Quasi-rhythmic features of hand gestures show unique modulations within languages: Evidence from bilingual speakers

Wim Pouw
Radboud University

Asli Özyürek
Max Planck Institute for Psycholinguistics and Radboud University

Quasi-rhythmic features of hand gestures show unique modulations within languages: Evidence from bilingual speakers

Wim Pouw[1, 2, 3]*, Jessica J. Jaramillo*[4]*, Asli Ozyurek*[5,2,3]*, James A. Dixon*[1,4]

1. Center for the Ecological Study of Perception and Action, University of Connecticut
2. Donders Institute for Brain, Cognition, and Behavior, Radboud University
3. Max Planck Institute for Psycholinguistics Nijmegen
4. Department of Psychological Sciences, University of Connecticut
5. Centre for Language Studies, Radboud University

w.pouw@psych.ru.nl, jessica.j.jaramillo@uconn.edu, asli.ozyurek@mpi.nl, james.dixon@uconn.edu

Abstract

Hand gestures are known to beat (i.e., oscillate) with the quasi-rhythmic aspects of speech. It is unknown however whether speakers modulate their oscillatory properties of gesture in unique ways depending on which language they are speaking. Leveraging wavelet analysis and motion tracking we find that gestures produced by bilingual speakers are more similar in their quasi-rhythmic properties when produced under the same rather than a different language. This suggests that gestures' oscillatory aspects are likely to be in part defined by the language spoken. More cross-language research is needed to understand whether oscillatory properties of gesture are indeed uniquely specific to the spoken language and how such oscillatory properties might relate to properties in speech.

Index Terms: Multimodal Language, Motion Tracking, Wavelet Analysis, Bilingualism

1. Introduction

Spoken language can be characterized as operating on nested levels of periodicity, such as faster scale vocal folds oscillations, meso-scale mouth opening-closing oscillations, and slower scale respiration cycles[1–3]. The meso-scale is an important time scale as it constitutes the syllable cycle, which is defined by a 2-8Hz range across languages, with particular modulations around the 4-5 Hz range[1]. The amplitude envelope is the acoustic marker of such meso-scale syllable cycles, closely correlating with labial kinematics[1, 4, 5]. Although speech is not strictly rhythmic in the sense of being isochronous, it has intrinsic (e.g., biomechanical) dynamics that drives the system to attract to (or "prefer") oscillations at 4-5 Hz range. In this sense speech can be said to be quasi-rhythmic[1]. It precisely this quasi-rhythmicity that neural ensembles in the brain seem to resonate to during speech processing[1, 6]. Interestingly, sign-language is also modulated at the 2-8 Hz range with slightly more dominant periodicities at the slower timescales (2-2.5Hz) as compared to speech, and signers similarly neurally resonate to 2-8Hz frequencies during sign-language processing[7]. Thus, it seems that the 2-8Hz range is an important time scale for carrying language-dedicated information regardless of modality.

Not much is known however about the dominant timescales that define the hand-movements which spontaneously co-occur with spoken language (referred to as co-speech gestures). This is surprising as co-speech gestures are considered by experts to be temporally coordinated with the quasi-rhythmic and intonational aspects of speech[8]. In fact, no matter what hand-movements depict, the acceleration profiles of gestures are often well timed with moments of emphasis in speech[9–13]. It has been argued that the explanation for this tight timing relation may have its origin in biomechanic linkages of upper limb movements with the respiratory system. For example, producing steady-state vocalizations or mono-syllable utterances during mechanical loading of movements onto the body increases intonation (F0) and the amplitude envelope of such utterances[14–16]. That there is a tight link between hand gestures and speech is further corroborated by recent advances in machine learning, where it has been shown that neural networks trained on associations of speech acoustics and hand gesture kinematics can produce very convincing synthetic gestures based on novel acoustic input[17, 18]. Finally, we know that gesture movement rate and speech rate covary reliably with acoustics and production rate in speech to a point that they can be characterized as one expressive system[19]. While such research supports the conclusion that gesture and speech are temporally coordinated, it is not clear whether quasi-rhythmicity in gesture is a relevant informational property similar to sign language and speech.

Here we probe whether quasi-rhythmic aspects of gesture might be an important feature of gestures. In this exploratory study, we assessed whether English-Spanish bilinguals who were asked to retell cartoon scenes modulated their gestures differently at the 2-8Hz range when speaking English versus Spanish. We know already that co-speech gestures often follow the systematic rules of the spoken language (i.e., syntax). A gesture iconically expressing a SUBJECT or a VERB is often produced in an order depending on the order that a particular spoken language favors[20, 21]. It is further known that speakers' gesture rate can vary depending on the language spoken (e.g., Turkish incorporate more gestures than the Dutch)[22]. Even *bilinguals* gesture more when speaking Turkish versus Dutch, suggesting gesture rate is constrained by the language spoken rather than by which language(s) you speak. But it is an open question whether gestures have time-varying (dynamic) features that are modulated by the language you speak, which is entirely orthogonal to gesture rate. Such unique modulations may possibly be supportive of speaking Spanish versus English as they may resonate to subtly different quasi-rhythmic features. If we find that gestures periodicities at the 2-8Hz range are in part defined by the language spoken, this shows that movement variability of gestures are defined by different time scales. If this is the case, further research is needed whether these quasi-rhythmic modulations are resonating to some feature in speech, or such differences arise due to other constraints associated with that spoken language (e.g., the context that the language is usually spoken in might affect gesturing style when speaking that language).

The current exploratory study serves two goals. The primary goal is to showcase a particular analysis pipeline to studying quasi-rhythmic features of gestures by leveraging wavelet analysis of motion tracking data. Such analysis can (and should) first be performed within participants, thereby accommodating for differences in gesturing styles in search for person-unique modulations of spoken language. For example, as mentioned it has been found that neural networks are able to synthesize gestures based on novel speech acoustics from the person the model is trained with. Performance drops, however, when such models are tasked to synthesize gestures based on novel speech from a different person [17]. Thus, it is entirely possible that modulations of quasi-rhythmic features of gesture with spoken language is person-specific. Yet there might be subtle differences between languages that constrain gestures across speakers, and we will show how to test for generalizable quasi-rhythmic gesture patterns depending on the language spoken. The second goal is to report on promising findings from a convenience sample of bilingual speakers of Spanish and US-American English, where we show that whether someone speaks Spanish or English show concomitant changes in the time-scales that define the gestural movement, both within and across speakers. While promising, since we have no speech and only motion tracking data for this sample the onus for future research lies in explicitly relating possible differences of quasi-rhythmic features of gestural movement with said features in speech.

2. Methods and Materials

2.1. Participants and design

We tested 15 self-reported English-Spanish bilingual participants (14 right-handed; *M* Age = 19.13 years, *SD* Age = 1.12; 2 Males) for 30 minutes. This sample produced 1332 gestures that were amenable for within-subject analysis (469 gestures under English vs. 863 under Spanish). Three participants who did not gesture were excluded from the analyses.

2.2. Language skills

All participants reported speaking Latin American Spanish (Mexico *N* = 5, Ecuador *N* = 4, Puerto Rico *N* = 2, Chili *N* = 1, Peru *N* = 1, Guatemala *N* = 1, Salvador *N* = 1). 53% of participants reported learning Spanish only at home where Spanish was the primary language. 46% reported that they learned additional Spanish in Kindergarten, or also Middle School, or also High School, next to learning Spanish at home. Table a. provides the key information about the language skills of the participants. It can be seen that language skills are not markedly different for Spanish or for English, although there is a small advantage for English (as to be expected for US citizens).

Table 1. Key self-reported language skills

0 = low proficiency 5 = high proficiency		English *M (SD)*	Spanish *M (SD)*
Self-reported Proficiency	Writing	3.66 *(0.61)*	3.00 *(0.75)*
	Speaking	3.6 *(0.63)*	3.06 *(0.70)*
	Listening	3.7 *(0.62)*	3.7 (0.49)
Age of Fluency (years)		7.8 (4.99)	6.2 (4.98)
How often do you switch to the other language? [never] 1-5 [often]		2.21 (0.97)	2.47 (0.74)

2.3. Materials

To elicit gesture and speech we used cartoons that are often used in gesture research: the Tweety and Sylvester cartoons "Canary Row" and "Snow Business". These cartoons were segmented in twelve vignettes of comparable length (*M* duration scenes = 59.42seconds; *SD* = 32.11). These cartoon vignettes were segmented so that participants could easily memorize and retell the content of the cartoon in a piecemeal fashion.

2.4. Procedure

Upon entering the lab participants first completed a survey about their language competence (as reported in table 1). Participants then entered into the main part of the experiment where they were seated and two motion tracking sensors were attached to the left and right index finger. Participants were asked to retell 12 vignettes consecutively, by first watching a vignette and then immediately retelling what they saw. After each vignette watching and retelling, participants were asked to prepare to switch language (English or Spanish) for the retelling of the next cartoon, so that participants retold vignettes in English and Spanish in alternating fashion. Language switching order was counterbalanced, so that the 12 scenes were assigned in a balanced way to Spanish or English retellings. Participants were asked to retell the cartoon as accurately as possible without the experimenter mentioning our particular interest in gestures (although the motion tracking sensors may have implied such interest). If participants took longer than 30 minutes, the experiment was ended after the finishing of the latest scene. Speech could unfortunately not be recorded due to IRB constraints, but the experimenter was present at all times during data

2.5. Motion tracking

We used a Polhemus Liberty (Vermont, USA) sampling at 240Hz at a resolution of 0.0012mm (at ideal conditions) with two wired sensors. Each sensor was attached to the nail of the left and right index finger. All position traces and time derivatives were smoothed with a low-pass 33Hz first order Butterworth filter.

2.6. Signal-based gesture coding

We automatically coded co-speech hand movements based on the motion-tracking data. When a movement of either hand exceeded a 15 cm/s 3D speed threshold this was coded as a gesture activity. This threshold is comparable to other work on gesture initiation detection (e.g., [23]). We considered the gesture as terminated when the speed went under the 15cm/s threshold. This ensured a fully objective coding of a co-speech movement event.

2.7. Data Analysis

For this study we want to assess whether within participants there were unique quasi-rhythmic features when speaking Spanish versus English. If this is the case, we can further probe whether gestures spoken under Spanish or English have across-participant modulations at particular frequency ranges.

2.7.1 Oscillatory properties of gesture

Figure 1 shows the data processing steps of the current procedure. We first determined for each gesture produced (only including gestures lasting longer than 500ms, i.e., > 2Hz) how its speed time series was defined over 2Hz (500ms; slowest oscillation) to 8Hz (200ms period; fastest oscillation) range. A fast alternating gesture would have higher power at faster time scale (e.g., > 4Hz), while slower gestures would have higher power at slower frequencies (e.g., < 4 Hz). We used wavelet analysis (using `R` package `Waveletcomp` [24], with a Morlet as the daughter wavelet, to estimate the smoothed power spectrum over the 2-8Hz frequency domain. The power estimates were rescaled from 0 = lowest power to 1 = highest power so as to enable later comparisons between gestures.

2.7.2 Spectral comparison between gestures

After extracting the power spectrum over the 2-8Hz range based on the speed time series for each gesture event, we then computed a distance measure between all gesture events produced by a single participant. Specifically, we assessed for each gesture the difference in the 2-8Hz power spectrum with all other gestures produced by this participant. This distance measure was calculated using equation 1:

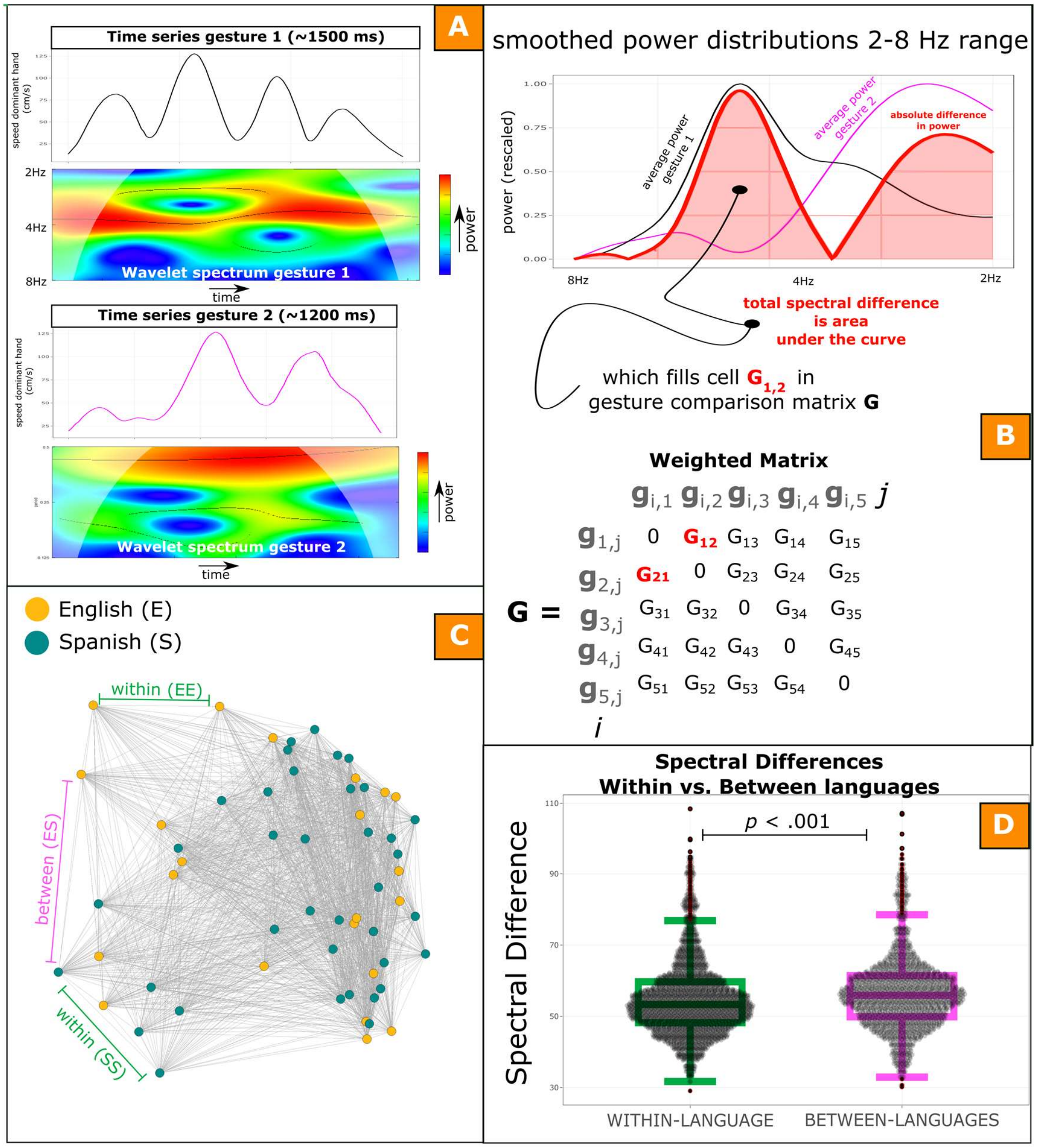
$$\text{Distance } \mathrm{G}_{\mathrm{i,j}} = \sum_{f=2\,Hz}^{f=8\,Hz} |\text{Power}_{f,i} - \text{Power}_{f,j}| \tag{1}$$

Equation 1 denotes that a spectral comparison was made between gesture *i* with another gesture *j*, by adding the absolute differences in power for gesture *i* versus gesture *j* at frequency *f* (with steps of 0.01Hz bins). This yields a low score when gestures had similar quasi-rhythmic features, while a higher score when gestures were differently defined over the 2-8Hz range.

For each participant, we saved distance computations in a matrix **G** with number of rows/columns equal to the number of gestures produced, thus containing all gesture comparisons $\mathrm{G}_{\mathrm{i,j}}$ for this participant. Such matrices can be represented as fully connected networks with each gesture as a node and the length of its edges indicating the “spectral” distance with other gestures (see Figure 1). Importantly we extracted from these matrices the distances between gestures that were produced under the same language (Spanish or English) so as to compare this to the spectral differences of gestures that were produced under the different language. This allowed us to assess whether within-language gestures have lower spectral distances than between language gestures.

2.8. Data availability

Anonymous data, analysis code this paper can be found on https://osf.io/2epwd/.



$$G = \begin{array}{c|ccccc} & g_{i,1} & g_{i,2} & g_{i,3} & g_{i,4} & g_{i,5} \\ \hline g_{1,j} & 0 & G_{12} & G_{13} & G_{14} & G_{15} \\ g_{2,j} & G_{21} & 0 & G_{23} & G_{24} & G_{25} \\ g_{3,j} & G_{31} & G_{32} & 0 & G_{34} & G_{35} \\ g_{4,j} & G_{41} & G_{42} & G_{43} & 0 & G_{45} \\ g_{5,j} & G_{51} & G_{52} & G_{53} & G_{54} & 0 \end{array}$$

Figure 1: *Procedure analysis and main results*

The key methodological steps and main findings are provided here. **A)** For each gesture produced by a participant, we performed wavelet analysis on the 3D speed time series of the dominant hand. In this example, gesture 1 has faster oscillations at around 4Hz as compared to gesture 2 which has slower oscillations at around 2.5 Hz. **B)** We then compared for each gesture pair the average power spectrum for the 2-8Hz range (power rescaled from 0 to 1). The difference in the power distribution between a gesture comparison pair was summed to get a "spectral difference" measure (see formula 1 give above). For each gesture pair (i,j). we computed this spectral difference measure, filling a matrix **G** containing all gesture distance comparisons (G_{ij} = Distance(g_i, g_j)). **C)** This matrix can be represented as a network of gesture relations[26], wherein each node represents a gesture, and edges are the spectral difference between another gesture wherein the higher the distance the more spectrally dissimilar the gestures. This network is from participant 4 in our sample (network made with `Igraph`[27]); for all networks see https://osf.io/pgdjn/ (or Figure B in supporting information). We can then make comparisons between the distances (i.e., spectral differences) between gestures that are produced under the same language (English or Spanish) or between languages (English vs. Spanish). **D)** We find that gestures spectral properties are more similar when produced under the same language (WITHIN-LANGUAGE) as compared to higher distances when gesture are compared between languages (BETWEEN- LANGUAGE). Note the swarm comprises each spectral difference made for each participant's gesture (within subjects) and is shaped according to its distribution (made with ggbeeswarm[28]).

3. Results

3.1. Participant level analysis

Mixed `nlme`[25] regression analysis (participant as random intercept) was performed to compare the spectral differences between gestures produced under the same language (English or Spanish gestures) versus the spectral differences between gestures produced under different language (English vs. Spanish gestures). We find that within subject, gestures produced under a different language had higher spectral differences, $b = 1.68$, 95% CI[1.06, 2.29], $t\ (4377) = 5.33$, $p < .001$, as compared to gestures spectral differences within languages (see Figure 1, Panel D). There were no differences between spectral comparisons within English versus within Spanish gestures, $b = 0.063$, 95% CI[-0.76, 0.89], $t\ (2651) = 0.15$, $p = .881$, suggesting that spectral differences between gestures within a language were similarly distributed for both languages.

Since we found that gestures are more homogenous in their oscillatory properties within language spoken as compared to between language, the question is whether there are homogenous oscillatory properties across speakers. Fitted power levels over the frequency range (2-8 Hz) are shown below in Figure 2. It can be seen that there are subtle overall differences, whereby there is modulation at higher frequency range for gestures produced under English as compared to Spanish gestures.

3.2. Language-level analysis

To assess whether any differences in the power-frequency trajectories were statistically reliable at the language level, we performed generalized additive modeling (GAM; see Figure 2). GAM is a type of multilevel regression which can model non-linear trajectories utilizing a set of base smooth functions. With GAM we can perform within-subject analyses to see whether the non-linear power spectrum trajectories at the language level (Spanish versus English) are reliably different from each other, which would suggest that Spanish and English gestures have unique quasi-rhythmic properties. We performed GAM with `R`-package `gam` with participant as random intercept. We find that the slopes for the Spanish versus English were statistically different, $b = -0.010$, $SE = 0.002$, $p < .001$, suggesting that indeed the Spanish versus English gestures are to some extent differentiable on the 2-8Hz range. Note that the non-linear smooth terms were statistically reliable, p's < .001, indicating that the slopes were indeed non-linear with 29% explained variance for non-linear smooths.

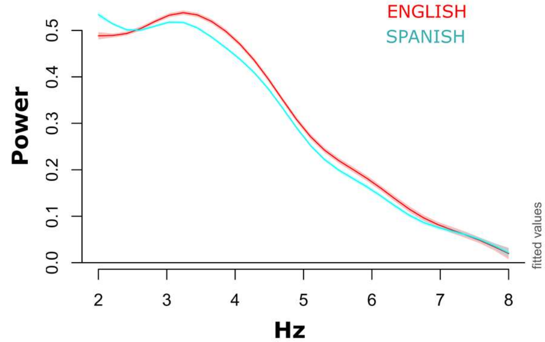


Figure 2*: Frequency-power slopes across the 2-8Hz range for all speakers.*

The GAM fitted power across all gestures shown for gestures produced during Spanish or English (ribbon indicates 95%CI). The are slight differences between languages, such that gestures produced under English generally are more defined by faster time scale oscillations (> 3Hz) as compared to gestures produced under Spanish. Gestures produced under Spanish, in contrast, showing modulations at slower scale oscillations (2-3Hz) as compared to gestures produced under English.

4. Discussion

We find that gestures produced by Spanish-English bilinguals are more homogenous in their quasi-rhythmic features of hand-movement speed when those gestures are compared within languages rather than between languages. We further obtained that gestures produced under Spanish speech versus English speech have reliably different modulations at the 2-8Hz range.

We speculate that the unique frequency compositions for gestures performed under Spanish versus English may be due to gestures patterning with differences in prosodic patterning between Spanish and English. For example, it has been shown that Spanish and English have divergent rhythmic structures[29], which have classically (but controversially) been equated with English being a stress-timed language and Spanish a syllable-timed language [30]. We also know that gesture kinematics are likely to spontaneously follow prosodic features in that they tend to align with stressed syllables[8, 12]. Furthermore, if there are syntactic differences in where stress is produced in a sentence between languages, gesture are likely to follow as well[31]. Note though, that the current analysis does not necessarily show that gestures are modulated at particular frequencies *because* they couple or resonate to quasi-rhythmic features of speech. It could very well be that gesture styles change based on the social contexts you are used to speak a particular language in, which likely differ between languages (e.g., the current bilinguals mainly spoke Spanish at home). More research is thus clearly needed to further asses the way in which speech quasi-rhythmic modulations are related to gesture patterning [see e.g., 32]. We have showcased a potential analysis one can employ to address these questions, and some promising results underlining the need for future research in this direction.

5. Author note

This research has been funded by The Netherlands Organisation of Scientific Research (NWO; Rubicon grant “Acting on Enacted Kinematics”, Grant Nr. 446-16-012; PI Wim Pouw). W.P. has further been supported by a DCC fellowship awarded by the Donders Institute for Brain, Cognition and Behaviour, and a postdoctoral position within the Language in Interaction Consortium (Gravitation Grant 024.001.006 funded by the NWO).

6. References

[1] Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 1–13. https://doi.org/10.1038/s41583-020-0304-4

[2] Kelso, J. A., & Tuller, B. (1984). Converging evidence in support of common dynamical principles for speech and movement coordination. *The American Journal of Physiology*, *246*,R928-935.https://doi.org/10.1152/ajpregu.1984.246.6.R928T.

[3] McNeilage, P. (2008). *The origin of speech*. New York: Oxford University Press.

[4] Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

[5] Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, *134*(1), 628–639. https://doi.org/10.1121/1.4807565

[6] Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

[7] Brookshire, G., Lu, J., Nusbaum, H. C., Goldin-Meadow, S., & Casasanto, D. (2017). Visual cortex entrains to sign language. *Proceedings of the National Academy of Sciences*, 201620350. https://doi.org/10.1073/pnas.1620350114

[8] Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–232. https://doi.org/10.1016/j.specom.2013.09.008

[9] Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3*(1), 71–89. https://doi.org/10.1515/lp-2012-0006

[10] Shattuck-Hufnagel, S., & Prieto, P. (2019). Dimensionalizing co-speech gestures. In *Proceedings of the International Congress of Phonetic Sciences 2019* (p. 5). Melbourne, Australia.

[11] Rusiewicz, H. L., Susan, S., Iverson, J., & Szuminsky, N. (2013). Effects of prosody and position on the timing of deictic gestures. *Journal of Speech, Language, and Hearing Research*, *56*(2), 458–470. https://doi.org/10.1044/1092-4388(2012/11-0283)

[12] Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J.-L. (2008). The speech focus position effect on jaw–finger coordination in a pointing task. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1507–1521. https://doi.org/10.1044/1092-4388(2008/07-0173)

[13] Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal Realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*(3), 850–864. https://doi.org/10.1044/1092-4388(2012/12-0049)

[14] Pouw, W., Harrison, S. J., & Dixon, J. A. (2019). Gesture-speech physics: The biomechanical basis of the emergence of gesture-speech synchrony. *Journal of Experimental Psychology: General*, *149*(2), 391–404. https://doi.org/10.1037/xge0000646

[15] Pouw, W., Esteve-Gibert, N., Harrison, S., & Dixon, J. A. (2019). Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. https://doi.org/10.31234/osf.io/rnpav

[16] Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2020). Acoustic information about upper limb movement in voicing. *Proceedings of the National Academy of Sciences*, *117*(21), 11364-11367. https://doi.org/10.1073/pnas.2004163117

[17] Ginosar, S., Bar, A., Kohavi, G., Chan, C., Owens, A., & Malik, J. (2019). Learning individual styles of conversational gesture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3497–3506). Retrieved from https://arxiv.org/abs/1906.04160

[18] Kucherenko, T., Hasegawa, D., Henter, G. E., Kaneko, N., & Kjellström, H. (2019). Analyzing Input and Output Representations for Speech-Driven Gesture Generation. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents - IVA ’19* (pp. 97–104). Presented at the the 19th ACM International Conference, Paris, France: ACM Press. https://doi.org/10.1145/3308532.3329472

[19] Alviar, C., Dale, R., & Galati, A. (2019). Complex Communication Dynamics: Exploring the Structure of an Academic Talk. *Cognitive Science*, *43*(3), e12718. https://doi.org/10.1111/cogs.12718

[20] Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

[21] Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes*, *22*(8), 1212–1236. https://doi.org/10.1080/01690960701461426

[22] Azar, Z., Backus, A., & Özyürek, A. (2019). Language contact does not drive gesture transfer: Heritage speakers maintain language specific gesture patterns in each language. *Bilingualism: Language and Cognition*, 1–15. https://doi.org/10.1017/S136672891900018X

[23] Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General*, *143*(3). https://doi.org/10.1037/a0036281

[24] Rosch, A., & Schmidbauer, H. (2014). WaveletComp 1.1: A guided tour through the R package, 59.

[25] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Team, R. C. (2019). *nlme: Linear and nonlinear mixed effects models*. R.

[26] Pouw, W., & Dixon, J. A. (2019). Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles. *Discourse Processes*. https://doi.org/10.1080/0163853X.2019.1678967

[27] Csárdi, G. (2019). *Package “igraph” Network Analysis and Visualization*. Retrieved from http://bioconductor.statistik.tu-dortmund.de/cran/web/packages/igraph/igraph.pdf

[28] Clarke, E., & Sherrill-Mix, S. (2017). *ggbeeswarm: Categorical Scatter (Violin Point) Plots*. Retrieved from https://CRAN.R-project.org/package=ggbeeswarm

[29] Ramus, F., Nespor, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition*, *75*(1), AD3–AD30. https://doi.org/10.1016/S0010-0277(00)00101-3

[30] Redford, M. A. (2015). *The handbook of speech production*. West Sussex: Wiley Blackwell.

[31] Kashiwadate, K., Yasuda, T., Fujita, K., Kita, S., & Kobayashi, H. (2020). Syntactic Structure Influences Speech-Gesture Synchronization. *Letters on Evolutionary Behavioral Science*, *11*(1), 10–14. https://doi.org/10.5178/lebs.2020.73

[32] Alviar, C., Dale, R., Dewitt, A., & Kello, C. T. (2020). Multimodal Coordination of Sound and Movement in Music and Speech. *Discourse Processes*, 1-21.