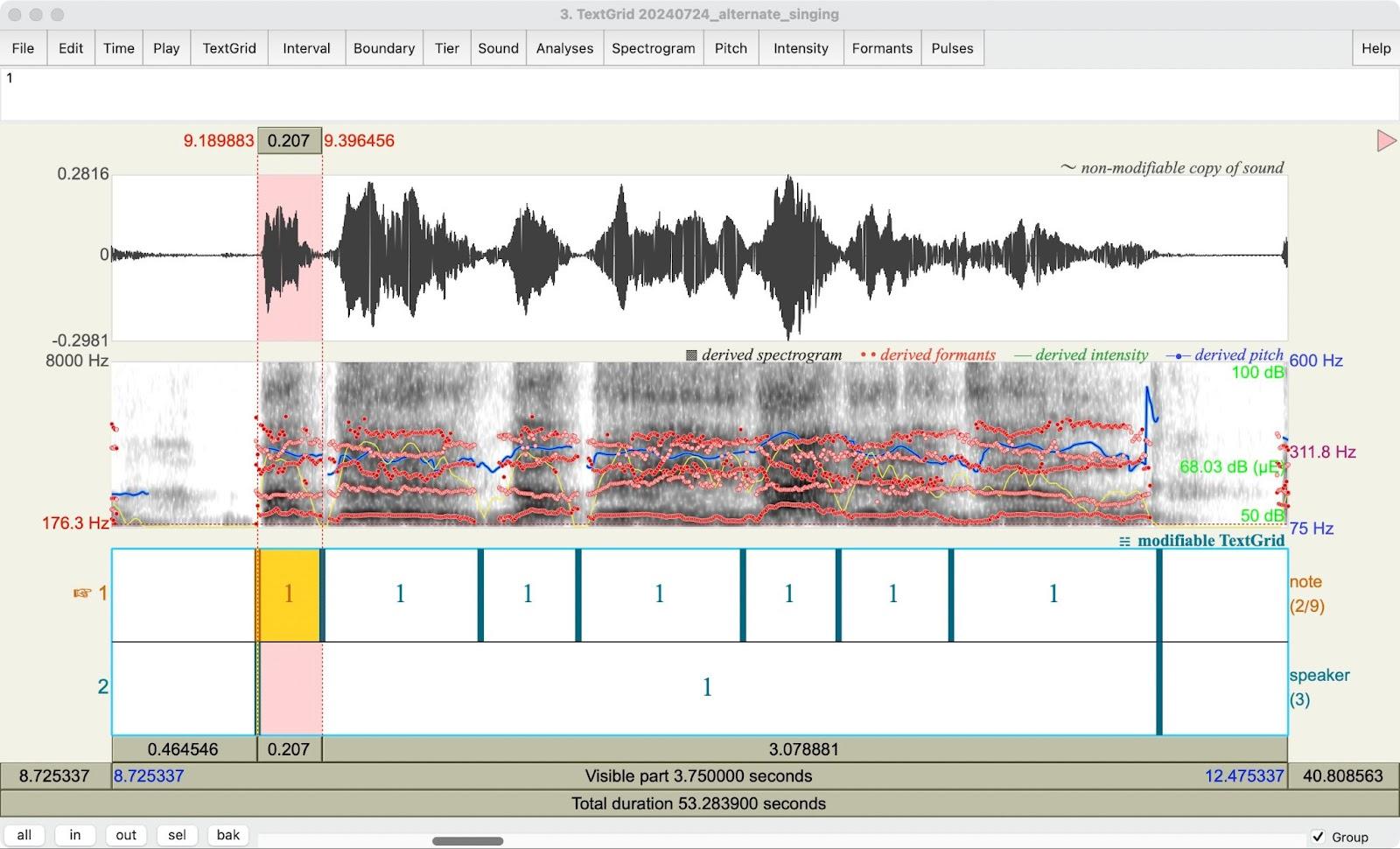
After clicking, a thick grey line will appear. Click the **small circle ⭕️ on top of the grey line** *(clicking on the grey line itself won't work)*.

When the grey line turns blue, it means the boundary has been successfully added.

1. **Add annotation text**:

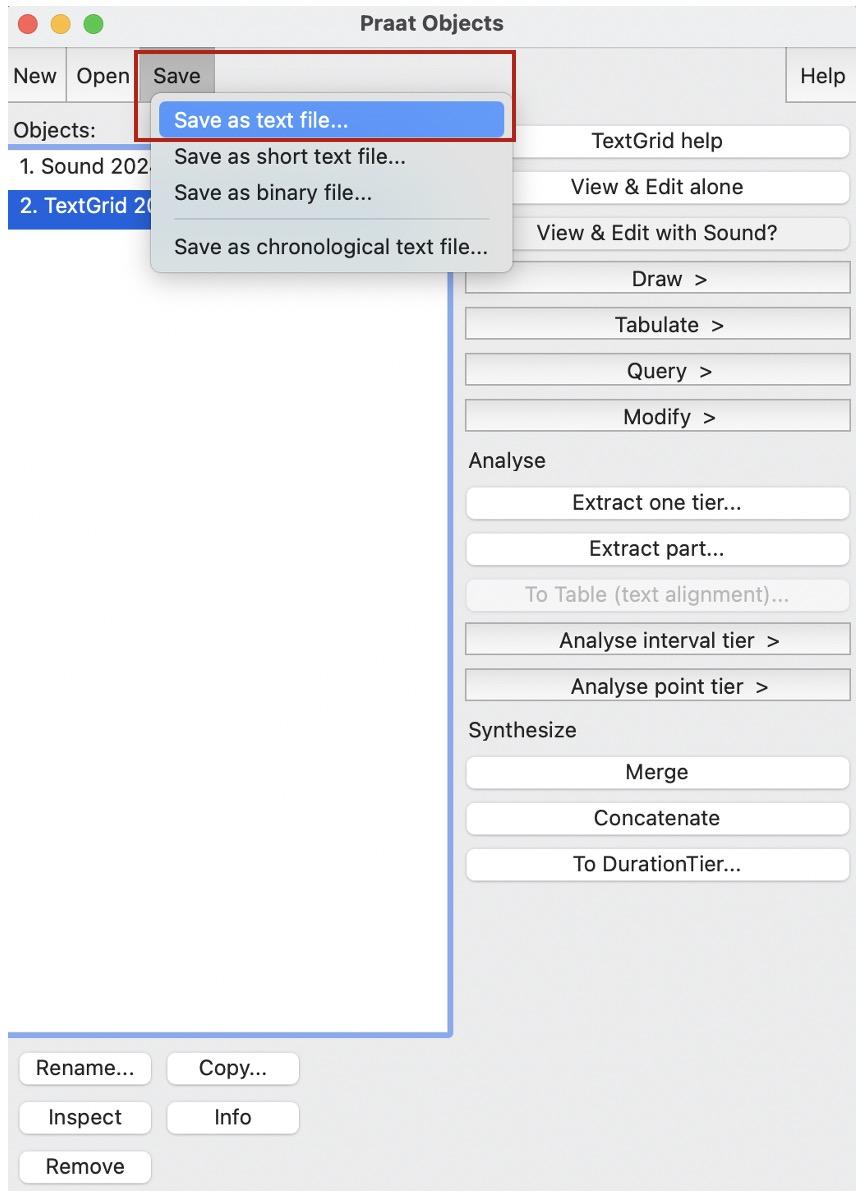
In the TextGrid area, select the region between two boundaries, and then type the text you want to add in the typing area.

(Since we ultimately need to know the duration of individual syllables or notes for different speakers, and also extract the complete speech segments for each speaker, it's important to annotate speaker information. We recommend labeling based on Participant ID, e.g., 1, 2, 3,..., up to 10.)



1. **Save the TextGrid file**:

After completing all annotations, go back to the Praat Objects window, select the TextGrid file, and choose ***Save > Save as text file...****.*



1. School of Psychology, University of Auckland, Auckland, New Zealand [↑](#endnote-ref-1)
2. Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa, Japan [↑](#endnote-ref-2)
3. School of Psychology, University of Auckland, Auckland, New Zealand [↑](#endnote-ref-3)
4. Keio Research Institute at SFC, Keio University, Fujisawa, Kanagawa, Japan [↑](#endnote-ref-4)
5. School of Psychology, University of Auckland, Auckland, New Zealand [↑](#endnote-ref-5)
6. School of Psychology, University of Auckland, Auckland, New Zealand [↑](#endnote-ref-6)
7. Department of Psychology, University at Buffalo, State University of New York, Buffalo, NY, USA [↑](#endnote-ref-7)
8. Department of Folk Music Research and Ethnomusicology, University of Music and Performing Arts–MDW, Wien, Austria [↑](#endnote-ref-8)
9. Neuroscience of Perception & Action Lab, Italian Institute of Technology (IIT), Rome, Italy [↑](#endnote-ref-9)
10. Centre of Excellence in Music, Mind, Body and Brain, Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland [↑](#endnote-ref-10)
11. Social Body Lab, Centre for the Study of Social Cohesion, University of Oxford, Oxford, UK [↑](#endnote-ref-11)
12. Heitor Villa-Lobos Municipal Conservatory/ BSB Musik, Brazil [↑](#endnote-ref-12)
13. Faculty of Social Sciences, Die Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU), Kaiserslautern, Germany [↑](#endnote-ref-13)
14. Department of Psychology, University of Toronto Mississauga, Mississauga, ON, Canada [↑](#endnote-ref-14)
15. Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa, Japan [↑](#endnote-ref-15)
16. Faculty of Social Sciences, Die Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU), Kaiserslautern, Germany [↑](#endnote-ref-16)
17. Faculty of Psychology, Babeș-Bolyai University, Cluj-Napoca, Romania [↑](#endnote-ref-17)
18. School of Psychology, University of Auckland, Auckland, New Zealand [↑](#endnote-ref-18)
19. Department of Psychology, University at Buffalo, State University of New York, Buffalo, NY, USA [↑](#endnote-ref-19)
20. Department of General Psychology, University of Padua, Padua, Italy [↑](#endnote-ref-20)
21. Institute of Formal and Applied Linguistics, Charles University, Czech Republic [↑](#endnote-ref-21)
22. Psychology Department, Center for Cognitive and Brain Health, Northeastern University, Boston, MA, USA [↑](#endnote-ref-22)
23. Centre of Excellence in Music, Mind, Body, & Brain, University of Jyväskylä, Jyväskylä, Finland [↑](#endnote-ref-23)
24. Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark [↑](#endnote-ref-24)
25. Interacting Minds Centre, Aarhus University, Aarhus, Denmark [↑](#endnote-ref-25)
26. Centre of Excellence in Music, Mind, Body and Brain, Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland [↑](#endnote-ref-26)
27. School of Psychology & Clinical Language Sciences, University of Reading, Reading, UK [↑](#endnote-ref-27)
28. Centre for Culture and Evolution, Department of Psychology, Brunel University, London, UK [↑](#endnote-ref-28)
29. Musicology Institute, Faculty of Arts Studies, University of Adam Mickiewicz, Poland [↑](#endnote-ref-29)
30. Institute of Language Studies, Estate University of Campinas, Brazil [↑](#endnote-ref-30)
31. South-West University Neofit Rilski, Blagoevgrad, Bulgaria [↑](#endnote-ref-31)
32. Phoenix Perpeticum Foundation, Sofia, Bulgaria [↑](#endnote-ref-32)
33. School of Psychology & Clinical Language Sciences, University of Reading, Reading, UK [↑](#endnote-ref-33)
34. Department of Music, Center for Cognitive and Brain Health, Northeastern University, Boston, MA, USA [↑](#endnote-ref-34)
35. Academy of Music, Dance and Fine Arts, Plovdiv, Bulgaria [↑](#endnote-ref-35)
36. Pontificia Universidad Católica de Chile, Villarrica, Chile [↑](#endnote-ref-36)
37. Independent Researcher, Brazil [↑](#endnote-ref-37)
38. Centre of Excellence in Music, Mind, Body and Brain, Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland [↑](#endnote-ref-38)
39. School of Arts, Music, University of KwaZulu-Natal, Durban, South Africa [↑](#endnote-ref-39)
40. Ethnological Institute of the Czech Academy of Sciences, Czech Republic [↑](#endnote-ref-40)
41. Department of Creative Arts, University of Lagos, Lagos, Nigeria [↑](#endnote-ref-41)
42. School of Arts, Music, University of KwaZulu-Natal, Durban, South Africa [↑](#endnote-ref-42)
43. School of Psychology, University of Auckland, Auckland, New Zealand [↑](#endnote-ref-43)
44. Memorial University of Newfoundland, St. John’s, NL, Canada [↑](#endnote-ref-44)
45. Department of Music and Dance, Kenyatta University, Nairobi, Kenya [↑](#endnote-ref-45)
46. Faculty of Society & Design, Bond University, Gold Coast, Australia [↑](#endnote-ref-46)
47. Department of Psychology, University at Buffalo, State University of New York, Buffalo, NY, USA [↑](#endnote-ref-47)
48. Musicology Institute, Faculty of Arts Studies, University of Adam Mickiewicz, Poland [↑](#endnote-ref-48)
49. Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, Vienna, Austria [↑](#endnote-ref-49)
50. Department of Philosophy, Macquarie University, Sydney, Australia [↑](#endnote-ref-50)
51. School of Psychology, University of Auckland, Auckland, New Zealand [↑](#endnote-ref-51)
52. Departamento de Humanidades y Artes, Universidad de Los Lagos, Osorno, Chile [↑](#endnote-ref-52)
53. Millennium Nucleus in Musical and Sound Cultures (NCS2022\_016), Santiago, Chile [↑](#endnote-ref-53)
54. Departamento de Humanidades y Artes, Universidad de Los Lagos, Osorno, Chile [↑](#endnote-ref-54)
55. Millennium Nucleus in Musical and Sound Cultures (NCS2022\_016), Santiago, Chile [↑](#endnote-ref-55)
56. Institute of Formal and Applied Linguistics, Charles University, Czech Republic [↑](#endnote-ref-56)
57. Faculty of Society & Design, Bond University, Gold Coast, Australia [↑](#endnote-ref-57)
58. Department of Psychology, Babeș-Bolyai University, Cluj-Napoca, Romania [↑](#endnote-ref-58)
59. Department of Psychology, University of Toronto Mississauga, Mississauga, ON, Canada [↑](#endnote-ref-59)