The Association of Second Language Proficiency with Nonverbal Behaviors

Katsunori Kotani†, Takehiko Yoshimi‡

†Kansai Gaidai University, ‡Ryukoku University

Abstract

This study investigates the association between nonverbal behaviors and second language proficiency. Proficient learners exhibit different nonverbal and verbal behaviors, such as gestures, speech rates, and vocabulary use. These distinctive properties can help analyze the language-learning process and develop evaluation tools for assessing proficiency levels. In this study, we compared the motion recognition results of learners' nonverbal behaviors using Principal Component Analysis. The results suggest that motion recognition outcomes can depict the distinctive properties of nonverbal behaviors. This research suggests that future studies should employ motion recognition for further research in this field.

1 Introduction

Second language (L2) learning is a complex process that involves nonverbal and verbal behaviors because of the importance of nonverbal behaviors in practical communication settings (Abercrombie 1968). Previous research (Gregersen et al. 2009; Franceschi 2018; Bonsignori and Cappelli 2022; Ishikawa 2022; Lopez-Ozieblo 2023) has examined L2 learners' nonverbal behaviors and their implications for L2 research and education. Some studies (Franceschi 2018; Bonsignori and Cappelli 2022) have focused on describing L2 learners' nonverbal behaviors and have pointed toward the advantages of multimodal analyses in understanding L2 learners' proficiency. Others (Gregersen et al. 2009; Lopez-Ozieblo 2022; Ishikawa 2022) have quantitatively investigated the relationship between proficiency and nonverbal behaviors.

However, the previous approaches have certain limitations. They often fail to account for the subtle movements that do not necessarily constitute gestures with pragmatic meanings. For instance, L2 learners frequently beat their fingers, an action that is irrelevant to the discourse, to solve linguistic problems by thinking about vocabulary/grammar or by ensuring fluent pronunciation (Lopez-Oizeblo 2022). Henceforth, this limitation is referred to as the Type Problem. Additionally, previous approaches may not take into account parallel nonverbal behaviors, particularly when learners produce independent gestures simultaneously. Henceforth, this limitation is referred to as the Token Problem. The Token Problem can manifest itself alongside the Type Problem, particularly because when the visibility of gestures varies, less-visible gestures (i.e., subtle movements) are often overlooked.

This study identifies a broad range of nonverbal behaviors generated by L2 learners and their association with proficiency levels to address this issue. Thus, we employed a motion-tracking technique to capture subtle facial and body movements in each frame of video data. This comprehensive multimodal analysis covers both gestures (Kendon 2004) and subtle nonverbal cues, offering insights into the intricate associations between nonverbal behaviors and L2 proficiency.

2 Previous research on proficiency and nonverbal behaviors

Previous research on the relationship between proficiency and nonverbal behaviors can be classified into two categories. One type of research (Franceschi 2018; Bonsignori and Cappelli 2022) has described the effective use of nonverbal behaviors and pointed toward its pedagogical advantages. The other type of research has statistically examined the relationship between proficiency and nonverbal behaviors (Gregersen et al. 2009; Lopez-Ozieblo 2023; Ishikawa 2022).

2.1 Qualitative analyses

Franceschi (2018) described 34 learners' nonverbal behavior and contextual gestures in lawyer–client interviews in pairs, as well as pronunciation and the use of vocabulary/grammar. The proficiency levels of the learners ranged between the B1 and B2 levels in the Common European Framework of Reference (CEFR) (Council of Europe 2001). The legal discourses were manually transcribed and annotated for nonverbal and verbal behaviors. The annotated information on nonverbal behavior describes the types and functions of gestures. The transcription and annotation tiers were accompanied by still-frame images of nonverbal cues at significant moments as judged by the annotators. Figure 1 shows a student acting as a lawyer speaking to a student acting as a client using metaphorically iconic hand movements that replicate the semantic content. The multimodal analysis appropriately identified the communication skills of a lawyer student who succeeded in constructing a rapport with a client student using nonverbal cues without verbal communication.

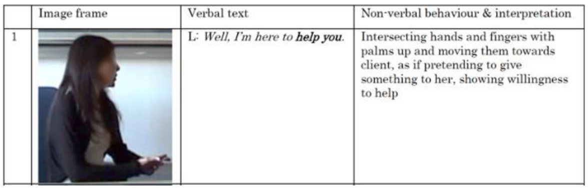

	Image frame	Verbal text	Non-verbal behaviour & interpretation
1		L: *Well, I'm here to* ***help you***.	Intersecting hands and fingers with palms up and moving them towards client, as if pretending to give something to her, showing willingness to help

Figure 2: The screenshot (Franceschi 2018), where the author of this paper has obscured the face

Bonsignori and Cappelli (2022) described 49 learners' nonverbal and verbal behaviors in tourism discourse. They compared the learners' nonverbal and verbal behavior before and after instruction. The tourism discourse was annotated according to the types and functions of nonverbal and verbal behaviors. As the goal of this multimodal analysis was to comprehensively capture both nonverbal and verbal behaviors, the discourse was represented fully with sound waves and orthographic transcriptions using the annotation tool ELAN (Lausberg and Sloetjes 2009), as shown in Figure 2. However, the nonverbal behaviors might have been unawarded at 00:00:06:965, when the red line is drawn on the tiers. Here, the learner had clasped their hands. This misidentification could be considered as a Type Problem with regard to manual annotation due to unobtrusive observations (Kipp et al. 2007). Here, the multimodal analysis should have identified hand-clasping as an ineffective nonverbal cue. Thus, despite a learner beginning a discourse, the hand-clasping gesture could be misinterpreted as a sign of the end of a proposition or topic sequence (Simmons-Mackie and Damico 1996).

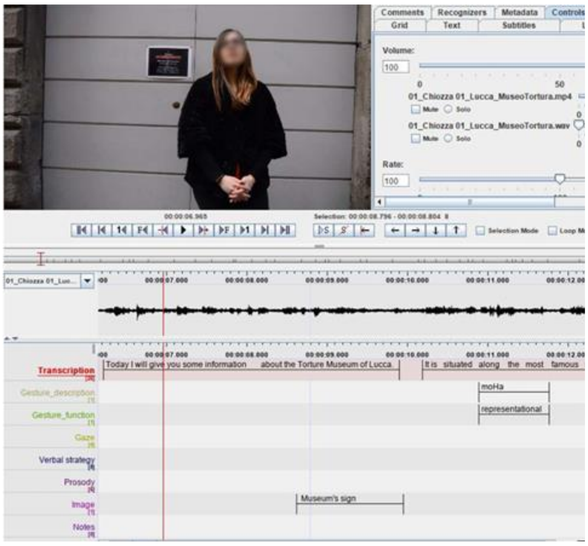

Figure 1: The screenshot (Bonsignori and Cappelli 2022), where the author of this paper has obscured the face

2.2 Quantitative analyses

Gregersen et al. (2009) confirmed a significant dependency between proficiency and the frequency and type of gestures. The experimental data consisted of interviews between two to six minutes, with an average interview duration of 2 minutes 55 seconds. The data were compiled from 75 English-speaking Spanish learners who were classified into beginner (n = 24), intermediate (n = 37), and advanced (n = 14) groups. The results revealed that advanced-level learners used significantly more speech-related and meaning-enhancing gestures than beginner- and intermediate-level learners.

Lopez-Ozieblo (2022) confirmed that changes in proficiency and the frequency and types of gestures depend on proficiency. The experimental data included two narrations: a 446-word story in 3.1 minutes and a 274-word story in 1.8 minutes. These narrations were performed by one learner at the beginning and end of the 6-month experimental period. Two annotators transcribed the gestures and noted their types and durations, which were subsequently checked by another annotator. The results revealed a negative correlation, in which the number of gestures per clause dropped from 0.38 to 0.35. Additionally, the referential- and deictic-type gestures increased, whereas that of beat-type gestures remained stable. Conversely, the number of discursive-type gestures decreased.

Ishikawa (2022) examined the correlation between learners' proficiency and the frequency and types of five types of gesture. The gestures included tilting or lifting the head, touching the head, moving the hand, and pointing at the picture, as shown in Figure 3. These gestures were manually annotated at a four-level frequency (0, 1, 2, and 3≤). The results indicated a weak correlation between fluency and gesture frequency (r = 0.28), whereas no correlation was found between fluency and gesture type (r = -0.15).

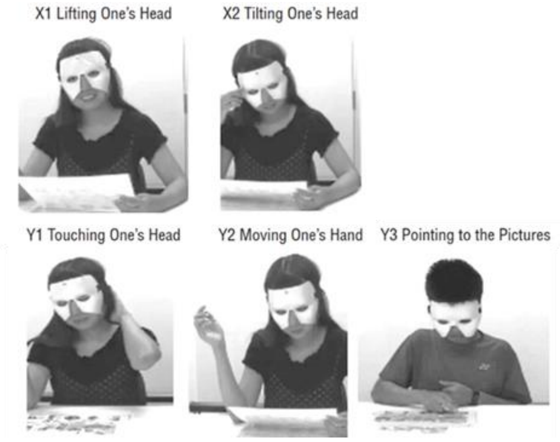

Figure 3: The screenshot (Ishikawa 2022)

3 Association between nonverbal behaviors and proficiency

3.1 Data for analysis

This study investigated how learners' nonverbal behaviors are associated with their proficiency levels. The experimental data consisted of English learners' monologues obtained from the International English Language Testing System (IELTS) mock exams on speaking proficiency available on YouTube (Christopher G. (2022). IELTS SPEAKING Mock Exam. IELTS Daily). Three learners were involved in the experiment: a beginner-level learner (Beg), an intermediate-level learner (Int), and an advanced-level learner (Adv). The learners' speaking proficiency levels were assessed based on the mock test scores, which ranged from 4.0 to 9.0. Each learner spoke for approximately 2 minutes on a given topic, following a one-minute preparation period. To ensure uninterrupted speaking, the interviewer refrained from interrupting the learners during the speaking session. Table 1 presents information regarding the experimental data, including details regarding the topics and the phonetic properties of the monologues. The monologue topics were carefully selected to facilitate the expression of opinions and information related to common experiences. The phonetic properties (De Jong and Wempe 2008) show that the duration of the monologues adhered to the designated time limit. The beginner-level learner produced the least number of syllables but the most pauses. The articulation rate (AR), which is defined as syllables per second in the given duration without pauses, depended on the proficiency levels.

	Beg	**Int**	**Adv**
Topic	Opinion: A book you want to read	Description: A person you know who is kind	Experience: A time when you were disappointed by someone/ something
Score	4.0	6.5	9.0
Duration (sec.)	120.36	102.68	97.92
Syllables	353	419	398
Pauses	47	21	23
AR.	4.39	4.63	4.67

Table 1: Details of the monologues

3.2 Method

The raw experimental data were preprocessed to identify relevant nonverbal behaviors. The learners' nonverbal gestures were recognized using the MediaPipe Holistic Landmarker (Grishchenko and Bazarevsky 2020). This gesture recognition tool generates landmarks for the face, body, and hands in each frame, and provides continuous streams of images at the rate of 25 frames per second. Of the 543 generated landmarks (468 face landmarks, 33 pose landmarks, and 21 hand landmarks per hand), 11 specific landmarks were analyzed in this study: the nose, left/right eye, left/right ear, left/right shoulder, left/right elbow, and left/right wrist. This selection was based on the visibility of the learners'

Figure 4: Screenshots of the learners' monologues

upper bodies and the absence of hand gestures in the intermediate and advanced-level learners. The screenshots in Figure 4 show the three learners' monologues, where beginner-level, intermediate-level, and advanced-level learners appear on the left, upper right, and lower right sides, respectively.

Each landmark is composed of *x*, *y*, and *z* three-dimensional coordinates. The *x* and *y* coordinates were normalized to [0.0, 1.0] based on the image width and height, respectively. The *z* coordinate represents the landmark depth. The origin points of the *x* and *y* were positioned at the top-left corner of the frame image, and the origin point of the *z* axis was positioned at the wrist of the target person. Landmark coordinates were employed to calculate the Euclidean distances (*x*, *y*, *z*) from the origin (0, 0, 0), and the Euclidean distance of each landmark (henceforth, the landmark distance) was measured in each frame of the video data.

The primary goal of this study was to investigate the relationship between landmark distances and L2 proficiency in the following ways. Initially, we visually assessed the relative differences in pitches by visually comparing the landmark distance plots without applying any normalization because we did not compare the magnitudes and intensities of the waveforms. This relative comparison helps to understand the differences in the distinctive patterns. Furthermore, Principal Component Analysis (PCA) was conducted on the landmark distances and proficiency using the R software (Mizumoto 2015). PCA allowed us to explore the inherent associations among the landmark distances. The primary objective was to identify the principal components that explain the variability in the landmark distances and interpret the relationships revealed through PCA. Finally, the PCA results of the learners were compared at different proficiency levels to determine the potential differences in landmark-distance patterns. The aim was to provide insights into how landmark distances may vary with language proficiency, and their potential implications for L2 research and education.

3.3 Results

Table 2 presents the number of frames (*n*) and the mean landmark distances across the discourse processes. As the origin was located at the top-left corner, the average landmark distances of the face landmarks were comparatively lesser than those of the pose landmarks, indicating a distinct pattern of spatial dynamics in nonverbal behaviors.

	Beg (*n*=3009)	**Int** (*n*=2567)	**Adv** (*n*=2448)
Nose	0.64	0.64	0.62
L_Eye	0.63	0.61	0.64
R_Eye	0.59	0.58	0.59
L_Ear	0.63	0.63	0.68
R_Ear	0.55	0.58	0.60
L_Shoulder	0.77	0.85	0.92
R_Shoulder	0.64	0.80	0.73
L_Elbow	1.04	1.09	1.18
R_Elbow	0.94	1.06	1.01
L_Wrist	1.05	1.34	1.26
R_Wrist	1.01	1.37	1.08

Table 2: Mean Euclidean distances of landmarks

Figure 5 presents the plots of the landmark distances of the Adv (top row), Int (middle row), and Beg (bottom row). The landmark plots represent the (dis)similarity of nonverbal behaviors across the beginner, intermediate, and advanced proficiency levels.

Table 3 shows the PCA results for the landmark coordinates. This revealed that the first principal component accounted for 53.9% (Beg), 69.5% (Int), and 79.6% (Adv) of the variance, indicating substantial contributions to capturing variability. Moreover, the combination of the first and second principal components explained 73.7% (Beg), 80.9% (Int), and 87.9% (Adv) of the variance, demonstrating their significant role in characterizing the data.

The eigenvalues of the principal components provide insights into the variance explained by these components. The eigenvalues declined steeply between the first and second components, indicating that these initial components effectively captured substantial portions of the variances. Thus, these principal components were selected to explore their associations with the proficiency levels.

The principal component loadings provide insights into the correlations between the original variables and the principal components. Figure 6 shows that the signs of the correlation coefficients in the first principal component (PC1) varied depending on the proficiency level. Specifically, positive correlations were observed for the L_Shoulder variable at the beginner level, numerous variables at the intermediate level, and all variables at the advanced level.

Figure 5: The plots of the landmarks in the Euclidean distance in Adv. (the top row), Int. (the middle row), and Beg. (the bottom row)

Beg	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalues	5.934	2.176	1.178	0.982	0.291	0.252
Standard Deviation	2.436	1.475	1.085	0.991	0.540	0.502
Proportion of Variance	0.539	0.198	0.107	0.089	0.027	0.023
Cumulative Proportion	0.539	0.737	0.844	0.934	0.960	0.983
	PC7	**PC8**	**PC9**	**PC10**	**PC11**	
Eigenvalues	0.101	0.054	0.030	0.001	0.001	
Standard Deviation	0.317	0.233	0.174	0.029	0.023	
Proportion of Variance	0.009	0.005	0.003	0.000	0.000	
Cumulative Proportion	0.992	0.997	1.000	1.000	1.000	
Int	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Eigenvalue	7.645	1.254	1.003	0.394	0.290	0.217
Standard Deviation	2.765	1.120	1.002	0.628	0.538	0.466
Proportion of Variance	0.695	0.114	0.091	0.036	0.026	0.020
Cumulative Proportion	0.695	0.809	0.900	0.936	0.963	0.982
	PC7	**PC8**	**PC9**	**PC10**	**PC11**	
Eigenvalues	0.147	0.042	0.006	0.001	0.000	
Standard Deviation	0.383	0.205	0.079	0.023	0.014	
Proportion of Variance	0.013	0.004	0.001	0.000	0.000	
Cumulative Proportion	0.996	0.999	1.000	1.000	1.000	
Adv	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Eigenvalues	8.755	0.915	0.690	0.348	0.202	0.038
Standard Deviation	2.959	0.957	0.830	0.590	0.450	0.195
Proportion of Variance	0.796	0.083	0.063	0.032	0.018	0.003
Cumulative Proportion	0.796	0.879	0.942	0.974	0.992	0.995
	PC7	**PC8**	**PC9**	**PC10**	**PC11**	
Eigenvalues	0.027	0.015	0.009	0.001	0.000	
Standard Deviation	0.165	0.122	0.094	0.025	0.020	
Proportion of Variance	0.002	0.001	0.001	0.000	0.000	
Cumulative Proportion	0.998	0.999	1.000	1.000	1.000	

Table 3: Importance of components in Beg, Int, and Adv

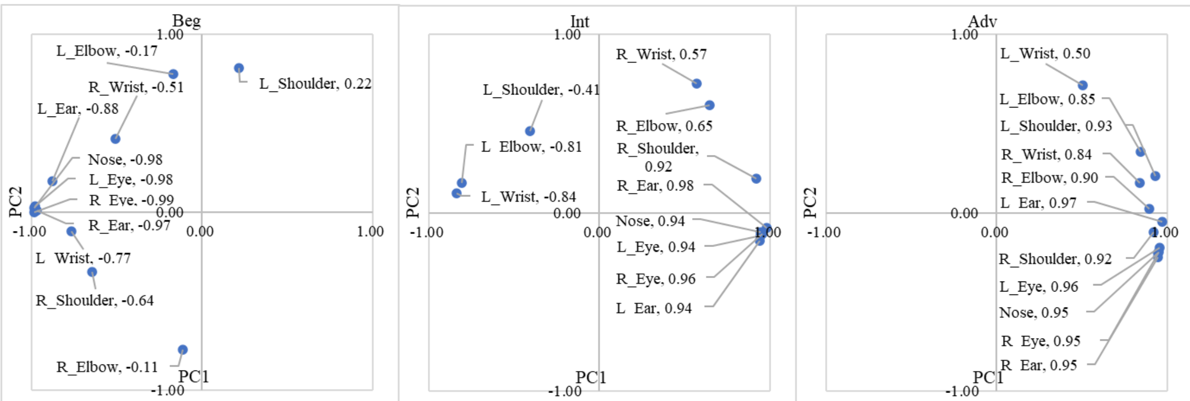

Figure 6: Principal component analysis plots for PC1 and PC2 in Beg, Int, and Adv

Figure 6 illustrates the PCA plots for PC1 and PC2 representing the principal components across the beginner, intermediate, and advanced proficiency levels. These plots visually represent the distribution and patterns of the data and provide valuable insights into the underlying associations.

3.4 Discussion

The results presented in Figure 5 suggest that analyzing gestures based on plots can effectively address the Type and Token Problems, as discussed in Section 1. The Token Problem is evidently resolved, as the plots clearly illustrate the movements of various body parts, including facial features, throughout the discourse process. Additionally, the Type Problem is also successfully managed, as the plots capture subtle movements of a learner's body parts. For instance, when a plot exhibits a low-pitch wave, as observed in the plots for the beginner-level learner's body movements (indicated by blue lines), it indicates slight movements of the body parts. Consequently, these low-pitch movements represent the motion of body parts that would be hard to identify manually. However, it is important to note that these plots also effectively capture readily identifiable movements, as seen in the advanced-level learner's plots for body movements (once again, indicated by blue lines).

The results in Table 3 indicate that a plot-based analysis of gestures can effectively account for the dependency of gestures on learners' proficiency levels. Based on the PCA results, we observed high cumulative proportions of PC1 and PC2, with values of 73.7%, 80.9%, and 87.9% for the Beginner, Intermediate, and Advanced proficiency levels, respectively. These substantial cumulative proportions confirm the significance of recognizing and analyzing facial and body movements in learners' discourse processes at these proficiency levels. A substantial portion of the total variance explained by PC1 and PC2 demonstrates that these principal components successfully capture essential information about learners' nonverbal behaviors during the discourse process.

The PCI results drawn in Figure 6 suggest the potential to estimate learners' proficiency levels based on landmark distances, in addition to considering linguistic properties such as speech rate, sentence length, and word types in learners' utterances. PC1, as the primary underlying factor, captures the highest variance and reflects the general trend of increasing landmark values throughout the discourse process. Comparing PC1 values among proficiency levels, we observed that advanced learners had the positive PC1 values, beginner learners had the negative values, and intermediate learners fell in between. These findings indicate a correlation between PC1 scores and proficiency levels, representing a dimension closely associated with nonverbal behaviors. Specifically, learners' gestures appear to be influenced by their proficiency levels. Proficient learners tend to employ more communicative gestures that enhance discourse understanding. Conversely, less proficient learners rely more on nonverbal cues, not necessarily for effective discourse communication, but as a means to release the stress of using the target language. Thus, the PCA results underscore the potential for estimating learners' proficiency levels based on landmark distances and linguistic properties, such as speech rate, sentence length, and word types in their utterances.

Finally, these results presented a direction of future research for multimodal learner corpus research. These plots, along with landmark positions, can facilitate the identification of gesture phases (e.g., preparation, stroke, and hold) based on distinctive features of ±movement, ±constant, and ±increase (Bressem and Ladewig 2011).

4 Conclusion

This study explores the association between language proficiency and nonverbal behavior. Specifically, we employed MediaPipe, a motion recognition tool, to identify nonverbal behaviors, and the *x-y-z* three-dimensional coordinates were converted into Euclidean distances for 11 facial and body landmarks. The distribution of the Euclidean distances in each frame exhibited variations among learners at different proficiency levels. Furthermore, we analyzed the Euclidean distance distributions using Principal Component Analysis. Both results indicate that motion recognition effectively captures the distinctive properties associated with language proficiency. These empirical findings support the integration of motion recognition into the multimodal analyses of learner corpora.

Limitations

This study examines the association between L2 learners' proficiency and nonverbal behaviors and demonstrates that motion recognition results effectively capture distinctive properties related to proficiency levels. However, we acknowledge the following limitations in our study.

First, the experimental data were derived from short monologues by only one beginner-, intermediate-, and advanced-level learner. Consequently, the sample size may not fully represent the diversity of proficiency levels among L2 learners.

Another limitation concerns the choice of motion recognition tool. Though we employed MediaPipe, we did not use other tools such as OpenPose (Cao et al. 2016), AlphaPose (Fang et al. 2022). Future research could compare multiple motion recognition tools to assess their impact on the results.

Moreover, we used a subset of facial landmarks provided by MediaPipe, focusing on only five primary facial landmarks (nose, left and right eyes, and left and right ears) of the 468 facial landmarks available. This selection might have limited the comprehensiveness of our analysis of nonverbal behaviors. Examining a broader set of facial landmarks could yield deeper insight into the relationship between proficiency and nonverbal cues.

In conclusion, though our study provides valuable insights into the association between proficiency and nonverbal behavior among L2 learners, it is essential to acknowledge and consider these limitations when interpreting the results. Future studies should address these limitations to offer more comprehensive and valuable insights.

Ethics Statement

This study addressed several ethical considerations. First, this study compiled the experimental data from publicly available videos on YouTube. In addition, the learners' privacy and identity was protected by obscuring their faces.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers, 22300299, 15H02940, 17K18679, and 21K18373.

References

David Abercrombie. 1968. Paralanguage. *International Journal of Language & Communication Disorders*, 3:55–59. https://doi.org/10.3109/13682826809011441

Veronica Bonsignori and Gloria Cappelli. 2022. Developing strategies for conceptual accessibility through multimodal literacy in the English for tourism classroom. *Multimodal Communication*, 11(1):31–47. https://doi.org/10.1515/mc-2020-0026

Jana Bressem and Silva H. Ladewig. 2011. Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica*, 184:53–91. https://doi.org/10.1515/semi.2011.022

Zhe Cao, Ginés Hidalgo, Thomas Simon, Shih-En Wei and Yaser Sheikh. 2019. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43:172–186. https://doi.org/10.48550/arXiv.1812.08008

Council of Europe. 2010. Guide for the Development and Implementation of Curricula for Plurilingual and Intercultural Education. Language Policy Unit. Strasbourg: Council of Europe.

Nivja de Jong and Ton Wempe. 2008. Praat Script Syllable Nuclei v2.

Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. 2022. AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(6): 7157-7173. https://doi.org/10.1109/TPAMI.2022.3222784

Daniele Franceschi. 2018. Simulating a lawyer-client interview: the case of Italian EFL University learners. *ESP Across Cultures*, 15:41–59.

Tammy Gregersen, Gabriela Olivares-Cuhat, and John Storm. 2009. An examination of LI and L2 gesture use: What role does proficiency play? *The Modern Language Journal* 93(2): 195–208. https://doi.org/10.1111/j.1540-4781.2009.00856.x

Ivan Grishchenko and Valentin Bazarevsky, R. E. 2020. *MediaPipe Holistic–Simultaneous Face, Hand and Pose Prediction, on Device*. Retrieved March 9, 2023.

Shin'ichiro Ishikawa. 2022. L2 English learners' gesture use in a picture description task: A study based on the ICNALE Dialogue Module. In Raúl Ruiz Cecilia and Juan Ramón Guijarro Ojeda (eds.). *Investigación e Innovación en Lengua Extranjera: Una Perspectiva Global*, pages 721–750.

Adam Kendon. 2004 *Gesture: Visible Action as Utterance*. Cambridge University Press: Cambridge.

Michael Kipp, Michael Neff, and Irene Albrecht. 2007 An annotation scheme for conversational gestures: how to economically capture timing and form. *Lang Resources & Evaluation* 41:325–339. https://doi.org/10.1007/s10579-007-9053-5

Hedda Lausberg and Han Sloetjes. 2009. Coding gestural behavior with the NEUROGES–ELAN system. *Behavior Research Methods*, 41(3):841–849. https://doi.org/10.3758/BRM.41.3.841

Renia Lopez-Ozieblo. 2023. Observations on the progression of gestures with L2 proficiency: A call for further research. *Journal of Psycholinguistic Research*. 52(2):381–404. https://doi.org/10.1007/s10936-022-09862-y

Atsushi Mizumoto. 2015. *Code for "Principal Component Analysis."* https://langtest.jp/shiny/pca/ Last accessed on 2023-7-9.

Nina N. Simmons-Mackie and Jack S. Damico. 1996. The contribution of discourse markers to communicative competence in Aphasia. *American Journal of Speech-Language Pathology*, 5(1):37–43.