İSTANBUL UNIVERSITY PRESS

Psikoloji Çalışmaları - Studies in Psychology
SP 2024, 44 (3): 349–363
DOI: 10.26650/SP2024-1443983

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

Multimodal Communication in Virtual and Face-to-Face Settings: Gesture Production and Speech Disfluency

Çevrimiçi ve Yüz Yüze İletişim Ortamlarında Multimodal Dil Kullanımı: Jest Üretimi ve Konuşma Akıcılığı Üzerine Bir Araştırma

Burcu Arslan[1], Can Avcı[2], Demet Özer[3]

[1] Dr., Koç University, Department of Psychology, Istanbul, Türkiye
[2] Koç University, Department of Psychology, Istanbul, Türkiye
[3] Assistant Professor, Kadir Has University, Department of Psychology, Istanbul, Türkiye

ABSTRACT

The COVID-19 pandemic has made online data collection a popular choice. It is important to evaluate how comparable online studies are to face-to-face studies, particularly in multimodal language research where modes of communication significantly impact the results. In this study, we examined individuals' rates and patterns of speech disfluency and gesture use across face-to-face and online videoconferencing settings as they described their daily routines (*N* = 64). We asked whether and how multimodal language is affected across different communication settings and gesture use, particularly iconic gestures, is associated with speech fluency regardless of the context. Our results have showed that the participants' overall disfluency rate was higher for the speech communicated via videoconferencing than the speech communicated face-to-face. However, the type of disfluencies changed across contexts, such that filled pauses and repairs were more common in online communication, whereas silent pauses were more common in face-to-face communication. These findings signal an interplay between the cognitive functions of different disfluency types and communicative strategies. Results indicate that the overall gesture frequency and iconic gesture use were similar in both settings. Furthermore, the use of iconic gestures was found to negatively predict the overall disfluency rate, regardless of the setting. This finding suggests that using iconic gestures might facilitate cognitive processes, paving the way for a more fluent speech. This study demonstrates that multimodal language and communication strategies may vary across different communication settings and nuanced understanding of the differences in multimodal language between online and face-to-face communication can be gained using different contexts. The findings contribute to understanding the impact of increasingly widespread online communication on multimodal language production processes and provide foundation for future research.

Keywords: Gesture production, speech disfluency, virtual communication, face-to-face communication

ÖZ

Çevrimiçi veri toplama, COVID-19 salgını nedeniyle öne çıkan bir seçenek haline gelmiştir. Çevrimiçi çalışmaların, özellikle bağlamın çok önemli bir etkiye sahip olduğu dil ve iletişim alanlarında, yüz yüze yapılan çalışmalarla ne ölçüde karşılaştırılabileceğini anlamak çok önemlidir. Bu çalışma, yüz yüze ve video konferans ortamlarında multimodal iletişimi araştırmak amacıyla, kişilerin (*N*= 64) günlük rutinlerini anlatırken kullandıkları konuşma akıcılıklarına ve sözlü dile eşlik eden jest üretimlerine odaklanmaktadır. Çalışmada, el jestlerinin ve sözlü dildeki akıcılığın farklı iletişim ortamlarında (çevrim içi ve yüz yüze) nasıl değiştiği ve iletişim ortamından bağımsız olarak jest kullanımının (özellikle ikonik jestlerin) konuşma akıcılığıyla ilişkisi araştırılmaktadır.

Corresponding Author: Demet Özer **E-mail:** demet.ozer@khas.edu.tr
Submitted: 29.02.2024 • **Revision Requested:** 26.06.2024 • **Last Revision Received:** 24.07.2024 • **Accepted:** 02.10.2024
Published Online: 29.11.2024

Çalışmanın sonuçları, konuşma akışındaki bozulma oranının video konferans yoluyla iletişim kuranlarda, yüz yüze iletişim kuranlara göre daha yüksek olduğunu göstermiştir. Fakat konuşma akıcılığındaki farklı bozulma türlerinin iki ortamda farklılık gösterdiği bulunmuştur. Konuşmacıların video konferans ortamında daha fazla dolgulu duraksama ve onarım kullanırken, yüz yüze iletişim ortamında daha fazla sessiz duraksama kullandığı bulunmuştur. Bu bulgular, konuşmanın akıcılığındaki bozulmaların iletişim ortamına göre değişebileceğini ve farklı iletişim stratejileri doğurabileceğini göstermektedir. Bunun yanında, genel jest kullanımının ve özel olarak ikonik (temsili) jest kullanımının iki ortam arasında fark göstermediği bulunmuştur. Ayrıca, iletişim ortamından bağımsız olarak, ikonik jest kullanım sıklığının konuşma akıcılığını artırdığı bulunmuştur. Bu bulgu, özellikle ikonik jest kullanımının bilişsel süreçleri kolaylaştırarak daha akıcı bir konuşmaya zemin hazırlayabileceğini göstermektedir. Bu çalışma, multimodal dil ve iletişim stratejilerinin farklı iletişim ortamlarında değişebildiğini göstermekte ve bu anlamda bağlamın araştırılmasının önemini vurgulamaktadır. Sonuçlar, günümüzde özellikle yaygınlaşan çevrimiçi iletişimin multimodal dil üretim süreçleri üzerindeki etkilerini anlamaya katkı sağlamaktadır ve ileride yapılacak çalışmalar için önemli bir temel sunmaktadır.

Anahtar Kelimeler: Jest kullanımı, konuşmanın akıcılığının bozulması, çevrimiçi iletişim, yüz yüze iletişim

Individuals' speech is accompanied by disfluent segments such as pauses, repetitions, or revisions. Since speech production requires a detailed planning process, disfluent segments in one's speech might be a byproduct of cognitive load (Bock, 1996; Fraundorf & Watson, 2014). Language is multimodal. People spontaneously gesture when they speak. It is argued that gesture and speech are closely linked mechanisms (Kita & Özyürek, 2003). Moreover, gestures facilitate speech production processes as they enhance the production of a more fluent speech by decreasing the cognitive load stemming from planning and word retrieval (Kita et al., 2017; Krauss et al., 2000). Then, one might expect gestures to come into play when the speech planning process becomes costly. Indeed, research has demonstrated that using gestures might aid speech production (Krauss et al.,, 2000; Özer et al., 2017; Rauscher et al., 1996).

The coronavirus (COVID-19) outbreak has severely affected the world since early 2020. Some precautions were taken to prevent the spread of the disease. Most governments have designated lockdowns and travel restrictions despite the increasing number of cases across the world. These precautions also included moving from face-to-face settings to virtual settings in education and business. Instead of meeting face-to-face, people and organizations started setting up online meetings using different videoconferencing tools. This situation impacted research as well since face-to-face data collection became almost impossible under these conditions. As a result, online data collection became a more prominent option for researchers. Studies indicate that videoconferencing might be a viable method for qualitative and quantitative data collection (Archibald et al., 2019; Glassmeyer & Dibbs, 2012; Torrentira, 2020). However, one might ask to what extent the findings obtained from online studies can be compared with the findings of face-to-face studies, particularly in language research on which the modes of communication can have a crucial effect. Research has indicated that, compared to face-to-face communication, communicating via video chat might require people to invest more attention in terms of ignoring distractors around and making sure that they are in the view of the camera (Cserző, 2021). Such a cognitive load might influence individuals' language production processes and communication strategies.

This study compares multimodal communication across face-to-face and virtual settings to understand the possible differences across communication contexts. To this end, we examined gesture production and speech disfluency in face-to-face and virtual settings. We ask whether (1) the rates and patterns of speech disfluency and gesture use are similar across face-to-face and virtual settings, and (2) gesture use is associated with speech fluency in either context.

Speech Disfluency

People become disfluent when they speak. Considering that speech results from a detailed planning process, disfluency rates in one's speech might be associated with cognitive load (Bock, 1996). Research suggested that when individuals engage with difficult tasks, they are prone to be disfluent (Bortfeld et al., 2001; Morsella & Krauss, 2004). Similarly, disfluencies are more likely to be observed at the beginning of sentences (Oviatt, 1995) and in grammatically complex sentences (Silverman & Ratner, 1997), supporting the link between planning load and disfluency. However, focusing on different types of disfluency might be necessary to understand the cognitive and communicative strategies associated with speech disfluency (Arslan & Göksun, 2022; Fraundorf & Watson, 2014).

Mclay and Osgood (1959) differentiated among the four types of disfluencies observed in speech. A *filled pause* refers to filling pauses with lexical items that do not carry a propositional content (e.g., um). A *silent pause* occurs where individuals temporarily pause within a sentence. Repetition involves repeating some parts of the message, such as words (e.g., behind behind the building). A *repair* refers to revising word choices or grammatical structures despite more plausible alternatives (e.g., at the exit – I mean, entrance). Fraundorf and Watson (2014) suggested that when individuals engage with difficult tasks, silent pauses and filled pauses are more likely to be observed during the planning process. In contrast, repetitions commonly occur after the speech plan is executed. They also indicated that filled pauses might be associated with conceptual issues, whereas silent pauses and repetitions might reflect issues linked to lexical and phonological access.

Speech disfluencies, particularly filled pauses, might also be interpreted as a communicative signal as well (Bortfeld et al., 2001). That is, using filled pauses might maintain the conversational floor by signaling the listeners that the speaker has the intention to continue speaking (Corley & Stewart, 2008; Smith & Clark, 1993). Bortfeld et al. (2001) demonstrated that the frequency of filled pauses is not necessarily higher in a difficult task (e.g., describing tangrams) as opposed to an easy task (e.g., describing pictures of children), suggesting that planning load alone might not be enough to explain the use of filled pauses.

Compared to face-to-face communication, communicating through online channels such as videoconferencing might require people to invest more attention in terms of ignoring distractors around them and making sure that they are in the view of the camera (Cserző, 2021). Because of such cognitive load, people might be more disfluent in virtual settings as opposed to face-to-face settings. However, an additional interpretation from a communicative perspective might be required to better understand the use of specific disfluency types. Online communication is likely to have some pitfalls related to the quality of the internet connection, which might affect the quality of sound and visual display. When there is a weak internet connection in a videoconference meeting, one's video image might suddenly freeze and there might be synchronization problems regarding the sound. In such a context, people in the meeting are likely to ask questions to each other to ensure that they can see and hear each other without interruption. From such a perspective, using filled pauses might be effective in maintaining the conversational floor in online settings by signaling listeners that there is not a connection related issue and the speaker intends to continue as soon as the planning process is handled. In contrast, using silent pauses might be interpreted by listeners as a connection problem. Therefore, individuals may prefer using fewer silent pauses in the virtual environment as opposed to face-to-face settings.

In sum, speech disfluencies across different communication settings, such as face-to-face vs. virtual communication, are yet to be explored. Disfluencies may result from the cognitive load associated with speech planning. However, disfluency types might differ from each other in terms of the strategies they reflect. These strategies should be interpreted in the light of not only the cognitive load but also the communicative intentions. Maintaining the conversational floor in online settings such as videoconferencing might require individuals to interchangeably use cognitive and communicative strategies.

Gesture Production

Although gestures may consist of hand, head, or body movements, we specifically focus on hand movements that accompany speech, which are known as co-speech hand gestures. McNeill (1992) created categories among different co-speech gestures. *Iconic* gestures refer to concrete objects and events (e.g., drawing a line with fingers to refer to a road), and *metaphoric* gestures refer to abstract concepts (e.g., leaving a small space between two fingers to refer to a small problem). In addition, there are *deictic* gestures (e.g., pointing gestures) and *beat* gestures, which are rhythmic movements without meaning. Finally, an *emblem* conveys a culturally shared conventionalized message on its own (e.g., waving hands to mean goodbye).

Gestures are closely associated with spatial cognition (Alibali, 2005), and individuals frequently gesture in a spatial context (Arslan & Göksun, 2021). The *gesture-for-conceptualization hypothesis* suggests that using gestures, particularly iconic gestures, facilitates cognitive processes and decreases cognitive load (Kita et al., 2017). For example, using gestures helps activate, maintain, and manipulate information for thinking and speaking purposes (Kita et al., 2017), packaging complex information into small chunks that can be verbalizable for speaking (Kita & Özyürek, 2003), and facilitates lexical retrieval (Krauss et al., 2000). In line with this argument, individuals are more likely to gesture when they engage with difficult tasks (Melinger & Kita, 2007) such as describing objects that are hard to conceptualize (Kita & Davies, 2009).

Planning speech is a cognitively demanding process as one must successfully retrieve target words, choose the right grammatical form, and maintain the conversational floor. Considering that gesture and speech are closely linked mechanisms (Kita & Özyürek, 2003), and gestures have self-oriented functions in human cognition (Kita et al., 2017), one might expect gestures to facilitate the speech production process as well. Research has suggested that using gestures might aid lexical access (Krauss et al., 2000). Moreover, individuals are more likely to be disfluent when their hand use is restricted (Morsella & Krauss, 2004; Rauscher et al., 1996). These findings suggest that using gestures might facilitate the production of a more fluent speech by decreasing the planning and word retrieval related cognitive load.

Given that virtual communication became prevalent, particularly after the COVID-19 outbreak, it is important to examine multimodal communication in virtual settings. Although there are studies that examined gesture production in virtual settings through data collection in Zoom (e.g., Avcı et al., 2022; Arslan et al., 2024; Hyusein & Göksun, 2023; Kandemir et al., 2023; Özder et al., 2023), to the best of our knowledge, there are no studies that directly compare gesture production across face-to-face and virtual settings. The current study attempts to fill this gap.

Considering that online communication through videoconferencing might create an additional cognitive load that does not exist in face-to-face communication (Cserző, 2021), individuals might gesture frequently to decrease cognitive load and aid their speech production process. Moreover, in online communication, sometimes one's video image freezes, but listeners receive the sound without an interruption. Therefore, using body language effectively and gesturing might emphasize the speaker's virtual presence in a meeting. Although it is difficult to differentiate between the cognitive and communicative motivations of gesturing, communicating through online channels might increase individuals' likelihood of producing gestures.

In conclusion, using gestures, particularly iconic gestures, might have self-oriented functions in human cognition, particularly speech production. Communicating through online channels might be challenging in terms of connection and synchronization issues. Individuals might benefit from gestures to decrease cognitive load and emphasize their virtual presence in the case of possible connection issues.

This study aims to understand multimodal language use in face-to-face and virtual communication. We examined individuals' rates and patterns of speech disfluency and gesture use in face-to-face and online settings. We investigated whether language production is influenced by the channel of communication. We also investigated whether gesture use, particularly iconic gesture production, is associated with speech fluency, regardless of the experimental setting. We elicited speech and gesture samples by using a task in which participants described their daily routines either face-to-face or through videoconferencing.

In light of all this information, the hypotheses of the study are as follows:

H1. Communication via videoconferencing shows a higher rate of speech disfluency compared to face-to-face communication.

H2. Communication via videoconferencing shows a higher rate of filled pauses among all disfluencies compared to face-to-face communication.

H3. Communication via videoconferencing shows a higher rate of gestures, particularly iconic gestures compared to face-to-face communication.

H4. Iconic gesture use negatively predicts the overall disfluency rate, regardless of the communication channel, by reducing cognitive load.

Since *H4* is based on gestures' self-oriented functions (Kita et al, 2017), suggesting that gestures decrease cognitive load, we chose the disfluency rate as the outcome variable. Research indicates that gestures, particularly representational gestures, precede their lexical affiliates (Seyfeddinipur & Kita, 2001; Ter Bekke et al., 2024). We argue that gestures might be precursors of cognitive load even before the temporary disruptions occur in spontaneous speech. By decreasing the cognitive load, we argue that gestures pave the way for a more fluent speech.

Methods

Participants

This study was conducted with 64 native Turkish speakers between 18 and 28 years of age. To prevent any bias that might arise due to the familiarity with technology, we focused on a healthy young adult population. Thirty adults (17 females) (M_{age} = 21.43, *SD* = 1.38) participated in a face-to-face study as a part of a larger project that investigated gesture production in younger and older adults (Arslan & Göksun, 2021). On the other hand, 34 adults (17 females) (M_{age} = 23, *SD* = 2.72) participated in an online study via the videoconferencing method as a part of a larger project that investigated the relationship between speech disfluency and gesture (Avcı et al., 2022). All participants were recruited via convenient sampling. They were right-handed, and they reported not having a record of a neurological disorder. Participants' informed consent was obtained before the experiment. Seventeen participants were recruited through Koç University subject pool, and they received one course credit in return. The rest of the participants were recruited based on convenience and did not receive a reward for their participation. This study was approved by the Institutional Review Panel for Human Subjects of Koç University (2018.276.IRB3.195 and 2021.159.IRB3.068).

Materials

We used a seat to welcome the participants in the face-to-face sessions in a laboratory room. We used Zoom Video Conferencing (Zoom Video Communications Inc., 2016) to arrange online meetings with participants. In the online sessions, the participants had their own internet connection. Each participant used a laptop with a functioning camera and microphone. They also used a table on which they could place their laptops along with a seat in front of the camera.

Procedure

Participants were welcomed either face-to-face or online. In the face-to-face sessions, after the participants were seated, they answered some demographic questions. Then, the experimenter asked them to describe what they would do on a regular day. For the online sessions, participants received the Zoom meeting link via email and joined the online session. The experimenter opened the camera and microphone to welcome the participants and provide instructions. Participants were required to leave their camera and microphone open during the session. The experimenter asked them to sit in front of their camera in a way that their upper body could be seen. Anything regarding hand use was not mentioned to prevent bias. The participants answered the demographic questions. Then, they were asked to describe what they would do on a regular day. The camera and microphone of the experimenter were open from the beginning to the end of the session. In both the face-to-face and online sessions, there was not a time limitation, and the participants could describe their routines as they wanted. Each session took approximately 10 minutes.

Coding

Speech and Disfluency

We transcribed speech and coded speech disfluencies using the ELAN software (Lausberg & Sloetjes, 2009). In line with Maclay and Osgood (1959), we identified and coded silent pauses, filled pauses, repetitions, and repairs. A trained assistant coded all participants' speech and another trained assistant coded only 20% of the participants for reliability. The disfluency rates indicated by the coders revealed a strong correlation (r = .93, p < .001) and there was high inter-rater reliability in categorizing disfluencies (κ = .91, p < .001). The overall disfluency rate was calculated for each participant by dividing the total number of disfluencies by the total word count. The disfluency rate for each category was calculated as proportions by dividing the number of disfluencies that belonged to a specific category by the total number of disfluencies.

Gesture

We identified and coded the gestures using the ELAN software (Lausberg & Sloetjes, 2009). In line with McNeill (1992), we coded iconic, metaphoric, deictic, beat, and emblem gestures. A primary trained assistant coded all gestures of all participants, whereas another second trained assistant coded 20% of the participants for reliability. There was a strong correlation between two coders in terms of gesture rates (r = .89, p < .001). There was a high interrater agreement in categorizing gestures (κ = .87, p < .001). We calculated the overall gesture frequency for each participant by dividing the total number of gestures by the total word count. The gesture frequency for iconic gesture category was calculated as proportions by dividing the number of iconic gestures by the total number of gestures.

Results

Speech Disfluency

For *H1* and *H2*, we used independent samples t-tests to examine disfluency rates across virtual and face-to-face settings[1]. Results showed that the overall disfluency rate was significantly higher for the participants who communicated via videoconferencing (M = .17, SD = .07) than those who communicated face-to-face (M = .12, SD = .08), t(62) = 2.33, p = .011. Moreover, we found that the proportion of filled pauses among all disfluency types was significantly higher for individuals who communicated through videoconferencing (M = .49, SD = .20) as opposed to face-to-face (M = .37, SD = .24), t(59) = 2.06, p = .022. Similarly, the proportion of repairs among all disfluency types was significantly higher in the virtual (M = .06, SD = .11) than in the face-to-face setting (M = .01, SD = .03), t(59) = 2.61, p = .006. On the other hand, the proportion of silent pauses among all disfluency types was significantly higher in the face-to-face (M = .60, SD = .21) than in the online setting (M = .44, SD = .24), t(59) = -2.74, p = .004. The proportion of repetitions, however, was comparable across the two settings, t(59) = .69, p = .247. We also applied the Benjamini-Hochberg procedure for performing multiple comparisons with different outcome variables by controlling the false discovery rate at 0.05. Benjamini-Hochberg correction supported the results reported above.[2]

For *H3*, we conducted independent samples t-tests to investigate the gesture frequencies across the virtual and face-to-face settings. Results demonstrated that the overall gesture

[1] We did not find any sex differences in overall disfluency rate, t(62) = -1.22, p = .227, or in overall gesture frequency, t(62) = 0.88, p = .383.

[2] For Benjamini-Hochberg procedure, we put all the p-values in the ascending order and assigned ranks to each of them. We calculated each individual p-value's Benjamini-Hochberg critical value by dividing the individual p-value's rank to the number of tests (i.e., comparisons) and then multiplying with the false discovery rate of 0.05. Then, we found the largest p-value that is less than or equal to its corresponding BH critical value (4th rank) and considered all p-values up to and including this point as significant.

Tests	p-values	Rank	BH critical value
Silent pauses	.004	1	.01
Repairs	.006	2	.02
Overall disfluency	.011	3	.03
Filled pauses	.022	4	.04
Repetitions	.247	5	.05

frequency ($t(62) = .64$, $p = .261$), and the proportion of iconic gestures among all gestures ($t(43) = .64$, $p = .261$) were comparable across video communication and face-to-face communication contexts.

Gesture and Speech

For *H4*, a hierarchical linear regression analysis was used to examine the total disfluency rates in spontaneous speech (see Table 1). The predictor variables were the experimental setting (virtual or face-to-face) and the proportion of iconic gestures among all gestures. Results suggested a significant model, $F(2,44) = 4.94$, $p = .012$, with an R^2 of .190. Only iconic gesture use ($\beta = -.330$, $p = .026$) significantly predicted the disfluency rate in spontaneous speech. The addition of the interaction term between the experimental setting and iconic gesture use did not improve the model. The interaction term was not significant ($\beta = -.021$, $p = .903$).

Table 1. *Regression Analysis Summary for Predicting Overall Disfluency Rate*

Predictors	β	p	ΔR^2	F-change
Step 1			.190	4.94
Experimental setting (ES)		.220	.130	
Iconic gesture use (IGU)	-.330	.026		
Step 2			.000	.015
Experimental setting (ES		219	.140	
Iconic gesture use (IGU)		-.318	.072	
ES X IGU		-.021	.903	

Note. $N = 64$

When we conducted the same regression analysis replacing the proportion of iconic gestures with the overall gesture frequency, the model was not significant, $F(2,63) = 2.99$, $p = .058$, with an R^2 of .089. The addition of the interaction term between the experimental setting and overall gesture frequency did not improve the model, and this interaction was not significant ($\beta = .091$, $p = .579$) (see Table 2).

Table 2. *Regression Analysis Summary for Predicting Overall Disfluency*

Predictors	β	p	ΔR^2	F-change
Step 1			.089	2.99
Experimental setting (ES)	.277	.028		
Gesture frequency (GF)	-.092	.454		
Step 2			.005	.312
Experimental setting (ES	.277	.028		
Gesture frequency (GF)	-.153	.356		
ES X GF	.091	.579		

Note. $N = 64$

Discussion

This study investigated multimodal language use in face-to-face and virtual communication in the context of describing daily routines. We examined individuals' rates and patterns of speech disfluency and gesture use in face-to-face and videoconferencing settings. We also investigated whether gesture use, particularly iconic gesture production, was associated with speech fluency, regardless of the experimental setting. Our findings provide support for *H1*. The overall disfluency rate was significantly higher for those who communicated via videoconferencing than those who communicated face-to-face. *H2* is also supported. The use of specific disfluency types among all disfluencies also differed across the two settings, with the proportion of filled pauses and repairs being higher in the videoconferencing setting than in the face-to-face setting. On the other hand, the proportion of silent pauses was higher in face-to-face communication than in online communication. The use of repetitions, however, was comparable across the two settings. Contrary to what we have expected in *H3*, we demonstrated that the overall gesture frequency and the proportion of iconic gestures among all gestures were similar in online and face-to-face communication. Finally, *H4* is supported. We found that the overall gesture frequency was not associated with overall speech disfluency. However, individuals' iconic gesture use negatively predicted their likelihood of being disfluent in speech, regardless of the experimental setting.

The higher overall disfluency rate observed in the online setting supports the argument that videoconferencing might have a cognitively more demanding nature compared to face-to-face communication (Cserző, 2021). Previous research has suggested that people are more disfluent when they engage with a difficult task (Bortfeld et al., 2001; Morsella & Krauss, 2004). Being present in a videocall might require individuals to invest more attention in terms of ignoring the distractors around them and ensuring that they are in the view of the camera. Moreover, participants of an online meeting might be prone to constantly check whether any connection issue interrupts the communication (Bailenson, 2021). Videoconferencing can also be more exhausting and cognitively demanding compared to face-to-face communication due to the overload of nonverbal cues. Communication in videoconferences requires constant monitoring of nonverbal signals when both sending and receiving, such as maintaining eye contact with multiple interlocutors simultaneously (Bailenson, 2021). All these factors might create a load on the speech production system, paving the way for higher disfluency rates.

Another factor that influences the speech production process might be related to individuals being exposed to their own video feed. In videoconferencing, when participants open their camera, they are exposed to their own visual displays. Research demonstrated that team members communicated and performed less effectively when they were exposed to their own video feed than when they were not (Hassell & Cotton, 2017). This is because when individuals see their own video feed, their objective self-awareness increases and as a result, they perform worse (Geller & Shaver, 1976; Liebling & Shaver, 1973; Xu & Behring, 2014). Participants' analyzing themselves might create an extra cognitive load (Hassell & Cotton, 2017), which might in turn affect the speech production process, particularly in the form of fluency.

As we expected, filled pause use was more prominent in the online communication and silent pause use was more prominent in face-to-face communication. The higher proportion of filled pauses observed in the videoconferencing setting suggests that filled pauses did not solely result from the planning load. There might also be communicative motivations behind the production of filled pauses (Corley & Stewart, 2008; Fraundorf & Watson, 2014). That is, using filled pauses in online communication might help speakers maintain the conversational floor while signaling that they are present at the meeting in the case of any connection issue. On the other hand, the use of silent pauses being more prominent in the face-to-face setting might be an indicator of the natural flow observed in face-to-face communication. Communicating face-to-face might not require individuals to frequently emphasize their presence as they are physically present in front of their listener(s). Therefore, when it comes to achieving successful communication, using silent pauses might be more acceptable in face-to-face than in online contexts.

The proportion of repairs among all disfluencies was also more prominent for those who communicated via videoconferencing. Repairs may occur when speech planning is not well-executed (Arslan & Göksun, 2022; Bock, 1996). One might suggest that the increased cognitive load in video communication might result in not allocating enough cognitive resources for the speech planning process. Therefore, predicting and preventing a possible error in speech might be less possible in videoconferencing, as individuals are already busy with reading and producing communicative signals to maintain a natural flow in online communication.

Unlike we expected, the overall gesture frequency and the proportion of iconic gestures were comparable across the two settings. Individuals produce more gestures, particularly iconic gestures in a spatial context (Alibali, 2005; Arslan & Göksun, 2021), or when task difficulty increases (Kita & Davies, 2009). Then, a task where individuals described their daily routines might not have prompted hand use in a way that can reveal differences in gesture production across the two contexts. Further research is needed to understand whether the mode of communication impacts gesture production in a spatial context.

Regardless of the experimental setting, we found a significant negative association of speech disfluency with iconic gesture use, but not with overall gesture frequency. This finding is in line with gestures' self-oriented functions (Kita et al., 2017), suggesting that using iconic gestures might facilitate cognitive processes, paving the way for a more fluent speech. Although it is only an indirect evidence of iconic gestures' facilitative roles in speech, observing such a link in a daily routine description context raises questions. Previous research targeting gestures' role in lexical access mainly focused on a spatial context (Morsella & Krauss, 2004), in line with gestures' close link with spatial cognition (Alibali, 2005). As being closely associated mechanisms (Kita & Özyürek, 2003), gesture and speech might communicate regardless of the context. Then, gestures' self-oriented functions might be reflected in speech in a less spatial context as well. Understanding whether manipulating to what extent a task is spatial alters the facilitative roles of gestures in speech warrants further investigation.

We also acknowledge previous studies showing no relationship between speech (dis)fluencies and representational gesture use (Arslan & Göksun, 2022; Arslan et al., 2024;

Graziano & Gullberg, 2018; Kısa et al., 2022; Ünal et al.,2022). For example, Ünal and colleagues (2022) examined the co-occurrence of gestures and speech disfluencies, finding that gestures were equally likely to co-occur with disfluent and fluent speech. Likewise, studies also showed that preventing individuals from gesturing may not necessarily increase their disfluency rates compared with spontaneous gesturing (Avcı et al., 2022), even for literal or metaphorical spatial content (Kısa et al., 2022). Although these findings seem to contrast with our findings, it is important to note that our methodology did not include any coding of gesture-disfluency co-occurrences or manipulation of gesture use. Instead, our analyses indicate a relationship between the overall representational gesture use and speech disfluencies in spontaneous speech. We suggest that using representational gestures might indirectly pave the way for a more fluent speech overall by decreasing the cognitive load.

This is among the first studies to target multimodal language production across face-to-face and virtual settings. There is an increased tendency to use online data collection methods in research due to the COVID-19 pandemic. It is important to understand to what extent the findings obtained from online studies can be compared with the findings of face-to-face studies, particularly in language research on which the modes of communication can have a crucial effect. Our results indicate a difference between the two settings in speech disfluency, but not in gesture production. Using different tasks and contexts is required to observe whether multimodal language differs between face-to-face and online communication.

Acknowledgement: We thank Busenur Caba and Zeynep Acar for their contribution to transcription and coding.

Ethics Committee Approval: Approval was received from the Koç University Ethics Committee (Date 07.11.2018, Decision No: 2018.276.IRB3.195).

Informed Consent: Informed consent was obtained from all participants for the study.

Peer-review: Externally peer-reviewed

Author Contributions: Conception/Design of Study- B.A., C.A., D.Ö.; Data Acquisition- B.A., C.A.; Data Analysis/Interpretation- B.A., C.A., D.Ö.; Drafting Manuscript- B.A., C.A.; Critical Revision of Manuscript- B.A., D.Ö.; Final Approval and Accountability- B.A., C.A., D.Ö.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declare that this study has received no financial support.

References / Kaynakça

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation, 5*(4), 307-331. https://doi.org/10.1207/s15427633scc0504_2

Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., & Lawless, M. (2019). Using zoom videoconferencing for qualitative data collection: Perceptions and experiences of researchers and participants. *International Journal of Qualitative Methods, 18*, 1-8. https://doi.org/10.1177/1609406919874596

Arslan, B., & Göksun, T. (2021). Ageing, working memory, and mental imagery: Understanding gestural communication in younger and older adults. *Quarterly Journal of Experimental Psychology, 74*(1), 29-44. https://doi.org/10.1177/1747021820944696

Arslan, B., & Göksun, T. (2022). Aging, gesture production, and disfluency in speech: A comparison of younger and older adults. *Cognitive Science, 46*(2), Article e13098. https://doi.org/10.1111/cogs.13098

Arslan, B., Avcı, C., Yılmaztekin, A., & Göksun, T. (2024). Do bilingual adults gesture when they are disfluent?: Understanding gesture-speech interaction across first and second languages. *Language, Cognition and Neuroscience, 39*(5), 571–583. https://doi.org/10.1080/23273798.2024.2345306

Avcı, C., Arslan, B., & Göksun, T. (2022). Gesture and speech disfluency in narrative context: Disfluency rates in spontaneous, restricted, and encouraged gesture conditions. In Culbertson J., Perfors A., Rabagliati H., & Ramenzoni V. (Eds.), *Proceedings of the 44th annual conference of the cognitive science society* (pp. 1912–1917). Cognitive Science Society.

Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior, 2*(1), 1-6. https://doi.org/10.1037/tmb0000030

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*, 395–421. https://doi.org/10.3758/BF03214545

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123-147. https://doi.org/10.1177/00238309010440020101

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass, 2*(4), 589-602. https://doi.org/10.1111/j.1749-818X.2008.00068.x

Cserző, D. (2021). Discourses and practices of attention in video chat. *Multimodal Communication, 10*(2), 143-156. https://doi.org/10.1515/mc-2020-0010

Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience, 29*(9), 1083-1096. https://doi.org/10.1080/01690965.2013.832785

Geller, V., & Shaver, P. (1976). Cognitive consequences of self-awareness. *Journal of Experimental Social Psychology, 12*(1), 99-108. https://doi.org/10.1016/0022-1031(76)90089-5

Glassmeyer, D. M., & Dibbs, R. A. (2012). Researching from a distance: Using live web conferencing to mediate data collection. *International Journal of Qualitative Methods, 11*(3), 292-302. https://doi.org/10.1177/160940691201100308

Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology, 9*, Article e00879. https://doi.org/10.3389/fpsyg.2018.00879

Hassell, M. D., & Cotton, J. L. (2017). Some things are better left unseen: Toward more effective communication and team performance in video-mediated interactions. *Computers in Human Behavior, 73*, 200-208. https://doi.org/10.1016/j.chb.2017.03.039

Hyusein G., & Göksun, T. (2023). The creative interplay between hand gestures, convergent thinking, and mental imagery. *PLOS ONE, 18*(4), Article e0283859. https://doi.org/10.1371/journal.pone.0283859

Kandemir, S., Özer, D., & Aktan-Erciyes, A. (2023). Multimodal language in child-directed versus adult-directed speech. *Quarterly Journal of Experimental Psychology, 77*(4), 716-728. https://doi.org/10.1177/17470218231188832

Kısa, Y. D., Goldin-Meadow, S., & Casasanto, D. (2022). Do gestures really facilitate speech production?. *Journal of Experimental Psychology: General, 151*(6), 1252-1271. https://doi.org/10.1037/xge0001135

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review, 124*(3), 245-266. https://doi.org/10.1037/rev0000059

Kita, S., & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes, 24*(5), 761-775. https://doi.org/10.1080/01690960802327971

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

Krauss, R., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 261–283). Cambridge University Press.

Lausberg, H., & Sloetjes, H. (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

Liebling, B. A., & Shaver, P. (1973). Evaluation, self-awareness, and task performance. *Journal of Experimental Social Psychology, 9*(4), 297-306. https://doi.org/10.1016/0022-1031(73)90067-X

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*(1), 19-44. https://doi.org/10.1080/00437956.1959.11659682

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes, 22*(4), 473-500. https://doi.org/10.1080/01690960600696916

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, 411-424. https://doi.org/10.2307/4149008

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language, 9*(1), 19-36. doi:10.1006/csla.1995.0002

Özder, L. E., Özer D., & Göksun, T. (2022). Gesture use in L1-Turkish and L2-English: Evidence from emotional narrative retellings. *Quarterly Journal of Experimental Psychology, 76*(8), 1797–1816. https://doi.org/10.1177/17470218221126685

Özer, D., Tansan, M., Özer, E. E., Malykhina, K., Chatterjee, A., & Göksun, T. (2017). The effects of gesture restriction on spatial language in young and elderly adults. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1471-1476). Cognitive Science Society.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*(4), 226-231. https://doi.org/10.1111/j.1467-9280.1996.tb00364.x

Seyfeddinipur, M. & Kita, S. (2001, August 29-31). *Gesture as an indicator of early error detection in self-monitoring of speech*. ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech, Edinburgh, Scotland, UK.

Silverman, S. W., & Ratner, N. B. (1997). Syntactic complexity, fluency, and accuracy of sentence imitation in adolescents. *Journal of Speech, Language, and Hearing Research, 40*(1), 95-106. https://doi.org/10.1044/jslhr.4001.95

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*(1), 25-38. https://doi.org/10.1006/jmla.1993.1002

Ter Bekke, M., Drijvers, L., & Holler, J. (2024). Hand gestures have predictive potential during conversation: An investigation of the timing of gestures in relation to speech. *Cognitive Science, 48*(1), Article e13407. https://doi.org/10.1111/cogs.13407

Torrentira, M. C., Jr. (2020). Online data collection as adaption in conducting quantitative and qualitative research during the COVID-19 pandemic. *European Journal of Education Studies, 7*(11), 78-87. http://dx.doi.org/10.46827/ejes.v7i11.3336

Ünal, E., Manhardt, F., & Özyürek, A. (2022). Speaking and gesturing guide event perception during message conceptualization: Evidence from eye movements. *Cognition, 225*, Article e105127. https://doi.org/10.1016/j.cognition.2022.105127

Xu, Q., & Behring, D. (2014). The richer, the Better? Effects of modality on intercultural virtual collaboration. *International Journal of Communication, 8*, 2733-2754.

Zoom Video Communications Inc. (2016). *Security guide. Zoom Video Communications Inc*. Retrieved from https://d24cgw3uvb9a9h.cloudfront.net/static/81625/doc/Zoom-Security-White-Paper.pdf

How cite this article / Atıf Biçimi

Arslan, B., Avcı, C., & Özer. (2024). Multimodal communication in virtual and face-to-face settings: Gesture production and speech disfluency. *Psikoloji Çalışmaları - Studies in Psychology, 44*(3), 349–363. https://doi.org/10.26650/SP2024-1443983