Disfluency in Spontaneous Speech (DiSS) Workshop 2023
28-30 August 2023, Bielefeld, Germany

The interplay between speech fluency and gesture in L1 Finnish and L2 English task-based interactions

Pauliina Peltonen[1], *Loulou Kosmala*[2], *Sandra Götz*[3] & *Pekka Lintunen*[4]

[1]University of Turku, Finland & Philipps University Marburg, Germany
[2]Université Paris Nanterre, France
[3]Philipps University Marburg
[4]University of Turku

paupelt@utu.fi, lkosmala@parisnanterre.fr, goetz-lehmann@uni-marburg.de, pekka.lintunen@utu.fi

Abstract

In second language (L2) fluency research, (dis)fluencies are often examined from monologue speech without reference to the speakers' visual-gestural behavior or first language (L1) data. However, research on L1 and L2 gestures has revealed an intricate relationship between gestures and (dis)fluency, underscoring the need for more research on the topic. Extending these two lines of research, this pilot study examines the interplay between speech fluency and gesture in L1 Finnish and L2 English task-based interactions (N = 8). The data were collected as part of a larger project. The fluency-related features (fluencemes) examined in the present study included silent and filled pauses, prolongations, and four repair features. The functional gesture types included referential (deictic and representational) and pragmatic (discursive, interactive, and thinking) gestures. The gestures were also coded for the degree of overlap (full, partial, or no overlap) with fluencemes. Our results showed that the gestures and fluencemes were more frequently produced in the L2 than in the L1, but the distributions of gesture and fluenceme types were relatively similar across L1 and L2 interactions. The majority of gestures did not overlap with fluencemes, corroborating previous findings. The findings have implications for L2 speech fluency and gesture research.

Index Terms: fluency, task-based interaction, second language acquisition, gesture

1. L2 speech fluency

When speaking in an L2, planning pressure is very high [1]. To overcome these planning phases, learners have been found to use different kinds of strategies (or "fluencemes" [2]), which have traditionally been considered as disfluencies despite being multifunctional. While L2 speech fluency research has shown that silent pauses (SPs) are linked with L2 proficiency, less proficient speakers producing more SPs than more advanced speakers, other fluencemes, such as filled pauses (FPs; e.g., *er*, *um* in English; *öö* in Finnish), certain types of repairs (e.g., self-repetitions and corrections), and prolongations (i.e., sound lengthenings) tend to be more prone to individual variation and less clearly linked with L2 proficiency (e.g., [3, 4]).

Studies investigating the effect of task types on fluenceme use find a general trend towards more fluent speech in tasks with a tighter narrative structure (e.g., [5]), although findings vary based on the investigated fluencemes and task types. Previous learner corpus-based research, for example, has revealed significantly less fluent learner performances for a picture story retell task compared to a monologic and an interview task (cf. [6], focusing on speech rate [SR]). Furthermore, task-based research comparing L2 speech in monologic and dialogic conditions has tended to show higher fluency in dialogue (e.g., [7–9]), but these studies have explored other fluencemes besides SPs only to a relatively limited extent.

Apart from taking the task type into consideration, researchers have also started to acknowledge the influence of first language (L1) speaking behavior on speakers' L2 performance. In L2 fluency research, contrastive studies have been rather rare, possibly due to scarcity of parallel corpora data, and most studies have examined monologic rather than dialogic speech data. However, a few recent studies comparing L1 and L2 speech from the same participants have revealed clear connections between L1 and L2 speech fluency especially for temporal fluency measures (e.g., [10–12]). These studies have also tended to show more fluent performance in the L1 compared to the L2 but have rarely included advanced level L2 speakers. In a previous study conducted in the Finnish context, Peltonen [13] examined L1 Finnish and L2 English fluency among intermediate level learners at two school levels. She found that most of the L2 temporal aspects correlated with comparable L1 measures (e.g., SR, mean length of run, FPs, SPs per minute) and qualitative analyses revealed further connections in the use of certain fluencemes for individual speakers. Another recent comparative study is Larsson Aas and Rørvik's [14] research on the use of repeats by Norwegian learners of English in interactive speech production. Drawing on parallel interviews of six Norwegian "proficient users" of English (CEFR C1 to C1+), the study investigated the use of repeats in L1 Norwegian and L2 English. Their qualitative analyses revealed similar repeat behavior in both languages even though they report on high inter-speaker variation.

One of the few studies investigating correlations between L1 and L2 fluency at different task types is Jiránková et al.'s [15] research on SR in L1 Czech vs. L2 English. Their study investigated 31 speakers of a parallel corpus across three tasks (a monologue, a dialogue, and a story reconstruction) in both their L1 and their L2. For all three tasks, the results revealed significant correlations between the L1 and the L2 SRs.

10.21437/DiSS.2023-2

2. Fluency and gesture

Even though relatively few studies have tackled the relationship between fluency and gesture in L1 or L2 production, this topic is gaining more and more attention (read [16] for a full review) in different languages (e.g., English, French, Dutch, Italian, Spanish, German, and Turkish [17–19]). Notably, Graziano and Gullberg's [20] work on gesture and disfluency in oral narratives among different groups of speakers (L1 speakers, adult and child L2 learners in Dutch and Italian) revealed two major trends: (1) speakers rarely produce fluencemes during gestures, and (2) these co-occurring gestures can serve both referential and pragmatic functions. While only a small fraction of gestures tend to co-occur with fluencemes overall, their relationship should not be disregarded, as it can be informative of the various cognitive and interactive processes associated with fluency mechanisms. Building on the work of Graziano and Gullberg [20], Kosmala [21] recently argued in favor of this fluency-gesture relationship by examining the rate of fluencemes and gestures in L1 and L2 speakers of English and French in tandem interactions. Her findings corroborated Graziano and Gullberg's [20] results, which further addresses the need to explore this relationship in other comparative crosslinguistic investigations. Other studies have focused on the co-occurrence of gestures and specific fluencemes, such as repairs [19], cut-offs [18], and FPs [22], but without systematically comparing the types of fluencemes that may co-occur with gestures (but see [23] for a study with L1 speakers).

When it comes to differences between L1 and L2 in gesture use, studies [24, 25] have shown that L2 learners are likely to produce more gestures in their L2 than in their L1 for cognitive purposes to overcome language difficulties in their target language. Gesture use has thus often been related to cognitive load as a facilitative or compensatory process (e.g., [26, 27]). However, this theory has also been disputed in other works [20, 21]. In addition, fluenceme and gesture production may be influenced by situational factors (such as task type or setting; e.g., [22]). These aspects should not be overlooked, given the different types of procedures used to collect data in the studies mentioned above, ranging from experimental to corpus-based approaches, elicited narratives, task-based interactions, and tandem interactions, among others. Along with situational factors, gesture use might also be influenced by the participants' L2 proficiency level, potentially interacting with individual differences (e.g., [28, 29]).

Building on this line of research, the aim of our descriptive pilot study is to shed light on the interplay between speech fluency and gesture in L1 Finnish and L2 English task-based interactions among advanced learners. We focus, in particular, on the degree of temporal overlap between fluencemes and gestures. Our study addressed the following research questions: 1) To what extent does advanced English learners' *overall* gesture and fluenceme production differ across L1 and L2 interactions? 2) What kinds of differences can be observed in the distribution of *specific* gesture and fluenceme types across L1 and L2 interaction? 3) To what extent do gestures overlap with fluencemes in L1 and L2 interaction?

3. Data and method

The participants (N = 8, four pairs) were selected randomly from a larger pool of *Fluency and Disfluency Features in L2 Speech* (FDF2) project participants. The participants were L1 Finnish first year English majors at a university in Finland. Three of the participants were female, four male, and one did not want to report their gender. Their average age was 19.38 (SD = 1.77). Based on the participants' LexTale [30] scores in English (mean 85.78%), six represented the C1/C2 (scores over 80%) level in the Common European Framework of Reference for Languages [31] and two represented the B2 level (scores 60–80%). Thus, the participants can be characterized as advanced English speakers. None of the participants reported language-related impairments.

Of the complete data set collected for the project, the dialogue tasks were used in the present study. The participants completed the dialogue tasks in their L1 Finnish and L2 English with the same pair based on comparable problem-solving task prompts. The order of the tasks and the languages was counterbalanced. The participants were given two minutes of individual preparation time for both tasks, followed by six minutes of time for completing the task together. The task performances were audio and video recorded by research assistants. The research assistants announced when 5 minutes had passed but did not otherwise intervene in the discussion. In both tasks, the participants received pictures of sixteen items (printed on an A4 paper) that they were asked to rank in the order of usefulness for a) survival on a desert island (see [3]) or b) walking to the mothership after a crash landing of their spaceship on the lighted side of the moon. The task instructions also specified that the participants should justify their choices and agree on the order of importance for all of the items.

Initial transcriptions of the task performances were produced and cross-checked by two MA level students of English. The other stages of the speech fluency analysis were completed in Praat: first a Praat script [32] was used to distinguish speaking time and SPs of at least 0.25 seconds (an upper limit for SPs was set manually at 3 seconds following [9] as long silences are likely to reflect other factors than individual disfluencies). Speaking time was manually allocated to each speaker, and SPs were coded manually as within-turn pauses ("belonging" to an individual speaker) or between-turn pauses. Only within-turn pauses were included as SPs in the analysis for the present paper. Along with pauses occurring between speakers' turns, pauses over 2 seconds were regarded as shared and included in between-turn pauses, thus being excluded from the present study. In addition to SPs, six fluenceme types were manually annotated, including four aspects of repair (false starts, repetitions, replacements, and reformulations), (non-lexicalised) FPs, and prolongations. The features have been widely studied in previous L1 and L2 speech fluency research. The categorization for repair was adopted from Foster and Skehan [33]: false starts included cut-off words or longer utterances, repetitions repeated words without further modifications, replacements one-word substitutions of another word, and reformulations longer than one-word modifications.

All manual gestures were annotated in the data using a functional typology following previous work [34–36]. Our annotation scheme (based on [16]) included two broad categories, and five subcategories. The two higher-level categories were (1) referential (gestures related to discourse content) and (2) pragmatic (gestures related to discourse itself), with five subtypes (see Table 1). A subset of the Finnish data was jointly annotated for gestures by the first and second authors to increase the reliability of the annotation. The degree of temporal overlap between fluencemes and gestures was also annotated, distinguishing between *full overlap* (the stroke of the gesture is fully aligned with the fluenceme), *partial overlap* (only parts of the gesture; i.e., preparation, hold, or retraction

are aligned with the fluenceme), and *no overlap* (the gesture is not produced during the fluenceme). The temporal annotations were conducted with the ELAN software [37], which enables the study of temporal alignment across gesture and speech.

Table 1: *Gesture types and subtypes*.

Referential gestures	Pragmatic gestures
Deictic: used to point towards objects, persons, locations, time etc.	**Discursive**: used for discourse and segmentation purposes (emphasis, contrast etc.)
Representational: used to metaphorically or iconically depict a referent or an action through means of representation (size, shape, movement, mode etc.)	**Interactive**: used to regulate interaction (turn-taking, holding etc.) indicate a stance or a speech act.
	Thinking: used to indicate an ongoing search

To answer our RQ 1, we calculated descriptive statistics (means, standard deviations) and computed paired t-tests (difference in mean number of gestures/fluencemes per minute of individual total time[1]). Despite the small data set, Shapiro-Wilk's tests of normality indicated that the variables were normally distributed ($p > .05$). For RQ 2, we report the proportions and totals of gesture and fluenceme subtypes in L1 and L2 production. For RQ 3, we focus on the overlaps between gestures and fluencemes, including their subtypes, based on proportions and total frequencies. For RQs 2 and 3, descriptive statistics were preferred over inferential statistics due to low frequencies of features in some fluenceme/gesture categories.

4. Results

Regarding RQ1, the participants produced a total of 255 gestures in L1 Finnish and 407 gestures in L2 English. The average number of gestures was 13.10 per minute of individual total time in the L1 (SD = 10.35) and 19.27 in the L2 (SD = 12.84). According to the paired t-tests, this difference in the average number of gestures between the L1 and L2 interactions was statistically significant, $t(7) = -2.992$, $p = .020$. Regarding fluencemes, the participants produced a total of 392 fluencemes in their L1 and 573 fluencemes in their L2. The average number of fluencemes was 18.93 per minute of individual total time in the L1 and 25.93 in the L2. According to the paired t-tests, this difference in means between the L1 and L2 interactions was statistically significant, $t(7) = -4.505$, $p = .003$.

Regarding RQ 2, the participants exhibited a very similar gesture distribution in their L1 and L2, with a large proportion of deictic gestures, followed by discursive ones (see Table 2). The three other gesture types occurred more rarely in the data.

Table 2: *Proportions and frequencies of gestures*.

Gesture subtype	L1 Proportion (total)	L2 Proportion (total)
Deictic	54% (138)	51% (209)
Representational	9% (23)	5% (22)
Discursive	25% (65)	25% (102)
Interactive	11% (29)	14% (55)
Thinking	0% (0)	5% (19)

[1] Individual total time = total speaking time per participant + total SP time per participant.

Regarding fluencemes (see Table 3), SPs were the most frequent fluenceme type in both L1 and L2 interactions, followed by false starts in L1 interaction and FPs in L2 interaction. The proportions for the other fluenceme types were relatively low in both L1 and L2 interactions.

Table 3: *Proportions and frequencies of fluencemes*.

Fluenceme subtype	L1 Proportion (total)	L2 Proportion (total)
Silent pauses (SP)	66% (259)	59% (339)
Filled pauses (FP)	7% (27)	13% (72)
Prolongations (PR)	3% (7)	6% (36)
False starts (FS)	17% (66)	11% (64)
Repetitions (REP)	5% (20)	8% (47)
Replacements (REPL)	1% (4)	1% (8)
Reformulations (REF)	2% (9)	1% (7)

Regarding RQ 3, the majority of the gestures did not overlap with fluencemes in L1 or L2 interactions (see Table 4). Overall, only 11% of the gestures fully overlapped with gestures, with 18% partially overlapping. L2 interactions contained proportionally slightly more full or partial overlaps than L1 interactions, potentially due to higher numbers of gestures and fluencemes overall in the L2 interactions.

Table 4: *Overlap between gestures and fluencemes*.

Degree of overlap	L1 proportion (total)	L2 proportion (total)	Overall proportion (total)
Full overlap	9% (36)	12% (71)	11% (107)
Partial overlap	16% (61)	20% (112)	18% (173)
No overlap	75% (295)	68% (390)	71% (685)

The overlaps between fluenceme subtypes and overlapping gestures in Table 5 show that false starts and SPs were the most frequently co-occurring fluencemes with gestures in both L1 and L2 interactions. FPs and PRs co-occurred with gestures to a limited extent, and FPs did not fully overlap with gestures at all in the L1 productions. Apart from the previously mentioned false starts, other repair types co-occurred with gestures very infrequently in the L1 and L2 interactions.

Table 5: *Full / partial overlap with fluenceme types*.

	L1		L2	
Fluenceme	Full overlap	Partial overlap	Full overlap	Partial overlap
SP	24% (23)	46% (45)	20% (36)	43% (78)
FP	0%	2% (2)	2% (3)	5% (10)
PR	1% (1)	0% (0)	2% (3)	3% (5)
FS	7% (7)	10% (10)	11% (20)	5% (10)
REP	2% (2)	2% (2)	2% (3)	3% (6)
REPL	0% (0)	1% (1)	2% (4)	1% (2)
REF	3% (3)	1% (1)	1% (2)	1% (1)

In L1 Finnish interactions (see Table 6), false starts and SPs jointly accounted for 88% of fluencemes overlapping with gestures, more specifically deictic and discursive gestures. This

is in line with our finding of the most frequent fluenceme subtypes in the L1 being false starts and SPs, and the most frequent gesture subtypes being deictic and discursive ones.

Table 6: *Partial and full overlap between gesture subtypes and fluencemes in L1 Finnish interactions.*

Fluen-ceme	Deictic	Dis-cursive	Inter-active	Represen-tational	Total
SP	34% (33)	24% (23)	10% (10)	2% (2)	70% (68)
FP	0% (0)	1% (1)	1% (1)	0% (0)	2% (2)
PR	0% (0)	0% (0)	1% (1)	0% (0)	1% (1)
FS	9% (9)	6% (6)	1% (1)	1% (1)	18% (17)
REP	2% (2)	1% (1)	1% (1)	0% (0)	4% (4)
REPL	1% (1)	0% (0)	0% (0)	0% (0)	1% (1)
REF	4% (4)	0% (0)	0% (0)	0% (0)	4% (4)
Total	51% (49)	32% (31)	14% (14)	3% (3)	100% (97)

The results were fairly similar for the L2 English interactions (see Table 7): again, the majority of the overlapping gestures co-occurred with SPs, followed by false starts. Furthermore, the overlaps tended to occur especially with deictic and discursive gestures, other subtypes overlapping with fluencemes more rarely.

Table 7: *Partial and full overlap between gesture subtypes and fluencemes in L2 English interactions.*

Fluen-ceme	Deictic	Dis-cursive	Inter-active	Represen-tational	Think-ing	Total
SP	31% (56)	17% (31)	5% (9)	5% (9)	5% (9)	62% (114)
FP	4% (8)	1% (2)	1% (1)	1% (1)	1% (1)	7% (13)
PR	2% (3)	1% (1)	1% (2)	1% (2)	0% (0)	4% (8)
FS	6% (11)	5% (9)	4% (7)	1% (2)	1% (1)	16% (30)
REP	2% (3)	1% (2)	1% (1)	1% (1)	1% (2)	5% (9)
REPL	2% (3)	0% (0)	1% (2)	1% (1)	0% (0)	3% (6)
REF	1% (2)	1% (1)	0% (0)	0% (0)	0% (0)	2% (3)
Total	47% (86)	25% (46)	12% (22)	9% (16)	7% (13)	100% (183)

5. Discussion and conclusion

The present pilot study explored the interplay between speech fluency and gesture in L1 Finnish and L2 English task-based interactions. Regarding RQ 1, the results showed that the participants produced overall more gestures and fluencemes in the L2 than the L1 interactions. Since the participants were advanced learners of English, the result is slightly surprising, although not entirely unexpected based on previous fluency research [10, 13]. However, these studies have often examined intermediate rather than advanced level learners. Furthermore, the more frequent use of gestures in the L2 is in line with some of the previous gesture studies ([21, 24, 25]; but see [18]). Rather than being directly linked with L2 proficiency or resulting from a frequent use of compensatory gestures, the finding may be indicative of a more visible "expressive" trait in the L2. To confirm this, further analyses on individual profiles across the speakers' L1 and L2 gestures are needed.

Regarding RQ 2, the distribution of gesture types across L1 and L2 interactions was found to be relatively similar, deictic gestures being the most common type, followed by discursive gestures. The similarity in the distributions can be potentially explained with reference to the task-based nature of the interactions and the participants' advanced level of L2 proficiency. As the instructions of the tasks specified that the participants should discuss sixteen items and rank them in the order of usefulness, with task materials including pictures of the items printed on A4 sheets to facilitate the interaction, the deictic gestures most often involved pointing to specific items on the sheet. Conversely, in spontaneous interactions, interactants are more prompted to produce pragmatic gestures to manage turn-taking, cite the other's contribution, or provide metalinguistic comments, among other practices, which may have an effect on the L2 (see [20] and [21] who found more pragmatic gestures in L2 than in L1). Regarding the distribution of fluencemes, they were found to be relatively similar across L1 and L2 interactions as well, again being potentially linked with the task-based nature of the interactions and advanced level of L2 proficiency. SPs were found to be the most frequent type in both L1 and L2 interactions, which is in line with previous L2 fluency research (e.g., [3, 4, 8]). In L2 interactions, FPs were the second most common type, while false starts were the second most common type in the L1 interactions. The proportions for the other categories were relatively low. The fact that FPs formed a larger proportion of fluencemes in the L2 than in the L1 interactions, along with a similar tendency for prolongations and repetitions, could potentially be related to strategic uses of these features (e.g., [2, 3]), but would require further analyses of individual speakers' profiles to confirm this.

Regarding RQ 3, the majority of gestures were not found to overlap (either fully or partially) with fluencemes in the L1 or L2 interactions. This finding is consistent with previous research [17, 20, 21] suggesting that the majority of gestures tend to co-occur with fluent speech rather than SPs or other disfluencies. Our findings complement previous research by examining the types of fluencemes and gestures occurring during overlaps, as previous studies have rarely incorporated a range of fluency-related features or analyzed the specific gesture types during overlaps. Thus, while it is overall rare that gestures overlap with fluencemes, our analysis shows that when this happens, certain fluenceme types (SPs) and gesture types (deictic and discursive) tend to co-occur more often than others. This finding reflects the distributions of fluencemes and gesture types in the data overall (RQ 2).

The present study has provided insights into an underresearched area related to the interplay between fluency and gesture in L1 and L2 task-based interactions. Our focus in this pilot study was on descriptive overall tendencies, which can provide a starting point for further qualitative and quantitative research into individual differences and links between L1 and L2 fluency and gestures among individual participants, as L1 fluency tends to be connected with L2 fluency [10–13] and rates of gesturing can vary across individual speakers [16, 28, 38]. Due to the small sample size, the tendencies reported here should be regarded as tentative, and ideally confirmed with future studies involving larger sample sizes and more advanced statistical techniques (e.g., regression). Methodologically, integrating speech (dis)fluency and gesture analysis should be more widely adopted, as it can help us achieve a fuller understanding of the nature of L1 and L2 interaction.

6. Acknowledgments

This study received support from the Research Council of Finland (project *Fluency and Disfluency Features in L2 Speech*, decision number 331903).

7. References

[1] J. Kormos, *Speech Production and Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, 2006.

[2] S. Götz, *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins, 2013.

[3] P. Peltonen, "Individual and interactional speech fluency in L2 English from a problem-solving perspective: A mixed-methods approach," Ph.D. dissertation, University of Turku, Turku, Finland, 2020.

[4] P. Tavakoli, F. Nakatsuhara, and A.-M. Hunter, "Aspects of fluency across assessed levels of speaking proficiency," *The Modern Language Journal*, vol. 104, no. 1, pp. 169–191, Spring 2020.

[5] P. Tavakoli and P. Foster, "Task design and second language performance: The effect of narrative type on learner output," *Language Learning*, vol. 58, no. 2, pp. 439–473, Jun. 2008.

[6] T. Gráf, "Speech rate revisited: The effect of task design on speech rate," in *Learner Corpora and Language Teaching*, S. Götz and J. Mukherjee, Eds., Amsterdam: John Benjamins, 2019, pp. 175–189.

[7] M. C. Michel, "Effects of task complexity and interaction on L2-performance," in S*econd Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*, P. Robinson, Ed., Amsterdam: John Benjamins, 2011, pp. 141−174.

[8] P. Tavakoli, "Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency," *International Review of Applied Linguistics in Language Teaching*, vol. 54, no. 2, pp. 133–150, Jun. 2016.

[9] G. Witton-Davies, "The study of fluency and its development in monologue and dialogue," Ph.D. dissertation, Lancaster University, Lancaster, U.K., 2014.

[10] N. H. De Jong, R. Groenhout, R. Schoonen, and J. H. Hulstijn, "Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior," *Applied Psycholinguistics*, vol. 36, no. 2, pp. 223–243, Mar. 2015.

[11] A. Huensch and N. Tracy-Ventura, "Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time", *Applied Psycholinguistics*, vol. 38, no. 4, pp. 755–785, Jul. 2017.

[12] Z. Duran-Karaoz and P. Tavakoli, "Predicting L2 fluency from L1 fluency behavior: The case of L1 Turkish and L2 English speakers," S*tudies in Second Language Acquisition*, vol. 42, no. 4, pp. 671–695, Sep. 2020.

[13] P. Peltonen, "Exploring connections between first and second language fluency: A mixed methods approach," *The Modern Language Journal*, vol. 102, no. 4, pp. 676–692, Winter 2018.

[14] H. Larsson Aas and S. Rørvik, "Exploring the use of repeats in learners' native and interlanguage production," in *Complexity, Accuracy and Fluency in Learner Corpus Research*, A. Leńko-Szymańska and S. Götz, Eds., Amsterdam: John Benjamins, 2022, pp. 299–324.

[15] L. Jiránková, T. Gráf, and A. Kvítková, "On the relation between L1 and L2 speech rate," in *Widening the Scope of Learner Corpus Research. Selected Papers from the Fourth Learner Corpus Research Conference*, A. Abel, A. Glaznieks, V. Lyding, and L. Nicolas, Eds., Louvain-la-neuve: Presses universitaires de Louvain, 2019, pp. 19–41.

[16] L. Kosmala, "A multimodal contrastive study of (dis)fluency across languages and settings: Towards a multidimensional scale of inter-(dis)fluency," Ph.D. dissertation, Sorbonne Nouvelle, Paris, France, 2021.

[17] B. Arslan and T. Göksun, "Understanding multimodal communication: Gesture production and disfluency in speech," in *The 7th Gesture and Speech in Interaction (GESPIN2020) Conference*, Stockholm, Sweden, 2020.

[18] R. Lopez-Ozieblo, "Cut-offs and co-occurring gestures: Similarities between speakers' first and second languages," *International Review of Applied Linguistics in Language Teaching*, vol. 60, no. 3, pp. 647–677, Aug. 2022.

[19] M. Seyfeddinipur, "Disfluency: Interrupting speech and gesture," Ph.D. dissertation, Radboud University, Nijmegen, The Netherlands, 2006.

[20] M. Graziano and M. Gullberg, "When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons," *Frontiers in Psychology*, vol. 9, article 879, Jun. 2018.

[21] L. Kosmala, "Gestures in fluent and disfluent cycles of speech: What they may tell us about the role of (dis)fluency in L2 discourse," in *Proceedings of the 10th Workshop on Disfluency in Spontaneous Speech (DiSS 2021)*, St. Denis, France, 2021, pp. 77–82.

[22] L. Kosmala, "Exploring the status of filled pauses as pragmatic markers: The role of gaze and gesture," *Pragmatics & Cognition*, vol. 29, no. 2, pp. 272–296, Dec. 2022.

[23] S. Betz, N. Bryhadyr, O. Türk, O., and P. Wagner, "Cognitive load increases spoken and gestural hesitation frequency", *Languages*, vol. 8, no. 1, article 71. Mar. 2023.

[24] M. Gullberg, *Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish*. Lund: Lund University Press, 1998.

[25] G. Stam, "Thinking for speaking about motion: L1 and L2 speech and gesture," *International Review of Applied Linguistics*, vol. 44, no. 2, pp. 145–171, Aug. 2006.

[26] M. Gullberg, "A helping hand? Gestures, L2 learners, and grammar," in *Gesture: Second Language Acquisition and Classroom Research*, S. G. McCafferty and G. Stam, Eds., New York: Routledge, 2009, pp. 185–210.

[27] C. Avcı, B. Arslan, and T. Göksun, "Gesture and speech disfluency in narrative context: Disfluency rates in spontaneous, restricted, and encouraged gesture conditions," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 44, 2022, pp. 1912–1917.

[28] M. Gullberg, "Multilingual multimodality: Communicative difficulties and their solutions in second-language use," in *Embodied Interaction: Language and Body in the Material World*, J. Streeck, C. Goodwin and C. LeBaron, Eds., Cambridge: Cambridge University Press, 2011, pp. 137–151.

[29] P. Peltonen, "Gestures as fluency-enhancing resources in L2 interaction: A case study on multimodal fluency," in *Fluency in L2 Learning and Use*, P. Lintunen, M. Mutta, and P. Peltonen, Eds., Bristol: Multilingual Matters, 2020, pp. 111–128.

[30] K. Lemhöfer and M. Broersma, "Introducing LexTALE: A quick and valid lexical test for advanced learners of English," *Behavior Research Methods*, vol. 44, no. 2, pp. 325–343, Jun. 2012.

[31] Council of Europe, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press, 2001.

[32] N. H. De Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior Research Methods*, vol. 41, no. 2, pp. 385–390, May 2009.

[33] P. Foster and P. Skehan, "The influence of source of planning and focus of planning on task-based performance," *Language Teaching Research*, vol. 3, no. 3, pp. 215–247, Jul. 1999.

[34] A. Cienki, "Bush's and Gore's language and gestures in the 2000 US presidential debates: A test case for two models of metaphors," *Journal of Language and Politics*, vol. 3, no. 3, pp. 409–440, 2004.

[35] R. Lopez-Ozieblo, "Proposing a revised functional classification of pragmatic gestures," *Lingua*, vol. 247, pp. 1–47, Nov. 2020.

[36] A. Kendon, *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press, 2004.

[37] H. Brugman and A. Russel, "Annotating multi-media/multi-modal resources with ELAN," in *Proceedings of the Fourth International Conference on Language Resources and Language Evolution* (LREC '04), Lisbon, Portugal, May 2004, pp. 2065–2068.

[38] D. Özer and T. Göksun, "Gesture use and processing: A review on individual differences in cognitive resources," *Frontiers in Psychology*, vol. 11, article 573555, Nov. 2020.