CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Investigating verbal and nonverbal indicators of physiological response during second language interaction

Rachael Lindberg* [iD], Kim McDonough [iD] and Pavel Trofimovich [iD]

Concordia University, Montréal, Québec, Canada
*Corresponding author. Email: rachael.lindberg@mail.concordia.ca

Abstract

Second language (L2) researchers have long acknowledged that affective variables (e.g., anxiety, motivation, positive emotions) are essential in understanding L2 learner psychology and behavior, both of which influence communication and have implications for language learning. However, there is little research investigating affective variables during L2 interaction, particularly from a dynamic rather than a static, trait-oriented perspective. Therefore, this study examined 60 L2 English speakers' affective responses in real time during a paired discussion task using galvanic skin response sensors to capture speakers' anxiety. Analyses focused on speakers' speech, their behavioral reactions, and the content of their discussion while experiencing anxiety episodes (high vs. low arousals). Findings revealed that speakers glanced away, blinked, and used self-adaption gestures (touching face, hair-twisting) significantly more frequently during high arousals than low arousals, whereas head nods were found to occur significantly more often during low arousals. In comparison to low arousals, a larger proportion of high arousals occurred while discussing personal topics. Implications are discussed in terms of the role of affective variables in communication processes.

In classroom settings where communication and meaningful interactions are important, second language (L2) teachers often recognize the value of foregrounding the development of speakers' motivation, collaboration, and positive emotion to improve their language learning experiences. Affective variables, in particular, appear to be essential in understanding the learning process (Gregersen et al., 2014; MacIntyre & Mercer, 2014) because they influence how L2 speakers communicate (MacIntyre & Vincze, 2017) and how they react to the learning context (Gardner & MacIntyre, 1993). For example, negative emotions, such as nervousness, can be associated with L2 speakers' unwillingness to engage in communication (Liu & Jackson, 2008) and can even negatively impact their language performance, as

CrossMark

evidenced by poorer performance on vocabulary tests (MacIntyre & Gardner, 1989) and lower class grades (Horwitz, 1986). Furthermore, negative emotions can affect L2 speakers' long-term achievement, such as by inhibiting the development of L2 comprehensibility (Saito et al., 2018). Identifying and understanding behaviors associated with speakers' emotional states is therefore important for fostering successful interactions and facilitating L2 development.

One of the most widely studied emotions in language learning is anxiety, where one experiences a negative emotional reaction when learning or using an L2 (MacIntyre, 1999). Anxiety can be understood as both a trait, which is a stable personality characteristic reflecting a predisposition to be anxious (Scovel, 1978), and a state, which is a transient experience that occurs in response to a particular stimulus (Spielberger, 1983). L2 anxiety and related feelings of nervousness and tension manifest in relation to various skills, including listening and speaking (Horwitz et al., 1986), but they can be driven by various nonlinguistic factors as well, such as individual differences and other social-psychological factors. With respect to individual differences, L2 speakers characterized as having high trait language anxiety tend to also be more likely to experience state anxiety in L2 contexts (MacIntyre & Gardner, 1989). In terms of sociopsychological factors, more extensive use of the target language and greater self-confidence tend to lead to lower language anxiety (Baker & MacIntyre, 2000; Sevinç, 2018), whereas negative perceptions of the interaction or one's interlocutor may lead to negative affect. For instance, because listener responses, such as backchannels, vary across individuals in frequency and placement (Cutrone, 2005; Heinz, 2003), if those conventions are not shared, speakers may perceive their interlocutor as being impatient or interrupting (Cutrone, 2005), possibly causing miscommunication (Li, 2006), anxiety, or frustration. Overall, these social-psychological factors are often larger contributors to language anxiety than linguistic factors, such as language proficiency (Hashemi, 2011; Sevinç, 2018; Sevinç & Dewaele, 2018).

Much of prior work on L2 anxiety, however, tends to take a broad and retrospective approach to examining emotional dimensions in L2 contexts by using self-reports based on ratings, questionnaires, or interviews (Liu & Jackson, 2008; MacIntyre & Gardner, 1994; Matsuda & Gobel, 2004). Yet changes in emotion can occur during L2 communication (Gregersen et al., 2014), which would be compatible with a dynamic systems perspective on language learning and use (e.g., de Bot et al., 2007), where linguistic and nonlinguistic aspects of interaction would be subject to change over time, display interconnectedness, and tend to self-organize into preferred and dispreferred states (e.g., high and low affective states). This highlights the need for additional methods that could measure emotional responses more dynamically, especially since self-reports are often argued to be inaccurate (Caldwell-Harris et al., 2011; Cramer, 2003). One such method involves capturing moment-by-moment changes in L2 speakers' physiological responses as these signify activation (or arousal) of the autonomic nervous system that regulates emotions.

Along with heart rate (Kreibig, 2010), the most commonly used physiological measure of anxiety is skin conductance response, which is obtained by using galvanic skin response sensors to assess sweating. Increased arousal, linked to a higher skin conductance response, indicates increased activity in the sympathetic

nervous system (a branch of the autonomic nervous system known to be associated with response to stressors) which is otherwise known as the fight or flight response (MacPherson et al., 2017). Skin conductance has been widely adopted as a reliable measure of stress reaction related to anxiety (e.g., Santos Sierra et al., 2011; Setz et al., 2010) and shown to be the most effective in differentiating between anxious and nonanxious states compared to other physiological measures, including cardio-vascular or respiratory (Blechert et al., 2006).

Skin conductance research has shown that increased arousal tends to occur under more demanding or stressful conditions, typically associated with an increase in cognitive load (MacPherson et al., 2017), or can occur in response to speech-related anxiety, such as during public speaking (Clements & Turpin, 1996; Croft et al., 2004; Kreibig, 2010), especially for speakers with higher levels of trait anxiety (Witt et al., 2006). Similar findings have emerged in L2 contexts. Gregersen et al. (2014) found that L2 Spanish learners with high trait anxiety experienced higher arousal during a classroom presentation than their classmates with low trait anxiety. Focusing on Turkish-Dutch bilinguals, Sevinç (2018) found that bilinguals experienced higher arousal (i.e., greater levels of skin conductance) when speaking with a native speaker in their less dominant language, which was consistent with their self-reported anxiety.

To date, physiological measurements have been primarily used to detect anxiety in stress-inducing contexts (e.g., class presentations, public speaking) and compared across different conditions (e.g., manipulating cognitive difficulty of tasks). However, while controlled experimental studies are important, recording skin conductance during open-ended conversations allows for exploration of a wider range of behaviors and interactional features, providing further insight into what may be associated with stress reactions during interaction. It could be that the interlocutor's speech or the topics that arise during the conversation are related to an affective response. For example, during interaction, L2 speakers with lower oral proficiency tend to demonstrate higher skin conductance responses (e.g., Sevinç, 2018), possibly due to changes in skin conductance being directly related to linguistic performance (e.g., code-switching, errors) or certain stimuli, such as unknown words (Geen, 1989). The nature of the conversation may also be associated with arousal, where discussing more negative topics (e.g., situations when one felt sad or angry) can elicit a stronger skin conductance response than when discussing positive topics (e.g., situations when one felt excited or happy), relative to neutral topics (Burbridge et al., 2005). However, it is less clear how interactional features (e.g., backchanneling, questions) or the type of discussion topics (i.e., personal experiences, language-focused content) may influence the arousal experienced by L2 speakers. While prior research has compared speech events (e.g., public speaking, completing a complex task, speaking in a less dominant language) that may be stress-inducing (e.g., Sevinç, 2018; Wallbott & Scherer, 1991), it has not been examined what happens within a communicative event that is associated with stress reactions. Therefore, an examination of interactional features and discussion topics may provide further insight into the components of interaction that contribute to L2 speakers' feelings of anxiousness.

In addition to using physiological measures to detect anxiety, researchers have also investigated visual cues as a way of identifying anxiety-driven behaviors.

For example, by comparing facial expressions and body movements between relaxed, neutral, and anxious states as manipulated through elicitation tasks that varied in emotional strength, Giannakakis et al. (2017) found that small, rapid head nods, a higher blink rate, and more brow movement and mouth activity were more common in the anxious state. Gregersen (2005) also investigated nonverbal behaviors, where during an oral foreign language exam, self-reported high-anxious learners blinked more, made less eye contact, and more frequently touched their face, hair, or fidgeted with an object. These examples of self-contact and object manipulation, which are categorized as self-adaption gestures within Ekman and Friesen's (1969) gesture category of "adaptors" (i.e., unintentional body movements), are widely recognized as responses to stress or negative feelings about oneself or others (Gregersen, 2005; Richmond & McCroskey, 2004). Furthermore, adjusting one's posture and restlessness (Ekman & Friesen, 1974) and leaning away from the interlocutor (Burgoon & Koper, 1984) have been associated with negative affect, such as feelings of awkwardness, tension, or anxiety.

To date, most research investigating nonverbal behaviors of autonomic arousal has focused on experimentally induced stress conditions. For example, Wallbott and Scherer's (1991) high cognitive stress condition was associated with more physiological indicators of anxiety (higher pulse and respiration rate) compared to the low stress condition, along with more brow lowering and more smiling, likely a defensive masking of negative emotions (see also Giannakakis et al., 2017). In contrast, other research has suggested that maintaining tenser facial muscles, thus less smiling and less brow movement, is more common for high- than low-anxious learners (Gregersen, 2005). One reason for these mixed findings could be that nonverbal behaviors, especially facial expressions, may be influenced by culture, and thus have multiple meanings (Gregersen, 2009; Hashemi, 2011). Different cultures may have different conventions for controlling facial expressions when feeling anxious in social situations, such as smiling to mask certain expressions (Ekman, 1977). Therefore, research that includes speakers from diverse multicultural backgrounds is necessary to identify more generalizable relationships between physiological measures and nonverbal behaviors that characterize autonomic arousal.

Despite calls for research to explore the emotional experiences of L2 speakers, which are essential to understanding learner psychology, and to examine verbal and nonverbal cues associated with state language anxiety (Gregersen et al., 2014; MacIntyre & Mercer, 2014; MacIntyre & Vincze, 2017), there is little research investigating these affective variables during L2 interaction particularly from a dynamic rather than a static, trait-oriented perspective (e.g., Matsuda & Gobel, 2004). Therefore, the current study adopted a time-sensitive approach by measuring L2 speakers' skin conductance (i.e., sweating), which is considered a reliable indicator of anxiety (e.g., Kreibig, 2010), to assess their affective responses continuously in real-time interaction. The goal of the study was to explore the verbal and nonverbal characteristics accompanying L2 speakers' physiological responses. The following research questions were addressed:

1. Which verbal features (e.g., disfluencies, listener responses, questions, discussion topics) occur among L2 speakers when they experience high versus low autonomic arousal during L2 interaction?

2. Which nonverbal features (e.g., eye gaze, facial, and body gestures) occur among L2 speakers when they experience high versus low autonomic arousal during L2 interaction?

Method

Participants

The participants were 60 L2 English speakers with a mean age of 24.57 years (*SD* = 4.74, *range* = 18–44) who were studying at English-medium universities in Montréal, Canada, as undergraduate (30) or graduate students (30). These participants were sampled from the larger Corpus of English as a Lingua Franca Interaction (CELFI), which consists of a diverse sample of L2 English students carrying out three discussion tasks in pairs (McDonough & Trofimovich, 2019). Assuming that responses to arousal and anxiety may be affected by gender (e.g., Wallbott & Scherer, 1991) and culture (e.g., Ekman, 1977), the 30 pairs were selected to balance gender (10 male–male pairs, 10 female–female pairs, 10 male–female pairs) and to have a diverse representation of language backgrounds within each gender group. Overall, the speakers spoke 21 different first languages, with the majority being Mandarin (10), Spanish (7), Arabic (6), and French (5). Participants in each pair came from different language backgrounds and did not know each other prior to the study. They had been studying English for an average of 12.33 years (*SD* = 5.06, *range* = 2–20) and had been living in Canada for an average of 3.83 years (*SD* = 4.39, *range* = 2 weeks–22 years).

Materials

The materials included a communicative task, a post-task survey consisting of 10 rating scales, a background questionnaire, and an L2 communicative trait-anxiety questionnaire [study materials posted at http://www.iris-database.org].

Communicative task

Both participants were given a handout with the following discussion prompt: "What are three challenges or problems that are important for most international students arriving in Québec? Try to find one possible solution for each challenge or problem you have identified and give reasons for why you are suggesting that solution." This interactive task was chosen for analysis among the three discussion tasks used in CELFI (McDonough & Trofimovich, 2019) as it promoted collaborative brainstorming and problem solving and was a familiar topic to all participants who were international students themselves. In addition, during post-experiment debriefing, this task was described by participants as being relatively easy compared to the other two tasks (academic discussion, exchange of personal narratives), so task difficulty would not play a major role in triggering anxiety-related arousals.

Rating scales

After the discussion task, participants received ten 100-millimeter continuous scales (a line with labeled endpoints, where the left endpoint indicated negative ratings and the right indicated positive ratings). Participants marked an X on the scale to rate themselves and their discussion partner for five different criteria: comprehensibility, flow, motivation, collaboration, and state-anxiety. These ratings were part of the data corpus, but only participants' self-ratings of state-anxiety (perceived level of stress, worry, or nervousness during the task) were used in the current study.

Trait-anxiety questionnaire

Adapted from MacIntyre and Gardner's (1994) input, processing, and output anxiety scales, this questionnaire measured participants' trait-anxiety when using English by having them indicate their agreement with 18 different statements on a 6-point Likert scale (1 = *strongly disagree*, 6 = *strongly agree*). Both negatively- and positively-worded items were used as suggested by Dörnyei (2003). Statements were related to how anxious participants feel when speaking English (e.g., I feel relaxed when I have to speak in English), when learning and thinking in English (e.g., Learning English vocabulary does not worry me, I can acquire it in no time), and when listening to English (e.g., I get upset when English is spoken too quickly). Items were modified from MacIntryre and Gardner (1994) to focus on oral communication by changing statements mentioning tests or readings to refer to presentations and lectures instead.

Procedure

Experimental setup

The study took place in a lab setting, where the two participants were seated at a table across from each other with a FOVIO eye-tracking system placed between them to monitor their eye gaze behavior. Two Logitech webcams captured each participant's field of vision (i.e., the upper body of their conversation partner) and their eye movements, which were visually depicted by a green dot in the video. The FOVIO units and cameras were connected to two synchronized Dell laptops, where the field-of-vision data and eye movements were recorded using Eyeworks Record software. Participants' galvanic skin responses were measured using TEA Captiv T-sens sensors. The charged battery packs of the sensors were attached to a velcro wristband strapped to the participants' nondominant hand, with the two electrodes secured with a velcro strap on the distal (first) phalanges of their middle and index fingers. The electrodes measured participants' skin conductance, assessing sweat secretion, and thus captured the episodes of autonomic arousal experienced by participants during their interaction. The signal from the sensors was captured via Bluetooth using a T-Receiver box and recorded in Captiv software (http://www.teaergo.com) on a Dell laptop. Participants generally stated that the equipment did not preclude them from engaging in natural conversation, as they provided an average rating of 81.2 ($SD = 18.6$) when asked how distracted they were by the experimental setup (where 100 indicates they were not at all distracted). Although it remains possible that the lab equipment created an unnatural and hence

stressful environment, participants reported that they felt comfortable interacting with their partner ($M = 86.75$, $SD = 14.00$), that their overall experience was positive ($M = 89.58$, $SD = 14.61$), and that they perceived to have experienced only minimal anxiety overall during the task ($M = 70.50$, $SD = 22.00$), where 100 meant they felt very comfortable, their experience was very positive, and that they were not at all anxious, respectively.

Task procedure

After signing the consent form, participants were introduced to the rating scales and the definitions of the terms (5 minutes). Instruction was then given to participants on how to attach the sensors to their fingers, followed by a four-point calibration of the eye-tracking equipment (5 minutes). Once both the sensors and eye-trackers had started recording, the discussion topic was introduced with a brief warm-up, which was not analyzed as part of the main task (3 minutes). The discussion prompt was then explained to participants orally and given to them on a handout. To minimize any potential stress induced by the presence of an observer, the researcher left the room, giving participants 10 minutes to complete the task. Afterward, both participants were given the rating scales and rated themselves and their partner in terms of comprehensibility, state-anxiety, motivation, collaboration, and flow (2 minutes). Finally, participants filled out the background questionnaire and trait-anxiety questionnaire (10 minutes).

Data analysis
Autonomic arousal

The webcam field-of-vision videos with participants' recorded eye movements were synchronized for each pair using Captiv software, and each participant's sensor data were synchronized with the videos. This allowed participants' behavioral reactions and their speech (as seen and heard in the videos) to be linked to their physiological responses (i.e., arousal episodes). No task lasted less than 10 minutes; for the six longer task performances, only the first 10 minutes of the task were analyzed. The proprietary coding algorithm in Captiv identified five levels of arousals, categorizing each episode as either high, medium-high, medium, medium-low, or low, based on both the amplitude (peak value) and the slope of the recorded skin conductance function. These levels, which are specific to the TEA Captiv T-sens sensor, are expressed in microSiemens (µS), where a greater value is associated with a higher rate of sweat secretion (Dawson et al., 2016). The determined arousal level reflected the range of each individual's response, such that what constituted high, medium, or low arousal depended on the magnitude of absolute differences in skin conductance levels recorded for a given individual. Because the goal of this study was to document verbal and nonverbal features of high and low affective response, for this study, only the episodes of high and low arousals were analyzed to compare the two extremes. Thus, high arousals corresponded to responses with the steepest function slopes and highest amplitudes while low arousals corresponded to responses with the shallowest function slopes and lowest amplitudes, relative to all arousal episodes experienced by a given participant.

A response function in a sensor signal indicated either a specific skin conductance response, where the reaction was due to a specific external stimulus or event (e.g., due to something a speaker's partner said), or a nonspecific skin conductance response, where the reaction was related to an emotional component, such as feeling discomfort when speaking English (Setz et al., 2010). However, as individuals' skin conductance levels vary substantially (Setz et al., 2010), using raw amplitude measurements in µS for comparison between individuals was not plausible. For example, a skin conductance level of 8 µS may be considered high for one individual but may be relatively low for another (Dawson et al., 2016). To exemplify this in the context of our sample, participants' average low arousal was 7.25 µS ($SD = 5.22$; $range = 1.91$–22.85), while their average high arousal was 10.53 µS ($SD = 10.16$; $range = 1.01$–43.57). As shown by the high standard deviations and wide ranges, a value as low as 1.01 was a high arousal for one participant while 22.85 was low for another participant. Due to this variability and given that the frequency of skin conductance responses is used as a measure of psychophysiological activation (Setz et al., 2010), the proportions of high and low arousals (out of all instances of arousal experienced at the five recorded levels), which included both specific and nonspecific responses, were used in all further analyses as a standardized measure to compare the occurrence of high and low arousals between participants. In addition, because skin conductance response typically has a latency period of 1–3 seconds (Bradley, 2000; Figner & Murphy, 2011), the arousal episodes were identified within a moving window of 1–3 seconds prior to the first significant deviation in the signal. This response delay was taken into account to connect the behavioral and linguistic events in the videos to the physiological responses. Therefore, participants' verbal and nonverbal behaviors occurring within these 2-second windows were coded.

Interlocutor behaviors

To create a coding scheme for interlocutor behaviors, five participants' task performances (not included in this study) were first open coded by a trained coder for recurring verbal and nonverbal characteristics to capture as many features that could be potentially associated with autonomic arousal. These initial dimensions spanned 14 categories for verbal features and 20 categories for nonverbal features. Through iterative coding, the categories were reduced to four verbal features and nine nonverbal features, to include only the categories that occurred commonly in this data set and that were investigated previously in relation to anxiety, such as blinks (Wallbott & Scherer, 1991) or brow movement (Giannakakis et al., 2017).

Verbal features

The final coding framework for verbal features is summarized in Table 1, along with illustrative examples of high arousal episodes for each category, where bolding indicates who was experiencing the arousal and underlining highlights when it occurred. The dysfluency category included unfilled pauses, defined as silent pauses longer than 0.5 seconds within a sentence or phrase (Riggenbach, 1991), filled pauses (e.g., uh, um), repetitions (i.e., repeating the same word or phrase), and repairs

Table 1. Coding Framework for Verbal Features

Coding category	Variable	Example
Disfluency	Unfilled pause	**P140**: "I just arrived here when I was … nineteen."
	Filled pause	**P156**: "Especially the … government should hold a different uh biggest uh most important festival."
	Repetition	**P112**: "We can we can call that like one major problem."
	Repair	**P192**: "Like let's say, when you are in higher – uh high school […]"
Listener responses	Backchannel	**P95**: "These three thing can be manageable."P96: "Mhmm."
	Noncorrective repetition	**P105**: "It's like an interview kind of."P106: "Yeah yeah like an interview."
	Emotional response	**P275**: "When they did send me the card, they put me as male."P276: "Oh my god!"
Communication breakdown	Language-related nonunderstanding	P93: A thesis student. Thesis.P94: Thesis?**P93**: Uhhh I mean, you – the professor in-instruct with you … in the lab.
	Global nonunderstanding	P93: Like the university can um organize uh the – can make up some organi– organization about the students who – uh where they come from. **P94**: Um … I don't understand.
Questions	Asking question to partner	**P136**: "Are you learning French right now?"
	Receiving question from partner	P219: "Could you find friends? Finding friends was easy for you? It's not a challenge?"**P220**: "It's not that challenging."

Note. Underlining indicates when the high arousal (where appropriate) occurred, and bolding indicates who experienced it.

(i.e., self-corrections of pronunciation, grammar, or vocabulary). The listener response category included three types of reactions from each speaker's interlocutor: backchannels (e.g., mhmm, yeah, okay), noncorrective repetitions (i.e., repeating a word or phrase the interlocutor said to show agreement or understanding), and emotional responses (i.e., reactions of amazement, disbelief, disgust, etc.). The remaining categories included communication breakdowns (i.e., difficulties understanding language or the global message of the utterance) and questions (i.e., asking or receiving a question). All verbal variables were coded categorically in terms of whether they occurred or not. Another trained coder used this framework and independently coded arousals from 10% of the participants (6 participants, 50 arousal episodes). Coding reliability (Cohen's kappa) was .73 for disfluencies, 1.00 for listener responses, and .85 for questions, suggesting substantial reliability (Landis & Koch, 1977). There was no kappa value for communication breakdown, as it did not occur in the subset of data used to measure interrater reliability.

Topics of discussion

Assuming that a speaker's arousal might also be associated with the content discussed, thus influencing the occurrence and magnitude of skin conductance responses (Burbridge et al., 2005), additional variables were derived by coding the topics of participants' discussions. High and low arousals were examined within the broader context of the conversation (i.e., from the beginning of the idea that was being discussed when the arousal occurred) using bottom-up coding. Conversation themes that emerged included topics related to general difficulties in Montréal that international students tend to face and solutions to those problems (i.e., discussion directly related to the task prompt), personal difficulties participants have dealt with and personal (nonproblematic) experiences, discussing their home country, or when their partner was discussing a personal difficulty or experience or talking about their home country. Additionally, when participants diverged from the task, or when there was complete silence, it was coded as silence/off-topic. This was a relevant category for our multi-cultural sample, as silent behavior and the typical length of pauses between speaking turns varies across cultures (King, 2013; Ruetenik, 2013). Lastly, when they discussed language, such as the meaning of a word used, the content was coded as language related. Table 2 outlines these conversation themes with illustrative examples of high

Table 2. Coding Framework for Topic of Conversation

Coding Category	Example
General difficulty	**P67:** "Living conditions is important, especially when you're just coming and it's hard to find people, but I guess they have residences, but residences are so expensive."
Personal difficulty	**P188:** "I'm not learned the French so when I want to talk to some people, I tell the driver, this road, what this road? I don't know, I'm not pronounce this exactly. Yeah, so that is also the problem for me."
Personal experience	**P278:** "I had not seen snow for almost 20, 22 years and then when I came here it was like so much snow!"
Partner's personal difficulty	P227: "I don't have like a very big vocabulary in French. So like sometimes I just don't understand a word. But like based on the whole sentence, I understand. And based on the context I can guess what it means. But still, like you know, you feel really nervous because you don't know what it really means! So yeah." **(P228)**
Partner's personal experience	P84: "For example, I'm seeing this guy now…and he speaks three languages." **(P83)**
Home country	**P264:** "I'm from Bangladesh."
Partner's home country	P136: "I'm from Israel, so it's such a different culture. Like here it's so individualistic […] where like in Israel everyone is just like it's like a gathering." **(P135)**
Language-related	P94: A scholarship. P93: …Uhh what's the meaning of a scholarship? **P94:** It's like um financial help to a study
Silence/off-topic[example displays off-topic talk]	**P219:** So um are we done, it's ten minutes orP220: Uh I think so. Uh it's not ten minutes otherwise she could have come

Note. Underlining indicates when the high arousal (where appropriate) occurred, and bolding indicates who experienced it.

arousal episodes, where bolding indicates who experienced the arousal, and underlining highlights the utterance when it occurred. The same trained coder coded the discussion content, and a re-coding of the discussions during 50 arousal episodes by another blind coder revealed a κ coefficient of .70, indicating substantial interrater reliability (Landis & Koch, 1977).

Nonverbal features

The final coding framework for nonverbal features, along with the definition of each category, is shown in Table 3. Except for nods and blinks, which were numeric variables, the remaining variables were binary, where the occurrence of the feature was coded as 1 and the absence of the feature was coded as 0. The recorded eye gaze, depicted by green dots in the videos, allowed for precise coding of participants' eye movement. The initial coding was carried out by the same trained coder, and a subset of the data (10%) was coded by another blind coder. For the numeric variables, coding reliability (Cronbach's alpha) was .88 for blinks and .97 for nods. For the binary variables, coding reliability (Cohen's kappa) was greater than .74 for the eye gaze categories, greater than .81 for body gestures (body shift, self-adaption gestures), and greater than .72 for facial gestures (brow movement, smile/laugh), again suggesting substantial reliability (Landis & Koch, 1977).

Because participants varied in the number of arousals they experienced (see the section below on the occurrence of high and low arousal), to obtain a standardized, comparable measure for all analyses, the occurrences of each coded variable within every verbal, content, and nonverbal category type were normalized per participant for high and low arousal episodes separately, such that the frequency of each coded category was divided by the total number of high (or low) arousals for that participant.

Table 3. Coding Framework for Nonverbal Features During Episodes of High and Low Arousal

Coding category	Variable	Definition
Eye gaze	Gaze aversion	Never looking at partner (binary)
	Glance away	Shifting gaze to the side, up, or down (binary)
	Mutual gaze	Making eye contact (binary)
Facial and body gesture	Blink	Number of blinks (numeric)
	Nod	Number of head nods (numeric)
	Self-adaption gesture	Touching neck, scratching face, hair-twisting (binary)
	Smile/laugh	Smiling and/or laughing (binary)
	Brow movement	Raising or lowering of their eyebrows (binary)
	Body shift	Shifting position (while seated), leaning forwards or backwards (binary)

Ratings and questionnaires

The post-task ratings (i.e., self-ratings of state-anxiety) were measured (in millimeters) from the left endpoint to the cross marked by participants to obtain a rating between 0 and 100. Regarding the trait-anxiety questionnaire, positive statements were reverse-scored for analysis, such that higher scores (more agreement) indicated greater L2 communicative trait-anxiety. After internal consistency (Cronbach's alpha) was checked for each subconstruct (input, output, processing trait-anxiety), two items from processing anxiety and one item from input anxiety were removed due to their initial low reliability (George & Mallery, 2003). Mean scores of each component – L2 input trait-anxiety ($\alpha = .69$), L2 output trait-anxiety ($\alpha = .77$), L2 processing trait-anxiety ($\alpha = .71$) – were used for subsequent analyses.

Results

Occurrence of high and low autonomic arousal

Across the 60 participants, the data set was composed of a total of 444 arousal episodes (271 high, 173 low), identified by Captiv's coding algorithm. Participants on average experienced 4.52 high arousals ($SD = 3.43$; *range* $= 0–16$) and 2.88 low arousals ($SD = 3.51$; *range* $= 0–14$) during the 10-minute communicative task. Put differently, out of all the arousals experienced by a participant, which also included the medium-high, medium, and medium-low arousals that were not included in the main analysis, on average 10% were considered high ($SD = 7\%$; *range* $= 0–33\%$) and 6% were considered low ($SD = 7\%$; *range* $= 0–32\%$).

To explore how the occurrence of arousal may be related to anxiety, Pearson correlations (two-tailed) were conducted between the proportion of participants' high and low arousals (out of all arousals experienced), their trait-anxiety scores (input, output, and processing anxiety), and their self-ratings of state-anxiety (see Table 4). First, the associations between state and trait measures of anxiety were below the field-specific threshold (.25) for a weak relationship (Plonsky & Oswald, 2014), which suggested that (at least in this data set) the two sets of measures tapped into distinct aspects of anxiety. Second, the associations between the anxiety measures and the occurrence of high and low arousals were also weak to nonexistent (Plonsky & Oswald, 2014), which implied that the arousals may have occurred due to participants' specific task and partner experiences rather than because they were generally anxious (trait-anxiety) or perceived themselves as anxious during the task (state-anxiety).

Table 4. Correlations Between Anxiety Variables and Proportion of Arousals Experienced

Variable	Low arousal	High arousal	State-anxiety
Output trait-anxiety	−.09	.08	−.10
Input trait-anxiety	.07	−.14	−.08
Processing trait-anxiety	−.25	.21	−.08
State-anxiety	.03	−.06	—

Verbal features of autonomic arousal

To answer the first research question focusing on the verbal features that occur when L2 speakers experience high versus low arousal episodes in interaction, the instances of three verbal features were compared between high and low arousals to see if an affective response may be related to either language issues (e.g., disfluent speech), asking or receiving a question, or having an interlocutor who is an active listener giving verbal feedback. Communication breakdowns were ultimately removed from the analyses as they occurred for only seven of the 60 participants, making the data set too small to examine how communication breakdowns related to arousal. Although pairwise comparisons (Bonferroni corrected $\alpha = .017$) yielded no significant differences for any of the verbal features (see Table 5), there was a trend for speech disfluencies (e.g., filled and unfilled pauses, repairs, and repetitions) to occur more often during high arousals (one instance every four arousals) compared to low arousals (one instance every 10 arousals). Therefore, none of the coded verbal variables seemed to distinguish high versus low arousals.

In terms of possible relationships between participants' psychophysiological responses and the conversational content, as shown in Table 6, the majority of both

Table 5. Verbal Features of High Versus Low Arousal

Variable	Low arousal		High arousal		Paired-samples *t* test		
	M	*SD*	*M*	*SD*	*t*(34)	*p*	*d*
Disfluencies	0.10	0.27	0.25	0.36	2.48	.018	0.39
Listener responses	0.01	0.02	0.11	0.21	1.11	.280	0.21
Questions	0.11	0.23	0.10	0.21	−0.21	.840	−0.03

Note. Comparisons based on 35 participants that experienced both low and high arousal.

Table 6. Proportions of Topics (k Instances) Discussed During High Versus Low Arousal

Variable	Low arousal		High arousal	
	k	%	*k*	%
General difficulty	93	54	138	51
Personal difficulty	7	4	35	13
Personal experience	8	5	33	12
Partner's personal difficulty	25	15	32	12
Partner's personal experience	19	11	12	4
Home country	8	5	8	3
Partner's home country	9	5	2	1
Language related	0	0	4	2
Silence/off-topic	4	2	7	3
Total	173	100	271	100

high and low arousals happened while participants discussed difficulties experienced by international students and provided possible solutions to these difficulties, which was content directly related to the task goal. More noteworthy, however, was that a greater proportion of the high arousals occurred when participants spoke about their personal difficulties (13% vs. 4% of high and low arousals, respectively) or personal experiences (12% vs. 5% of high and low arousals, respectively). On the other hand, a larger proportion of the low arousals co-occurred with the speakers' partners describing their personal experiences (4% vs. 11% of high and low arousals, respectively).

Nonverbal features of autonomic arousal

The second research question focused on the nonverbal features that occur when L2 speakers experience high versus low arousal episodes in interaction. First, we compared the occurrences of the nine nonverbal features between high and low arousal episodes on the assumption that any variables that occurred significantly more frequently during high versus low arousals could be behavioral manifestations of a given affective response. As summarized in Table 7, paired-samples *t* tests (Bonferroni corrected α = .006) revealed that four nonverbal features differed significantly in their occurrence between high and low arousal episodes. More specifically, participants glanced away, blinked, and used self-adaption gestures more frequently when they were experiencing high arousal but nodded more often during episodes of low arousal, all with small-to-medium effect sizes (Plonsky & Oswald, 2014). In relative terms, participants were 2.5 times more likely to glance away, 1.7 times more likely to blink, and 20 times more likely to use self-adaption gestures during high than low arousals, but they were 2.7 times more likely to nod when experiencing low than high arousals.

Table 7. Nonverbal Features of High Versus Low Arousal

Variable	Low arousal		High arousal		Paired-samples *t* test		
	M	*SD*	*M*	*SD*	*t*(34)	*p*	*d*
Gaze aversion	0.10	0.16	0.13	0.22	0.79	.434	0.13
Mutual gaze	0.81	0.20	0.74	0.30	−1.25	.221	−0.20
Glance away	0.18	0.31	0.45	0.39	3.20	.003	0.55
Blink	0.66	0.63	1.13	0.76	5.43	.001	0.92
Self-adaption gesture	0.01	0.04	0.20	0.30	3.69	.001	0.61
Nod	0.56	0.65	0.21	0.40	−3.32	.002	−0.56
Smile/laugh	0.11	0.24	0.23	0.36	2.02	.051	0.33
Brow movement	0.06	0.10	0.19	0.30	2.49	.018	0.41
Body shift	0.01	0.05	0.12	0.27	2.22	.033	0.39

Note. Comparisons based on 35 participants that experienced both low and high arousal.

Discussion

The goal of this study was to investigate L2 anxiety from a dynamic perspective, examining L2 speakers' physiological responses in relation to various verbal and nonverbal features of interaction. From a dynamic systems perspective (e.g., de Bot et al., 2007), the complexity of L2 speakers' emotions requires a moment-by-moment approach in investigating speaker affect due to the interconnectedness of the variables involved in L2 communicative events and the potential for their influence on each other to change even over short periods of time. Our results therefore highlight how a speaker's cognition and emotion, an interlocutor's contributions, and nonverbal behavior are intertwined as part of a complex system (i.e., L2 peer interaction). In terms of the verbal features co-occurring with arousal, none of the coded variables appeared to distinguish between the episodes of high versus low arousal experienced by L2 speakers, although high arousals tended to occur when L2 speakers discussed their personal difficulties and personal experiences, while low arousals tended to coincide with the speakers' partner discussing their own difficulties and experiences. However, high arousals appeared to be distinguished from low arousals through nonverbal behaviors, with more frequent glancing away, blinking, and self-adaption gestures associated with high arousal episodes and more frequent nodding linked to low arousal episodes.

Nonverbal indicators of autonomic arousal

In this data set, episodes of high versus low arousal were distinguished through L2 speakers' nonverbal behaviors, such that glancing away, blinking, and self-adaption gestures (e.g., scratching one's face, hair twisting) were indicative of high arousal while frequent nodding coincided with low arousal. Regarding eye gaze behaviors, even though glancing away (breaking eye contact) was linked to autonomic arousal, gaze aversion (absence of any eye contact), and mutual gaze (consistent eye contact between interlocutors) did not differ between levels of autonomic response. It may therefore be that a physiological response is associated with the action of breaking eye contact rather than with its presence or absence. Assuming that maintaining eye contact indicates a nonverbal attempt for a speaker to reinforce the interaction with their interlocutor (Richmond & McCroskey, 2004), it could be that the more anxious L2 speakers, who were perhaps more reticent, glanced away more frequently to avoid engaging in prolonged eye contact. This behavior aligns with that of Gregersen's (2005) high-anxious L2 learners who maintained less direct eye contact with their teacher during an oral exam.

Glancing away could also be attributed to cognitive processing, as both speakers and listeners tend to look away when attempting to process complex information (Knapp & Hall, 2001). Consistent with this explanation, blinking was also increased during high arousals, which is typically a behavior associated with a higher cognitive load, thinking, and internal attention situations, such as when completing mental tasks (Eckstein et al., 2017; Karson et al., 1981) and when responding to questions (Hirokawa et al., 2004). As increased arousal also tends to be associated with an increase in cognitive load (e.g., MacPherson et al., 2017), a higher blink rate may be the involuntary behavior that manifests along with an emotional response

due to more complex cognitive processing and speakers shifting their attention inward when speaking and formulating their thoughts. Alternatively, blinking may have been a reflex to anxiety-evoking stimuli, as blink rate also tends to increase while in an anxious state or under stress conditions (Giannakakis et al., 2017; Harrigan & O'Connell, 1996). The present finding therefore contributes to previous work which lacked physiological measures but nonetheless used blinking as an index to measure interpersonal stress (e.g., Hirokawa et al., 2001).

Categorized as unintentional body movements (Ekman & Friesen, 1969) that tend to be responses to stress or negative feelings about oneself or others (Gregersen, 2009), self-adaption gestures occurred significantly more frequently during high arousals compared to low arousals. This result aligns well with prior findings showing the occurrence of self-adaption behaviors among high-anxious L2 learners (Gregersen, 2005). Based on Harrigan's (1985) findings that speakers tend to engage in self-adaption gestures just seconds before or after an utterance, it is also possible that this nonverbal behavior may reflect psycholinguistic processes, such that the accompanying physiological response may have occurred due to anxiety associated with language processing (before an utterance) or the speech content (after an utterance).

On the other hand, facial expressions, such as smiling and brow movement, were not found to be representative of high arousal, and both rarely occurred (only about once every five high arousal episodes). It is not surprising that a high affective response did not appear to be marked by facial cues considering that high-anxious L2 learners are less likely to use facial expressions compared to low-anxious learners (Gregersen, 2005) and that anxious speakers generally display rigid communicative behavior (Buck, 1984). Although body shifts tended to occur more often during high arousals, this nonverbal behavior did not distinguish reliably between low and high arousals, likely because this broad category encompassed movements associated with negative affect, such as leaning back (Burgoon & Koper, 1984), readjusting posture, and restlessness (Ekman & Friesen, 1974), along with other movements associated with positive affect, such as leaning forward (Harmon-Jones et al., 2011; Miragall et al., 2019), and more generally with nonanxious states (Gregersen, 2007). Finally, our finding that head nods occurred significantly more frequently during low arousals (about one every other episode) than high arousals (about one every five episodes) is consistent with previous work suggesting that nonanxious learners tend to use more head nods than anxious learners (Gregersen, 2005). Thus, head nods might be a particularly useful indicator that those displaying this type of behavior during interaction are likely not experiencing negative affect or struggling with communication reticence.

Verbal and content indicators of autonomic arousal

Unlike the visual behaviors, verbal features appeared to be distributed similarly across high and low arousal levels, but there was a trend for high arousals to occur more frequently during speech disfluencies (i.e., unfilled and filled pauses, repetitions, repairs). It is widely acknowledged that individuals generally have more difficulties with language production when feeling negative emotions, such that more anxious students tend to exhibit poorer output quality (MacIntyre & Gardner, 1994;

MacIntyre et al., 1997). The trend highlighted in our results points to the possibility of negative affect being associated with speech fluency through increased physiological arousal, similar to how high arousal could impair cognitive task performance (Anderson, 1994). Just as certain nonverbal behaviors (blinking, glancing away, self-adaption gestures) reflected a possible relationship between physiological arousal and cognitive processing, participants' increased disfluencies may have emerged as verbal indicators of high arousal impairing cognitive and language functions (Burbridge et al., 2005).

Although none of the verbal features in our study distinguished high arousals from low arousals, high arousals occurred more often when speaking compared to listening (56% speaking, 44% listening), but low arousals occurred more frequently while listening (35% speaking, 65% listening). This suggests that different conversational roles and different levels of interactional engagement may in fact contribute to physiological arousal, but it remains for future research to explore further which aspects of speech production and comprehension may be associated with a speaker's emotional response. Furthermore, although listener responses (backchannels, noncorrective repetitions, emotional reactions) may lead to negative perceptions during interaction if speakers misunderstand the intention of such listener contributions (Cutrone, 2005), there was no evidence in this data set that listener responses were associated with certain levels of a speaker's arousal. Perhaps listener responses were positively received by the speakers and therefore played no role in their affective reaction, particularly because all interactions were nonconfrontational and collaborative.

In terms of the speech content associated with high versus low autonomic arousal, apart from conversations directly related to the task goal, the speakers tended to experience more high arousals while sharing a personal story (positive or negative). On the other hand, low arousals occurred more commonly while the speakers' partner was sharing a personal experience (positive or negative). In other words, the speakers tended to have a high affective response while recounting stories about themselves but felt calmer when listening to their partner share their experiences or challenges. Although existing findings are mixed, positive emotions can reflect autonomic nervous system activity (Kreibig, 2010), so the high arousal experienced by the speakers could have been a response to feelings of happiness, amusement, or pride when describing their positive experiences. Experiencing high arousal while discussing personal difficulties also aligns with prior research showing a higher frequency of skin conductance responses while discussing affectively negative topics (Burbridge et al., 2005). It is likely that the speakers were simply more anxious when sharing stories about themselves with an interlocutor whom they had never met before, given that speaker familiarity with interlocutors can lead to fluctuations in anxiety (Shirvan & Talebzadeh, 2017).

Alternatively, the emotionality of the content discussed may have contributed to high arousal, especially because higher levels of skin conductance have been shown to be associated with receiving empathetic, emotional responses from listeners compared to inattentive behavior (Finset et al., 2011). Therefore, the speakers' partner may have remained attentive and expressed statements of empathy in return, triggering a high affective response for the speaker, just as the empathetic remarks from the interlocutor in the Finset et al. study (2011) reinforced the speakers' distress

related to their own emotionally charged experience. While receiving empathy may trigger a high affective response for the speaker, feeling empathy toward others may have the opposite effect, which could explain why low arousals were more common while listening to others' personal stories. Indeed, identification (which involves acknowledging someone else's emotion and adopting it as one's own) appears to be a defense mechanism to stress associated with lower skin conductance levels (Cramer, 2003). Because all speakers could relate to the discussion topic (challenges experienced by students arriving in a new context), they could identify with the experience described by their partner and empathize with any emotions conveyed through the story, which was then reflected through their low arousal.

Limitations and future work

Although promising, the findings of this study have several limitations. A major limitation pertains to our use of sensor-specific skin conductance levels to capture autonomic arousal, which implies that verbal and nonverbal indicators of a speaker's affective response are particular to the relative differences between high versus low autonomic arousal and are specific to the skin conductance sensor used. Similarly, we only examined the two extreme levels of arousals, disregarding intermediate values. For instance, it could be that some behaviors may not have been coded because they did not co-occur with the highest level of arousal, but they may still have been present as a reaction to an intermediate-level arousal. In addition, although it was not necessary for our current comparison of skin conductance responses associated with high and low autonomic arousal, we did not have a baseline measurement of skin conductance. To determine how speakers' skin conductance levels change over time in response to particular stimuli, future studies should include a baseline measurement. Therefore, researchers should consider taking a finer-grained approach to include moderate levels of arousal or should instead compare arousal levels across resting and target conditions as a way of including baseline data and increasing the range of affective states investigated. Finally, future work should examine the extent to which experimental lab settings, where speaker interaction is monitored through video recording, eye-tracking, and skin conductance measurement, potentially contribute to additional anxiety experienced by speakers, compared to communication in less controlled contexts.

The current exploratory study offers several avenues for future work using skin conductance measures to capture L2 speakers' affective responses dynamically in relation to various speaker and task dimensions. For example, by taking a more controlled, within-group experimental approach, future studies could compare L2 speakers' skin conductance across different conditions, such as in response to different corrective feedback types, when discussing various content in different tasks, or when conversing in different languages, to determine whether and to what extent speakers' affective states predict their performance. In addition, skin conductance measures offer L2 research guided by the dynamic systems theory (e.g., de Bot et al., 2007) to explore how interrelated variables (e.g., affect, speech input and output, nonverbal behavior) and their influence on each other change over time within a complex system (e.g., L2 peer interaction). To extend work on nonverbal

behaviors, future research could also study specific body, facial, or gestural indicators of affective states. For instance, closed and tense body posture, crossed arms, and leaning away tend to indicate anxiety (Burgoon & Koper, 1984), while leaning forward and maintaining a relaxed and open posture characterize nonanxious individuals (Gregersen, 2007). Although a given nonverbal behavior might not convey "a specific message outside of the restrictive environs of the context or culture in which it occurs" (Gregersen, 2005, p. 394) in this study, high arousal behaviors (blinking, glancing away, self-adaption gestures) and low arousal behaviors (nodding) occurred among L2 speakers from various language backgrounds. The extent to which such nonverbal markers of arousal are universal must be examined in future work.

Finally, future work should examine the relationship between autonomic arousal and trait- and state-anxiety (Gregersen et al., 2014; Sevinç, 2018; Witt et al., 2006). Although no such relationship emerged in the present data set, this does not necessarily mean that high arousal was not an indicator of underlying feelings of anxiousness or negative affect during the speaking task, as autonomic arousal does not always correspond to one's subjective experience (Gross, 1998), especially when it is a retrospective judgment. For instance, the speakers' arousal and possible feelings of anxiousness could have been related to their task-specific experiences, such as from negative perceptions of their speech during the task (Gregersen & Horwitz, 2002; Szyszka, 2011), or linked to partner-specific experiences, such as from impressions of their partner's language competence (Heng et al., 2012). Thus, in future research, it would be important to focus on speakers' interpretation of their affective states (e.g., through interviews or retrospective recall procedures) to understand how autonomic arousal during interaction is associated with various affective states (in terms of state-anxiety) and how this association might differ across various speakers and listeners (in terms of trait-anxiety).

Conclusion

Motivated by work on language anxiety (Gregersen, 2005; MacIntyre & Gardner, 1989; Szyszka, 2011), this study extended prior research by investigating L2 speakers' emotional responses during interaction in real time using sensors to record skin conductance, which is an established physiological index of anxiety. The goal of this dynamic approach was to draw connections between autonomic arousal and verbal and nonverbal behaviors shown by L2 speakers in interaction and to explore how arousal may be associated with self-reported anxiety. The main findings showed that high arousals were characterized by significantly more blinks, glances away, and self-adaption gestures, with a trend for more speech disfluencies, while low arousals involved more nodding. Our findings contribute to discussions of how emotions can interact with cognitive processing, such that negative affect may induce a stress response leading to changes in arousal, which in turn may affect language and cognitive functions (Burbridge et al., 2005). In terms of practical considerations, our findings provide implications for both L2 educators and learners by raising their awareness of nonverbal indicators of language anxiety expressed by L2 speakers. If teachers or learners are able to recognize that their interlocutor is blinking

excessively, breaking eye-contact continuously, or repeatedly touching their face or hair during interaction, steps can be taken immediately to address the possible source of this reaction (e.g., by speaking more slowly, changing the topic of discussion, providing encouragement, etc.). In addition, as our results descriptively support prior research suggesting that high arousals are more likely to occur when discussing negatively valenced topics (Burbridge et al., 2005), educators may wish to provide appropriate scaffolding, coping, or exit strategies for learners engaged in interactive tasks that prompt conversations about personal challenges or difficulties. We call for further examination of possible triggers of arousal and its behavioral manifestations, which would be important for educators who seek pedagogical strategies to reduce L2 speakers' negative affect as a way of building their confidence and supporting their L2 development.

References

Anderson, K. J. (1994). Impulsitivity, caffeine, and task difficulty: A within-subjects test of the Yerkes-Dodson law. *Personality and Individual Differences*, **16**, 813–829.

Baker, S. C., & MacIntyre, P. D. (2000). The role of gender and immersion in communication and second language orientations. *Language Learning*, **50**, 311–341.

Blechert, J., Lajtman, M., Michael, T., Margraf, J., & Wilhelm, F. (2006). Identifying anxiety states using broad sampling and advanced processing of peripheral physiological information. *Biomedical Sciences Instrumentation*, **42**, 136–141.

Bradley, M. (2000). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 602–642). Cambridge University Press.

Buck, R. (1984). *The Communication of Emotion*. The Guilford Press.

Burbridge, J. A., Larsen, R. J., & Barch, D. M. (2005). Affective reactivity in language: The role of psychophysiological arousal. *Emotion*, **5**, 145–153.

Burgoon, J., & Koper, R. (1984). Nonverbal and relational communication associated with reticence. *Human Communication Research*, **10**, 601–626.

Caldwell-Harris, C. L., Tong, J., Lung, W., & Poo, S. (2011). Physiological reactivity to emotional phrases in Mandarin–English bilinguals. *International Journal of Bilingualism*, **15**, 329–352.

Clements, K., & Turpin, G. (1996). Physiological effects of public speaking assessed using a measure of palmar sweating. *Journal of Psychophysiology*, **10**, 283–290.

Cramer, P. (2003). Defense mechanisms and physiological reactivity to stress. *Journal of Personality*, **71**, 221–244.

Croft, R. J., Gonsalvez, C. J., Gander, J., Lechem, L., & Barry, R. J. (2004). Differential relations between heart rate and skin conductance, and public speaking anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, **35**, 259–271.

Cutrone, P. (2005). A case study examining backchannels in conversations between Japanese–British dyads. *Multilingua*, **24**, 237–274.

Dawson, M., Schell, A., & Filion, D. (2016). The Electrodermal System. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of Psychophysiology (Cambridge Handbooks in Psychology)* (pp. 217–243). Cambridge: Cambridge University Press.

de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, **10**, 7–21.

Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. Lawrence Erlbaum.

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, **25**, 69–91.

Ekman, P. (1977). Facial Expression. In A. Siegman & S. Feldstein (Eds.), *Nonverbal Communication and Behavior* (pp. 97–126). Lawrence Erlbaum.

Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica*, **1**, 49–98.

Ekman, P., & Friesen, W. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, **29**, 288–298.

Figner, B., & Murphy, R. O. (2011). Using skin conductance in judgment and decision making research. In M. Schulte-Mecklenbeck, A. Kuehberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research* (pp. 163–184). Psychology Press.

Finset, A., Stensrud, T., Holt, E., Verheul, W., & Bensing, J. (2011). Electrodermal activity in response to empathic statements in clinical interviews with fibromyalgia patients. *Patient Education and Counseling*, **82**, 355–360.

Gardner, R., & MacIntyre, P. (1993). A student's contributions to second-language learning. Part II: Affective variables. *Language Teaching*, **26**, 1–11.

Geen, R. G. (1989). Alternative conceptions of social facilitation. In P. B. Paulus (Ed.), *Psychology of group influence* (2nd ed., pp. 15–51). Lawrence Erlbaum.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference 11.0 update* (4th ed.). Allyn & Bacon.

Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P. G., Marias, K., & Tsiknakis, M. (2017). Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control*, **31**, 89–101.

Gregersen, T. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals*, **38**, 388–400.

Gregersen, T. (2007). Breaking the code of silence: A study of teachers' nonverbal decoding accuracy of foreign language anxiety. *Language Teaching Research*, **11**, 209–222.

Gregersen, T. (2009). Recognizing visual and auditory cues in the detection of foreign-language anxiety. *TESL Canada Journal*, **26**, 46–64.

Gregersen, T. S., & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal*, **86**, 562–570.

Gregersen, T., MacIntyre, P.D., & Meza, M.D. (2014). The motion of emotion: Idiodynamic case studies of learners' foreign language anxiety. *The Modern Language Journal*, **98**, 574–588.

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, **74**, 224–237.

Harmon-Jones, E., Gable, P., & Price, T. (2011). Leaning embodies desire: Evidence that leaning forward increases relative left frontal cortical activation to appetitive stimuli. *Biological Psychology*, **87**, 311–313.

Harrigan, J. A. (1985). Self-touching as an indicator of underlying affect and language process. *Social Science and Medicine*, **20**, 1161–1168.

Harrigan, J. A., & O'Connell, D. M. (1996). How do you look when feeling anxious? Facial displays of anxiety. *Personality and Individual Differences*, **21**, 205–212.

Hashemi, M. (2011). Language stress and anxiety among the English language learners. *Procedia – Social and Behavioral Sciences*, **30**, 1811–1816.

Heinz, B. (2003). Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics*, **35**, 1113–1142.

Heng, C. S., Abdullah, A.N., & Yusof, N. B. (2012) Investigating the construct of anxiety in relation to speaking skills among ESL tertiary learners. *3L: The Southeast Asian Journal of English Studies*, **18**, 155–166.

Hirokawa, K., Yagi, A., & Miyata, Y. (2004). Comparison of blinking behavior during listening to and speaking in Japanese and English. *Perceptual and Motor Skills*, **98**, 463–472.

Hirokawa, K., Yamada, F., Dohi, I., & Miyata, Y. (2001). Effect of gender-type on interpersonal stress measured by blink rate and questionnaires: Focusing on stereotypically sex-typed and androgynous types. *Social Behavior and Personality*, **29**, 375–384.

Horwitz, E. K. (1986). Preliminary evidence for the reliability and validity of a foreign language anxiety scale. *TESOL Quarterly*, **20**, 559–562.

Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, **70**, 125–132.

Karson, C. N., Berman, K. F., Donnelly, E. F., Mendelson, W. B., Kleinman, J. E., & Wyatt, R. J. (1981). Speaking, thinking, blinking. *Psychiatry Research*, **5**, 243–246.

King, J. (2013). Silence in the second language classrooms of Japanese universities. *Applied Linguistics*, **34**(3), 325–343.

Knapp, M. L., & Hall, J. A. (2001). *Nonverbal communication in interaction*. Holt, Rinehart and Winston.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, **84**, 394–421.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.

Li, H. Z. (2006). Backchannel responses as misleading feedback in intercultural discourse. *Journal of Intercultural Communication Research*, **35**, 99–116.

Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal*, **92**, 71–86.

MacIntyre, P. D. (1999). Language anxiety: a review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: a practical guide to creating a low-anxiety classroom atmosphere* (pp. 24–45). McGraw-Hill.

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second language learning: Toward a theoretical clarification. *Language Learning*, **39**, 251–275.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, **44**, 283–305.

MacIntyre, P. D., & Mercer, S. (2014). Introducing positive psychology to SLA. *Studies in Second Language Learning and Teaching*, **4**(2), 153–172.

MacIntyre, P. D., & Vincze, L. (2017). Positive and negative emotions underlie motivation for L2 learning. *Studies in Second Language Learning and Teaching*, **7**, 61–88.

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning*, **47**, 265–287.

MacPherson, D., Abur, D., & Stepp, C. (2017). Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load. *Journal of Voice*, **31**, 504.e1–504.e9.

Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, **32**, 21–36.

McDonough, K., & Trofimovich, P. (2019). *Corpus of English as a Lingua Franca Interaction (CELFI)*. Montréal, Canada: Concordia University.

Miragall, M., Vara, D., Cebolla, A., Etchemendy, E., & Baños, R. (2019). Leaning forward to increase approach motivation! The role of joy, exercise, and posture in achieving goals. *Current Psychology*. Published online 14 February 2019.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, **64**, 878–912.

Richmond, V., & McCroskey, J. (2004). *Nonverbal behavior in interpersonal relations (Fifth Edition)*. Pearson Education, Inc.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, **14**, 423–441.

Ruetenik, F. T. (2013). Negotiating effectively and "politely" across Cultures: Comparison of linguistic strategies among international students in Japan. *Journal of Modern Educational Review*, **3**(8), 604–617.

Saito, K., Dewaele, J-M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, **68**, 709–743.

Santos Sierra, A., Sánchez Ávila, C., Guerra Casanova, J., & Bailador del Pozo, G. (2011). A stress-detection system based on physiological signals and fuzzy logic. *IEEE Transactions on Industrial Electronics*, **58**, 4857–4865.

Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning*, **28**, 129–142.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine*, **14**, 410–417.

Sevinç, Y. (2018). Language anxiety in the immigrant context: Sweaty palms? *International Journal of Bilingualism*, **22**, 717–739.

Sevinç, Y., & Dewaele, J. M. (2018). Heritage language anxiety and majority language anxiety among Turkish immigrants in the Netherlands. *International Journal of Bilingualism*, **22**, 159–179.

Shirvan, M. E., & Talebzadeh, N. (2017). English as a foreign language learners' anxiety and interlocutors' status and familiarity: An idiodynamic perspective. *Polish Psychological Bulletin*, **48**, 489–503.

Spielberger, C. D. (1983). *Manual for the State-Trait-Anxiety: STAI (form Y)*. Consulting Psychologists Press.

Szyszka, M. (2011). Foreign language anxiety and self-perceived English pronunciation competence. *Studies in Second Language Learning and Teaching*, **1**, 283–300.

Wallbott, H. G., & Scherer, K. R. (1991). Stress specificities: Differential effects of coping style, gender, and type of stressor on autonomic arousal, facial expression, and subjective feeling. *Journal of Personality and Social Psychology*, **61**, 147–156.

Witt, P. L., Brown, K. C., Roberts, J. B., Weisel, J., Sawyer, C. R., & Behnke, R. R. (2006). Somatic anxiety patterns before, during, and after giving a public speech. *Southern Communication Journal*, **71**, 87–100.