



HAL
open science

A Multimodal Study of How Pronunciation-Induced Communication Breakdowns are Managed During Tandem Interactions

Loulou Kosmala, Céline Horgues, Sylwia Scheuer

► **To cite this version:**

Loulou Kosmala, Céline Horgues, Sylwia Scheuer. A Multimodal Study of How Pronunciation-Induced Communication Breakdowns are Managed During Tandem Interactions. *Research in Language*, 2023, 21 (3), pp.291-312. 10.18778/1731-7533.21.3.05 . hal-04367366

HAL Id: hal-04367366

<https://hal.science/hal-04367366v1>

Submitted on 30 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A MULTIMODAL STUDY OF HOW PRONUNCIATION-INDUCED COMMUNICATION BREAKDOWNS ARE MANAGED DURING TANDEM INTERACTIONS

LOULOU KOSMALA

Université Paris Nanterre;
loulou.kosmala@gmail.com

CÉLINE HORGUES,

Université Sorbonne Nouvelle
celine.horgues@sorbonne-nouvelle.fr

SYLWIA SCHEUER

Université Sorbonne Nouvelle
sylwia.scheuer-samson@sorbonne-nouvelle.fr

Abstract

This paper offers quantitative and qualitative findings from the exploration of communication breakdowns in English tandem interactions, by adopting a multimodal perspective. It focuses on the ways in which pronunciation-induced CBs are managed by language peers in a tandem setting. This study shows cases where it was the non-native participant's output that was the main communicative stumbling block, with a view to reporting on pronunciation-induced breakdowns. More specifically, our analyses target the ways in which CBs are signaled to the interlocutor with different multimodal cues (verbal / vocal / visual). Those pronunciation issues are dealt with in a highly collaborative manner, through multimodal communication strategies, revealing recurrent visual patterns involving different visible body articulators (i.e., the face, the trunk, and the hands) which differ according to participants' status (native versus non-native).

Key words: pronunciation, second language acquisition, multimodality, tandem interactions, communication breakdowns

Introduction

Most L2 pronunciation researchers and pedagogists agree that the focus of pronunciation teaching in ESL and EFL contexts should shift from the unrealistic goal of L2 learners reaching “native-like” pronunciation to that of helping learners

secure a reasonable level of intelligibility in their L2 productions (the ‘intelligibility principle’, e.g. Levis 2005; 2018). Yet, researchers are still trying to explore the nature of the complex relationship between non-native pronunciation and L2 intelligibility especially to Native-Speaker (NS) interlocutors (especially in the English as a Second Language (ESL) context, Derwing & Munro 2009; Trofimovich & Isaacs 2012; Saito et al., 2017) also to Non-Native Speaker (NNS) interlocutors (English Lingua Franca (EFL) or English as an International Language (EIL) contexts (Gardiner & Deterding, 2017; Mauranen, 2006).

In this study, we wish to take a very broad take on the concept of (un)intelligibility of L2 speech through the various phenomena pertaining to miscommunication (including, in the same vein as Mauranen 2006, misunderstanding and non-understanding). We therefore do not restrict our study to the rather narrow definition of intelligibility adopted by Munro & Derwing (1995, p. 76): “intelligibility may be broadly defined as the extent to which a speaker’s message is actually understood by a listener, but there is a no universally accepted way of assessing it”. We refer to the umbrella term intelligibility to also include sub-components such as interpretability (interpreting the meaning behind a word/utterance) and (perceived) comprehensibility (in its definition given by Munro & Derwing, 1995 or Trofimovich & Isaacs, 2012, i.e. interlocutors’ impression of the ease or difficulty with which they understand speech), sub-components which some researchers prefer to treat separately. Furthermore, we will use the term ‘communication breakdown’ (henceforth CB) as a synonym for miscommunication or ‘broad’ (un)intelligibility to refer to any instance where at least one of the participants in a dialogue experiences comprehension difficulties which they attended to collaboratively, although the flow of communication may not be entirely broken as such. It is our interest to look at cases when pronunciation issues effectively hamper NNS-NS communication in real-life face-to-face exchanges and focus our attention on the role of the multimodal resources (especially visual cues) used by NS and NNS interactants when such pronunciation-related CBs arise.

Therefore, this study is grounded in an interactional and multimodal approach to language (Stivers & Sidnell, 2005; Mondada, 2019; Goodwin, 2010) whereby participants display a number of visible and audible communicative cues, such as intonation, facial expressions, manual gestures, and torso movement to co-construct meaning in interaction. In this view, language is regarded as a complex interactive plurichannel system made of vocal-aural and visual-spatial modalities which are continuously deployed together in ways that are relevant to the interaction. Several authors have pointed out the multimodal dimension of tandem interactions, and have studied gestures with regards to corrective feedback, fluency mechanisms, or chains of reference (Kosmala, 2021; Debras & Beaupoil-Hourdel, 2019; Debras et al. 2015, 2020). In second language discourse, hand

gestures are truly relevant as they can be used as a compensatory device with “expressive power and rich semiotic affordances” (Gullberg, 2011, p. 138). Several studies have also pointed out the role of gestures during lexical retrieval (Stam 2001), or the role of representational gestures to elicit lexical help from the interlocutor (Gullberg, 2014). Moreover, manual gestures can be used to signal turn-taking by displaying a request for a turn, or projecting a concept or an action in the beginning of a turn (Mondada, 2007). They can also be used to indicate aspects of an illocutionary force, discourse structure, or a stance (Kendon, 1995, 2004). Other visual features, such as gaze direction and facial expression, also play a key part in interaction (Kendon 1967; Rossano 2013). Visible displays of misunderstanding, for instance, can be marked through specific facial displays, such as frowns (Kaukomaa et al., 2014), or combinations of squints and frowns to mark questionhood (Nota et al., 2021). More recently, a conversation-analytic study focused on the visible bodily resources used by interactants to demonstrate understanding (Jokipohja, 2023) and has revealed recurrent displays: checking understanding was marked by gaze and body suspension, which made the communication breakdown visible (see Mondada, 2011). Trouble cues were also displayed with a number of facial and bodily resources, such as frowning, lowered chin, head tilted to the side, and raised eyebrows.

In line with this research, the aim of the present study is to explore how communication breakdowns are managed outside the classroom in actual face-to-face tandem interactions, from the perspective of both the native and non-native speakers, adopting a multimodal and interactional approach. In particular, we report on pronunciation-induced communication breakdowns and the display of visible cues at different phases of the CB, and the ways in which the mis- or lack of comprehension is signaled by NS to NNS. This paper is structured as follows: we first present our research background which reviews work in pronunciation and miscommunication in L2 speech, the role of tandem-learning settings to explore such phenomena, and research on gesture in L2 productions (Section 1). We then introduce our data and methodology (Section 2), present our results (Section 3), and discuss our findings (Section 4).

1. Research Background

1.1. The study of L2 speech, miscommunication and the role of pronunciation

In the ESL context, a seminal study on NS-NNS miscommunication was carried out by Varonis & Gass (1985b; 1985a) who looked at cases of misunderstandings emerging in phone conversations between NNS and NS engaged in a fairly hierarchical relationship with the NS often having a higher social role (being the shop assistant or teacher they had more authority over the NNS in their customer or student role). The study was conducted in a conversation-analytic framework, and the focus was not particularly on pronunciation features, but they found that

communication issues were more frequent between NNS and NS (compared to the NS-NS) because they did not share as much common ground (gaps in their linguistic and cultural systems) and they hence had to engage in more negotiation routines.

Other researchers adopted a more experimental approach trying to elicit retrospective responses from external NS listeners on audio stimuli produced by NNS (e.g. Munro & Derwing 1995, Trofimovich & Isaacs 2012, Saito et al., 2017). The NS judges participating in these perceptual tests would typically attest to the amount of speech they actually understood (elicited through their orthographic transcriptions of NNS output) and would also self-report on their impressions of the degree (Likert scale) of foreign accentedness of the stimuli on the one hand, and of ease/difficulty which we they think they understood non-native speech (perceived comprehensibility) on the other hand. However, the speech material used was rarely dialogic (mainly picture description monologues in Munro & Derwing, 1995 and Trofimovich & Isaac, 2012, Saito et al., 2017) and the impact of non-native pronunciation on NS's comprehension was primarily investigated through external and retrospective assessment rather than real-life, communicative responses of NS interlocutors engaged in an actual conversation with NNSs.

More recently, researchers' attention has centered on factors ensuring or else hindering mutual intelligibility in NNS-NNS, i.e. in an ELF or EIL context (Mauranen, 2006). Some studies have focused on the impact of pronunciation features in particular (Deterding, 2013; Gardiner & Deterding, 2017; Jenkins, 2000). These studies explore the communicative impact of pronunciation during actual conversations through the identification of detectable misunderstandings and report results that are valid for some specific L1-L2 pairings (NNS's L1 is often one of the South Asian languages). Being based on audio recordings, they can 'only' take into account the contribution of verbal and vocal modalities and *de facto* leave aside the role of visual cues in the triggering, signaling or resolving of "communicative turbulence" (term borrowed from Mauranen, 2006).

Although resorting to multimodal conversational data is often called for, to the best of our knowledge, no L2 intelligibility study has relied on a video-recorded conversational data which would allow for a more comprehensive observation of how pronunciation-induced CB are collaboratively managed in the context of tandem interactions.

1.2. Tandem learning face-to-face exchanges as a particular context for NNS-NS communication

We believe that tandem learning (Brammerts & Calvert, 2003) is a particularly valid language learning and communicative context for exploring the role of pronunciation on mutual intelligibility between NS and NNS. Indeed, it sets NS

and NNS in a low-hierarchy relationship (compared to the asymmetric teacher-learner relation) characterized by some neutralization of proficiency gaps between tandem partners thanks to the language switch occurring in the course of tandem conversations. In that respect, Vassallo & Telles (2006) refer to a ‘symmetrisation’ process operated through the role reversal shaping tandem exchanges, i.e. each participant’s role shifts from being the relative expert to the relative novice depending on whether they speak their L1 or their L2, respectively. Tandem participants engage in authentic, meaningful conversations in the L2 (rather than artificial experimental productive or perceptual tasks). The fact that they are deliberately and consciously committed to a co-learning ‘contract’ with their tandem partner may make them more ready to focus on form when they engage in negotiation of meaning (Varonis & Gass, 1985b; Long, 1991) or Language-Related Episodes (Swain & Lapkin 1998) i.e., when communication issues emerge, than if they were holding a purely informal conversation in their everyday lives. That is the reason our study will rely on a video-recorded corpus of tandem conversations between native speakers of French and native speakers of English (see. Method).

Our previous research on miscommunication in face-to-face English/French tandem conversations has shown that reduced intelligibility is a ‘two-way street’ in such exchanges (Horgues & Scheuer, 2018) as CBs could emerge not only from NNS output but also from NS output (proportion of $\frac{2}{3}$ for the former vs $\frac{1}{3}$ for the latter), so the responsibility does not only lie with the L2 learner’s non-native speech features. Our results have also shown that pronunciation - on its own or combined with other levels such as lexis - ranked among the key factors causing these CBs (Scheuer & Horgues, 2021; Horgues & Scheuer, 2023). These previous studies also pointed to the key role of suprasegmentals (especially syllable division, word stress, rhythm) in triggering CBs produced by these French learners of English, without ruling out the role -albeit less frequent- of segmental issues, especially initial consonants, vowels in monosyllables or involved in vowel reduction of polysyllabic words.

1.3. Gesture use in L2 productions: a look at communication strategies

While extensive research has been conducted in SLA research with regards to CBs, little is known about the visible types of behavior associated with CBs. When production difficulties arise in L2, certain visible strategies are mobilized by NNS to solicit NS’ help, known as multimodal communication strategies (Gullberg, 2011). Based on interactions between Swedish and Dutch learners of L2 French, and French learners of L2 Swedish, Gullberg (2011) distinguished between three major types of difficulties experienced by NNS, mainly lexical, grammatical, and interactional related difficulties. For each type, she investigated whether they would yield different types of visible behavior, involving the hands and gaze. For instance, when dealing with grammatical difficulties, e.g. tense marking,

she showed that learners tried to resolve these problems by using temporal adverbials (“yesterday”, “tomorrow”) or by making use of their gesture space. Regarding lexical difficulties, Gullberg analyzed one example in which the learner did not know the word for “to paint” in her target language (French) and produced a “painting” gesture (holding her right hand in a fist and moving it repeatedly up and down to represent the action of painting) and maintaining her gaze towards her interlocutor (Fig.1). The native-speaker, while providing corrective feedback, repeated the same gesture (Debras et al., 2015, 2020), leading to a joint co-construction.



Fig. 1. Lexical multimodal communication strategy (Gullberg, 2011, p. 140)

Building on the work of Gullberg, Kosmala (2021) examined an instance of pronunciation-related difficulty during an interaction between an American L2 learner of French and a native French speaker where NNS had difficulty with the pronunciation of a word in her target language, and visibly requested assistance by gazing towards her interlocutor, and keeping her hands in the same position (Fig.2). NS understood her partner’s request and took the floor to provide phonological repair.



Fig. 2 pronunciation-related multimodal communication strategy (Kosmala, 2021 p. 258)

In both examples, however, (in Gullberg’s and Kosmala’s), the communication strategies mobilized by the learners are *not* the result of communication breakdowns, and while research on multimodal features of communication strategies or language-related episodes in L2 discourse is gaining more and more

attention in the field of gesture studies overall (e.g. Harrison et al. 2018), studies targeting *communication breakdowns* in relation to gesture have been sparse (e.g. Mondada, 2011; Jokipohja, 2023). Studies on gesture and pronunciation, on the other hand, are slightly more prevalent, but have mainly been examined in experimental conditions so far, as the next section reviews.

1.4. Gesture and the teaching of L2 pronunciation

In the field of SLA, a few experimental studies have been conducted on the use of gestures for facilitating pronunciation features in L2 teaching, involving different perception experiments through video observations. Gluhareva & Prieto (2017) have focused on beat gestures i.e, gestures produced with a batonic up-and-down or back-and-forth movement, generally used to mark prosodic prominence and emphasis. Their study investigated the impact of beat gesture observation in L2 pronunciation learning, more specifically suprasegmentals. They examined the positive effects of a brief language training with or without beat gestures on participants' ratings of accentedness. The participants (learners of English at an undergraduate level) were instructed to watch a training video in which an L2 instructor gave answers to easy and difficult prompts. Participants also provided answers to the prompts, in pre-training and post-training conditions, and their answers were evaluated by native speaker judges. Results showed that beat gesture training significantly improved the participants' ratings of accentedness on discourse-demanding prompts, supporting the role of gestures for enhancing acquisition of suprasegmental aspects in a L2. Similarly, Hoetjes & Maastricht (2020) examined the role of gestures at a segmental level, to facilitate phoneme acquisition. Participants (learners of Spanish) were presented instructions on the production of two Spanish phonemes in three conditions: audio-only, audio-visual with a simple pointing gesture, or audio-visual with more complex, iconic gestures representing the relevant speech articulators. In the pre and post-training conditions, the participants were instructed to read aloud sentences containing the target phonemes, and their production was evaluated by native speakers in a perception study. Results showed that the combination of more complex iconic gestures and speech had a positive effect on L2 training. However, both gesture and phoneme complexity need to be taken into account: while the effect was positive for the less complex phonemes, it was negative for the more complex ones, suggesting that certain gestures are more beneficial for certain phonemes than others. This corroborates Kelly & Lee's (2012) study on the perception of Japanese word pairs by L2 learners which were presented with or without gestures. Their findings showed that when phonetic demands were too high, gestures had a negative effect on learning.

Perception studies have not only been concerned with the beneficial role of hand gestures, but also facial expressions. Tsunemoto et al. (2022) for instance, examined the role of visual cues (including facial expressions) in L2 speech

assessment. Participants (native speakers) were instructed to evaluate the comprehensibility, accentedness, and fluency of L2 English speakers who were narrating a story in three conditions (audio with the speaker's static face, audio with the speaker's static torso but dynamic face, and audio with dynamic torso and face). Their findings showed that raters who had access to the fully dynamic images tended to evaluate the speaker as more comprehensible and less accented. In a similar vein, Wheeler and Saito (2022) focused on "visual speech", which provides information about the place of articulation, with a clear focus on a speaker's mouth, lips, tongue and teeth. In their study, both L1 and L2 speakers were presented with audio-visual stimuli that combined different variables (vowel error, use of visual speech, and presence of an iconic gesture) and were asked to provide an orthographic transcription of the target words. Results showed that while visual speech did not have a significant effect on intelligibility, iconic gestures did, for both groups, and it had a positive effect on L2 listeners when speech contained no errors.

Taken together, these studies have shown the positive effect of facial cues and hand gestures (when phonetic complexity was not high) in pronunciation and their implication for instruction practices. However, these results were obtained through experimental procedures for strictly pedagogical purposes. Less is known about the use of gestures and facial expressions in natural face-to-face interactions in more ecological contexts, such as tandem settings, which involve a lower degree of hierarchy and asymmetry between participants. Conversely, other studies (cf section 1.3) have adopted a more interactional perspective of L2 productions, with a focus on the addressee's contribution and not only the learner's individual production, but without necessarily focusing on pronunciation sequences and with no clear relation to CBs. The aim of this paper is hence to address this gap by looking at the multimodal features of pronunciation-induced communication breakdowns in actual tandem interactions in a non-formal setting. More specifically we intend to explore:

- RQ1: How much CB resolution and pronunciation rectification happens during pronunciation-induced communication breakdowns?
- RQ2: How is the mis- or lack of comprehension visibly manifested and signaled by the native speaker to the non-native speaker?
- RQ3: What types of multimodal strategies are mobilized to collaboratively negotiate meaning or form?

2. Data and method

2.1. The SITAF Corpus

The SITAF corpus (Horgues & Scheuer, 2015) was recorded in 2013 and it consists of video-recorded face-to-face tandem conversations held in English and French by

twenty-one pairs of students. Each tandem pair was composed of a native speaker of English and a native speaker of French. All participants were undergraduate (local and international mobility) students at a French University and their L2 proficiency levels were self-assessed by participants upon registration to the tandem program, with a broad range from intermediate to advanced. The 21 native-English speakers represented a variety of different accents (British, American, Irish, Canadian, Australian English) and the French speaking participants were French students majoring in English studies for the most part. The tandem pairs met weekly for autonomous tandem conversations over the course of one academic semester. They were recorded twice, performing the exact same set of speaking tasks in the two languages (two collaborative game-like activities and one monitored reading task) at the beginning of their tandem experience (session 1) and then three months later (session 2). Game 1 (story-telling) consists in the L2 speaker narrating a personal story integrating three lies with their interlocutor trying to elucidate the lies through asking questions. Game 2 (debating) consisted in the two interactants giving their opinion about a set controversial topic to then decide on the degree of like-mindedness between them. It is characterized by overall more symmetry between the NS and the NNS's speaking time and contribution to discourse than Game 1.

In this paper, we only study the two semi-spontaneous activities (Game 1 and 2) held in English in the two sessions (15 pairs from the corpus, about 5.5 hours of video-recorded speech) and we analyzed CBs emerging from the output of the NNS or L2 speaker only (French learner of L2 English). CBs arising from NS output are therefore outside the scope of this study.

2.2. CB identification

We define a case of communication breakdown (CB) as a moment of misunderstanding, non-understanding or comprehension issue on the part of the interlocutor (their comprehension trouble being manifested verbally, vocally and/or visually). This perspective is in line with Nakahama et al. (2001) in relying on the observation of the *interlocutor's* verbal, vocal and visual reactions to problematic NNS speech, that is when the recipient *demonstrably* had difficulty, or was incapable of, grasping the meaning of an utterance as seemingly intended by the speaker. We focused on miscommunication attended to through interactional work (i.e. collaboratively managed in the following speech turns by the main speaker and the interlocutor). This therefore led us to exclude instances of communicative issues prevented from happening (Preemptive LREs; Strawbridge, 2021), or cases where the interlocutor may have self-resolved potential processing issues on their own, or even avoided engaging in miscommunication management ("let-it pass" strategy). To identify CB cases, the second and third authors independently inspected (auditorily and visually) all video-recorded semi-spontaneous sequences in English (2 games, 2 sessions).

Their identification coincided for about 80% of CB cases, and they further reviewed and discussed multimedia sequences of the remaining 20% of the less consensual cases to reach a joint decision.

We define a pronunciation-induced CB as a CB where pronunciation is hypothesized to be *the* factor or *one of* the factors (in combination with other factors such as lexical/morphosyntactic, pragmatic, cultural) triggering miscommunication. We further differentiated between “only factor” (when pronunciation is hypothesized to be the main trigger of CB) and “mixed” factor (when it is mixed with other factors, such as lexical, grammatical, pragmatic etc).

CBs sequences were analyzed in their chronological and co-constructed structure (Varonis & Gass, 1985a): *Trigger* (Speaker), *Indicator or Signal* (Hearer), *Response* (Speaker), *Reaction to Response* (Hearer), as in the example below:

NNS: [There were] Hens (['ɛns])	[TRIGGER]
NS: (<i>silence at first</i>) Ants?	[SIGNAL]
NNS: (<i>makes a clucking sound and a representational gesture of flapping wings</i>)	[RESPONSE]
NS: Chickens?	[REACTION TO RESPONSE]

2.3. Annotation of visual features

For each main phase of the sequence (trigger, signal, response), different visible body articulators (eyes, face, head, hands, and body) were analyzed for both the native and non-native speaker in the position of hearer or speaker:

- Gaze direction: as either “averted” (from interlocutor), “towards interlocutor”, “paper” (towards the piece of paper the participants were given for the task)
- Hand gesture: If no gesture was produced, it was annotated as “no gesture” i.e., when the hands are not moving and remain still in a resting position such as the lap. If a gesture was fully produced, then its function was annotated, following previous functional gesture typologies (Graziano & Gullberg, 2018; Kendon, 2004) distinguishing between “referential” (hand gestures related to discourse content, used to depict, iconically represent, or refer to an object, person, or action) and “pragmatic” (hand gestures related to discourse itself, to mark speech emphasis, indicate a speech act, a stance, or regulate turn-taking).
- Facial expressions: “neutral”, “frown” “smile” “raised eyebrows”
- Head movements: “still” “nod” “shake” “tilt”
- Body configurations: “still”, “lean forward” “straighten up”

To address RQ2 (see section 1.4) we also focused on NS' signaling phase more specifically, and differentiated between cases where NS displayed either: audible cues (i.e., vocal and verbal), visual cues (facial expressions, hand gestures, body movement etc.), or bimodal cues (both audible and visual together) to signal misunderstanding to NNS.

3. Results and discussion

We follow up on our previous study (Horgues & Scheuer, 2023) which reported 60 cases of CBs in the L2 speakers' output in English, among which 37 were judged to be pronunciation-induced, as pronunciation issues were hypothesized to be the factor or one of the factors triggering miscommunication. Out of these 37 cases, one third were mostly segmental, while two thirds had a suprasegmental layer (syllable division, word stress realization, rhythm). The majority of pronunciation-induced CBs were categorized as mixed (N=22/37), as pronunciation mostly combined with other factors (most frequently lexical or grammatical issues).

To answer RQ1, almost all of these CBs (32/37) were found to be successfully resolved in the next couple of turns, and 2/37 CBs were resolved with a delay (2'30 mins and 8'30 mins). However, these NNS mispronunciations did not systematically lead the NS to produce the target pronunciation form during the CB management sequence (20/37 cases), and they very rarely resulted in the L2 learner rectifying their initial mispronunciation (only 7/37 cases). This may suggest that tandem partners wish to focus more on meaning than on accuracy of form, in other words, on getting understood than on rectifying pronunciation explicitly, so long as the communication breakdown is resolved. One enlightening example is reported below:

Excerpt 1.

- 1 NNS: and my uncle has, ah, he has a yellow ferrari [**ferra'ri*]
[*NS blank face*]
- 2 NNS: fe-rra-ri, yes [**ferra'ri*]
- 3 NS: he has a (..) ? Say that ag-? [*frowns + leans forward*]
- 4 NNS: a car, a yellow [*frowns + fingers points down*]
- 5 NS : oh a yellow ferrari [*reinforcing word stress through vocal emphasis and head nods/beats*]
- 6 NNS : a car [*fingers points down*]
- 7 NS: ok, s-
- 8 NNS: yes?
- 9 NS: no, never mind, I was just gonna get you to repeat, but that's ok.
- 10 NNS : and.. [*pursues story*]

In this excerpt, NNS produces the word “ferrari” using an incorrect stress pattern with a right shift: she places the stress on the final syllable [**ferra'ri*] instead of the penultimate [*fe'rrari*]. NS’ reaction is a puzzled, blank face, also known as the ‘freeze-look’ (Manrique & Enfield, 2015). This absence of bodily movement can be associated with a CB signal, taken up by interlocutors who modify their preceding turn to resolve the communication problem (Mondada, 2011). Indeed, NNS repeats her initial output (1.2) by isolating every single syllable (“fe-rra-ri”) but without resolving the CB nonetheless. NS then explicitly expresses his misunderstanding (1.3), using verbal and bodily cues: he asks her to repeat (“say that ag-?”) while frowning and leaning forwards. These are recurrent visible cues associated with the response phase (see Tables 2 and 3). Once the CB is resolved, it is marked by a change-of-state token “oh” (Heritage 2002) and a corrective feedback provision (1.5). NS then closes the CB sequence with a clear withdrawal (1.9) from an explicit correcting posture which he prefers not to engage in further (beyond his implicit simple recast “oh a yellow ferrari”) so as to prioritize the flow of communication over focus on pronunciation rectification.

Focusing now on overall visual-gestural features (RQ2 and RQ3) associated with each stage of the communication breakdown sequence (based on the 37 sequences identified in total), results show different patterns of visible behavior for NS and NNS. First, if we look at NS’ displays of comprehension difficulty (RQ2) during the signaling phase (Fig. 3) results show that in a majority of cases (N=23/37) NS mobilized both audible (verbal-vocal) and visual cues. This points to the multimodality of communication breakdowns, which should not be overlooked. In this view, signaling communication breakdowns is an inherently visible and embodied practice, which cannot easily be displayed with words or intonation only, in line with Jokipohja (2023).

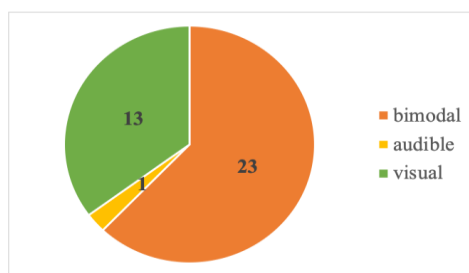


Fig. 3 Overview of miscommunication displays by NS during the signaling phase

Table 1 illustrates the distribution of hand gestures across the three different phases for the two speakers. It shows that hand gestures were more likely to be produced during the trigger phase (coming from NNS’ verbal output which caused the pronunciation-induced CB) and NNS’ response to NS’ signal than during the signal itself.

Table 1. Hand gesture behavior during CB sequence

	COMMUNICATION BREAKDOWN SEQUENCE			
	Trigger	Signal	Response ¹	
	NNS	NS	NNS	NS
pragmatic gesture	10	1	14	0
referential gesture	6	0	11	6
no gesture	21	36	12	31

No clear differences are found between pragmatic and referential gestures for both the trigger and response phase for NNS. What is relevant to note, however, is that NS almost never produced hand gestures in the signal phase, suggesting that gestures may not be the most salient visible expression of CB on the part of the NS interlocutor. In the response phase, on the other hand, only referential gestures were used by NS, and no pragmatic ones. Even though NSs do not show a tendency to gesture in the response phase overall (only 6 cases out of 37), when they do gesture, the latter are associated with discourse content, perhaps as a way to resolve the CB or provide feedback on NNS' initial output. This is illustrated in the following example:

Excerpt 2.

- 1 NNS: and it was a castle and you have (.) goose (.) goose inside
[gazes at NS; frowns, places hands in the gesture space]
- 2 NS: goose? *[frowns, head tilt]*
- 3 NNS: goose euh hens *[represents the action of flapping one's wings]*
- 4 NS: oh goose! *[raised eyebrows]*
- 5 NS: ok ok.
- 6 NS: uh you can say for (.) um there's one than more goose (.) they're geese.
[right hand curved into a U shape moved to the side + raised eyebrows]
- 7 NNS: geese. *[very stretched lips + vocal reinforcement of tense /i:/]*
- 8 NS: geese yeah it changes to 'ee' in the middle *[spells the vowel digraph in the air, looks at NNS cf Fig. 4]*
- 9 NNS yeah
- 10 NNS: so geese *[draws the letters in the air, looks up; cf Fig. 4]*
- 11 NS: they're geese.
- 12 NNS: geese ok. *[vocal reinforcement of articulatory tension]*

¹ The response phase includes both NNS' initial response, and NS' reaction to response.

This excerpt shows an example of mixed trigger, as the CB was not only caused by pronunciation but also morphosyntactic factors, more specifically the use of irregular plural form (from ‘goose’ to ‘geese’). NNS first introduces the word ‘goose’ (1.1) twice, by placing both her hands in the left periphery of her gesture space, as to introduce the discourse referent. The fact that her gaze is directed towards NS and that she is frowning may indicate that she is unsure of the word herself, and is perhaps seeking agreement or confirmation from NS. Her gesture is in fact held in the same position (as she initiates the first occurrence of the word ‘goose’ and the second one) until NNS’ reaction, which may illustrate a similar multimodal communication strategy as the one documented in Kosmala (2021) (cf section 1.3, and also Jokipohja, 2023). NS’ reaction is marked by his visible body behavior: he repeats the target word with a rising intonation, frowns, and tilts his head downward. In the response phase, NNS provides another word “hen”, and mimics the action of flapping one’s wings to make herself understood (1.3), which leads to NS’ understanding of the word, marked by ‘oh’, and raised eyebrows (1.4). The CB then leads to a more pedagogical-oriented type of resolution whereby NS, like a language instructor, takes some time to explain the plural form of ‘goose’ using his hands: he illustrates the shift from plural to singular with a specific U handshape by moving it from left to right (1.6). Then NNS repeats the target word (1. 7) in a hyperarticulated way, using visual speech, spreading her lips somewhat exaggeratedly, and showing her teeth. NS then relies on graphophonemics to illustrate the word’s spelling by writing the vowel digraph ‘ee’ in the air; this referential representational gesture is then repeated by NNS (Fig. 4) to help her visualize the word and perhaps better remember its plural form.

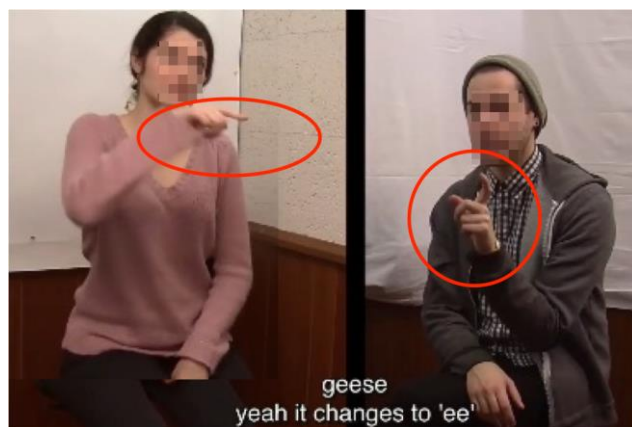


Fig. 4. Gestures related to graphophonemics (NNS on the left, NS on the right)

It should be noted however, that this type of pronunciation-related gesture and the presence of visual speech was extremely rare, overall, and is thus not representative of the whole sample. What is still relevant to note, is that the CB sequence led to NNS’ rectification of the plural form (from ‘goose’ to ‘geese’), so the gesture may have had an impact on NNS’ production, confirming previous

experimental studies. Conversely, in the next example, taken from another dyad, NS' posture is much less pedagogical-oriented, and he does not produce gestures to resolve the CB or provide feedback:

Excerpt 3.

- 1 NNS: euh the extra money is for euh the leisure ([*le'ʒɜ:R]) and the different activities.
- 2 NS: for what? [*gaze on NNS; head tilt; body moves forward*]
- 3 NNS: for the leasures ([*le'ʒɜ:RZ]) [*gazes at NS*]
- 4 NS: leisure? ([li:ʒəɹ]) [*gazes at NNS*]
- 5 NNS: leisure ([li:ʒəɹ]) sorry [*gazes at NS*]
- 6 NNS: um and for it's it's better to use this extra money for the leisure ([*le'ʒɜ:R]) and the quality of food.

In this excerpt, NS is a speaker of North American English, and the CB seems to be determined by language pronunciation norms. NNS first pronounces the word 'leisure' approximating the vowel quality of British pronunciation in the first syllable (l.1) but using an incorrect stress pattern (with another rightward shift and no vowel reduction). The misunderstanding may hence be caused by both segmental (for her American interlocutor who probably expected a long /i:/) and suprasegmental factors with the stress pattern, but also by the incorrect plural form (on line 3). Once again, NS' reaction in the signaling phase is visibly displayed by his bodily behavior with a head tilt and his body moving forward accompanying the verbal response "for what?" (l. 2). Unlike the previous example, NS simply repeats the target word with a different pronunciation (and using a different reference accent) but without lingering on it or providing any clear explanation through hand gesture (l.4). NNS first repeats the target word, using the American native speaker's pronunciation, apologizes for her own pronunciation, but returns to her initial pronunciation in the following turn (l. 6).

Turning now to facial expressions and gaze behavior (Tables 2 and 3), results show that frowns were the most common facial cues used by both NNS in the trigger phase and NS in the signaling phase, but presumably for two different reasons: while frowning may indicate a state of uncertainty and encoding difficulty on NNS' part, it displayed NS' mis- or lack of understanding. In addition, in the response phase, smiling was found to be a frequent visible cue on NS' part (N=20/37).

Table 2. Facial expressions during CB sequence

	COMMUNICATION BREAKDOWN SEQUENCE			
	Trigger	Signal	Response	
	NNS	NS	NNS	NS
frown	14	15	6	1
neutral	13	14	18	15
raised eyebrows	2	4	4	1
smile	8	4	9	20

Mutual gaze was also a common feature in all phases of the CB for both speakers, accounting for the intersubjective nature of CBs and the empathic relationship between tandem partners.

Table 3. Gaze direction during CB sequence

	COMMUNICATION BREAKDOWN SEQUENCE			
	Trigger	Signal	Response	
	NNS	NS	NNS	NS
averted	2	0	8	10
towards interlocutor	31	34	27	27
towards paper	4	3	2	0

Results included in Table 4 also show that NS sometimes leaned forward in the signaling phase (N=11), which is consistent with previous research showing that forward torso leans occur when speakers wish to seek clarification (Holler, 2022), which is a clear characteristic of interactions held face-to-face. Similarly, when CBs occur, NSs' displays of misunderstanding preceding the response phase may further initiate a clarification request.

Table 4. Body positions during CB sequence

	COMMUNICATION BREAKDOWN SEQUENCE			
	Trigger	Signal	Response	
	NNS	NS	NNS	NS
lean forward	0	11	5	1
still	37	26	26	26
straighten up	0	0	6	10

Instances of leaning forward in the signal phase were illustrated in Excerpts 1 and 3. Head nods (cf Table 5) were also predominant in the response phase on NS' part (N=32/37) which is consistent with previous studies conducted on understanding (e.g. Helmer et al., 2021). In particular, turn-final head nods can make affiliative displays relevant (Aoki, 2011).

Table 5. Head movements during CB sequence

	COMMUNICATION BREAKDOWN SEQUENCE			
	Trigger	Signal	Response	
	NNS	NS	NNS	NS
nod	6	9	2	32
shake	0	1	5	1
still	30	17	30	4
tilt	1	10	0	0

In summary, we can find recurrent visible patterns associated with different stages of the CB: while NNSs used a great deal of hand gestures during the trigger and the response phases to accompany the initial verbalizing and rephrasing of their output, NS rarely did it during the signaling and response phase (with the exception of a few referential gestures, one of which is illustrated in Excerpt 2.). Instances of frowning, body leaning forward, and head tilts were quite common in the signaling phase, as opposed to smiling and nodding in the response phase (for NSs). All these visible cues were found to be relevant for the progression of the exchange, as they enabled the tandem partners to resolve the CBs successfully through interactional work.

4. Conclusion

Pronunciation issues are a key factor in causing miscommunication emerging from L2 speech. However, this study has shown that pronunciation-induced CBs do not stem from pronunciation issues in isolation, as they often combine with lexical, pragmatic, morphosyntactic and/or cultural factors. This study has revealed several multimodal (vocal, verbal, and visual-gestural) components of communication breakdowns, both on NS and NNS' part, at different stages of the CB sequence, focusing on the interactive nature of CB management. Our main interest was to see how pronunciation-induced CBs get managed by tandem partners (non-teacher participants) in a non-formal setting, outside the classroom. Our results have shown that tandem partners tend to focus more on getting understood than on the NS correcting pronunciation or the NNS modifying their initial output. Mispronunciations were in fact rarely rectified by the L2 learner (only 7 cases out of 37), but a large majority of cases led to a resolution of the communication breakdown (RQ1). This points to the interactive and non-hierarchical nature of tandem interactions, based on friendliness, solidarity, and reciprocity (Brammerts & Calvert, 2003), unlike teacher-student interactions in institutional settings.

In addition, the mis- or lack of comprehension signaled by NS (RQ2) relied extensively on bimodal cues combining the verbal-vocal and visual-spatial modalities. It would be enlightening in future research to explore how these different cues are combined simultaneously in different configurations, forming “multimodal gestalts” (Mondada, 2018).

As an answer to RQ3, CBs were found to be collaboratively managed with an extensive use of hand gestures by NNS during both the trigger and response phases, and with displays of joint attention by both parties with mutual gaze, as well as vocal-verbal and visual tokens of understanding (smiling, nodding, ‘oh’ tokens etc.). In particular, a number of recurrent visual-gestural patterns were observed for NS in the signaling phase, in line with Jokipohja (2023) with instances of frowning, body leaning forward and head tilts. At a more qualitative level, different strategies were mobilized by NSs to manage the CB. In Excerpt 1, NS explicitly refrained from displaying an explicit corrective posture, while in Excerpt 2 the CB was managed collaboratively with a pedagogical-oriented resolution, which had a clear (short-term) impact on NNS' production. In Excerpt 3, however, the CB was not managed so collaboratively as the native speaker provided minimal feedback, so the CB was quickly resolved, but with no clear impact on NNS' output.

Finally, it should be noted that our analyses are based on a very small sample (only 37 tokens) taken from the English data only. It thus would be useful in the future to compare other types of CBs (following Horgues & Scheuer, 2018) and their visual affiliates in both English and French, as well as to explore in more depth the complex relationship between CB and corrective feedback at a multimodal level.

References

- Aoki, H. 2011. *Some Functions of Speaker Head Nods*. In J. Streeck, Ch. Goodwin and C. Le Baron (eds), *Embodied Interaction: Language and Body in the Material World*. Cambridge, UK: Cambridge University Press. 93–105.
Pronunciation: Issues and Practices, 97-107
- Brammerts, H. and Calvert, M. 2003. *Learning by Communication in Tandem. Autonomous Language Learning in Tandem..* Lewis, T. and L. Walker (eds). Sheffield: Academy Electronic Publication
- Debras, C., C. Horgues, and S. Scheuer. 2015. *The Multimodality of Corrective Feedback in Tandem Interactions*. *Procedia-Social and Behavioral Sciences* 212: 16–22.
<https://doi.org/10.1016/j.sbspro.2015.11.292>
- Debras, C. and P. Beaupoil-Hourdel. 2019. *Gestualité et Construction Des Chaînes de Référence Dans Un Corpus d'interactions Tandem*. *Cahiers de Praxématique* 72: 1-25
<https://doi.org/10.4000/praxematique.5576>
- Debras, C. Pauline Beaupoil-Hourdel, A. Morgenstern, C. Horgues and S. Scheuer. 2020. *Corrective Feedback Sequences in Tandem Interactions: Multimodal Cues and Speakers' Positionings*. In S. Raineri, M. Sekali, and A. Leroux (eds), *Linguistic Correction/Correctness*. Paris Nanterre: Presse de Paris Nanterre.
- Derwing, T. M. and M. J. Munro. 2009. *Putting Accent in Its Place: Rethinking Obstacles to Communication*. *Language Teaching* 42 (4): 476–90.
<https://doi.org/10.1017/S026144480800551X>
- Deterding, D. 2013. *Misunderstandings in English as a Lingua Franca: An Analysis of ELF Interactions in South-East Asia*. Berlin, Boston: De Gruyter Mouton.
<https://doi.org/10.1515/9783110288599>
- Gardiner, I. A. and D. Deterding. 2017. *Pronunciation and Miscommunication in ELF Interactions: An Analysis of Initial Clusters*. In J. Jenkins, W. Baker, and M. Dewey (eds) *The Routledge Handbook of English as a Lingua Franca*. London and New York: Routledge. 224–32
<https://doi.org/10.4324/9781315717173-19>
- Gluhareva, D. and P. Prieto. 2017. *Training with Rhythmic Beat Gestures Benefits L2 Pronunciation in Discourse-Demanding Situations*. *Language Teaching Research* 21 (5): 609–31.
<https://doi.org/10.1177/1362168816651463>
- Goodwin, C. 2010. *Multimodality in Human Interaction*. *Calidoscopio* 8 (2): 85-88
<https://doi.org/10.4013/cld.2010.82.01>
- Graziano, M. and M. Gullberg. 2018. *When Speech Stops, Gesture Stops: Evidence from Developmental and Crosslinguistic Comparisons*. *Frontiers in Psychology* 9: 860-879.
<https://doi.org/10.3389/fpsyg.2018.00879>
- Gullberg, M. 2011. *Multilingual Multimodality: Communicative Difficulties and Their Solutions in Second-Language Use*. In J. Streeck, Ch. Goodwin and C. LeBaron (eds), *Embodied Interaction: Language and Body in the Material World*. Cambridge, UK: Cambridge University Press: 137–51.
- . 2014. *Gestures and Second Language Acquisition*. In C. Müller, A. Cienki, S. H. Ladewig, D. McNeill, and S. Tessendorf (eds) *Body- Language- Communication: An International Handbook on Multimodality in Human Interaction*. Berlin, Boston: Mouton de Gruyter: 1868-1875
- Harrison, S., S. Adolphs, M. G. Dowens, P. Du and J. Littlemore. 2018. *All Hands on Deck. Negotiation over Gesture Forms in Collaborative Discourse*. *Lingua* 207: 1–22.
<https://doi.org/10.1016/j.lingua.2018.02.002>
- Helmer, H., E. Betz, and A. Deppermann. 2021. *Coordination of OKAY, Nods, and Gaze in Claiming Understanding and Closing Topics*. In E. Betz, A. Deppermann, L.Mondada, and M. Sorjonen

- (eds) *OKAY across Languages: Toward a Comparative Approach to Its Use in Talk-in-Interaction*. Amsterdam: Benjamins: 363–93.
<https://doi.org/10.1075/slsi.34.12hel>
- Heritage, J. 2002. 'Oh'-Prefaced Responses to Assessments: A Method of Modifying Agreement/Disagreement. In C. Ford, B Fox, and S Thompson (eds). *The Language of Turn and Sequence*, edited by New York: Oxford University Press: 196-224.
<https://doi.org/10.1093/oso/9780195124897.003.0008>
- Hoetjes, M. and L. V. Maastricht. 2020. *Using Gesture to Facilitate L2 Phoneme Acquisition: The Importance of Gesture and Phoneme Complexity*. *Frontiers in Psychology* 11: 1-25
- Holler, J. 2022. *Visual Bodily Signals as Core Devices for Coordinating Minds in Interaction*. *Philosophical Transactions of the Royal Society* 377: 1–15.
<https://doi.org/10.1098/rstb.2021.0094>
- Horgues, C. and Scheuer, S. 2015. *Why some things are better done in tandem?* In J. A. Mompeán and J. Fouz-González (eds.), *Investigating English pronunciation: Trends and directions*. Basingstoke and New York: Palgrave Macmillan. 47-87.
https://doi.org/10.1057/9781137509437_3
- Horgues, C. and Scheuer, S. 2018. "What are you not understanding?" *Managing communication breakdowns in native/non-native English/French tandem interactions*, in Babatsouli, E. (eds.) *Crosslinguistic research in monolingual and bilingual speech*. 129-146.
- Horgues, C. and Scheuer, S. 2023. *How L2 English pronunciation contributes to miscommunication in two French/English tandem conversation tasks*. In A. Henderson and A. Kirkova-Naskova (eds). *Proceedings of the 7th International Conference on English*
- Jenkins, J. 2000. *The Phonology of English as an International Language*. Oxford, UK: University Press.
- Jokipohja, A. 2023. *Demonstrating and Checking Understanding – Bodily-Visual Resources in Action Formation and Ascription*. *Journal of Pragmatics* 210: 87–108.
<https://doi.org/10.1016/j.pragma.2023.03.018>
- Kaukomaa, T., A. Peräkylä, and J. Ruusuvuori. 2014. *Foreshadowing a Problem: Turn-Opening Frowns in Conversation*. *Journal of Pragmatics* 71: 132–47.
<https://doi.org/10.1016/j.pragma.2014.08.002>
- Kelly, S. D., and A. L. Lee. 2012. When Actions Speak Too Much Louder than Words: Hand Gestures Disrupt Word Learning When Phonetic Demands Are High. *Language and Cognitive Processes* 27 (6): 793–807.
<https://doi.org/10.1080/01690965.2011.581125>
- Kendon, A. 1967. *Some Functions of Gaze-Direction in Social Interaction*. *Acta Psychologica* 26: 22–63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)
- . 1995. *Gestures as Illocutionary and Discourse Structure Markers in Southern Italian Conversation*. *Journal of Pragmatics* 23 (3): 247–79. [https://doi.org/10.1016/0378-2166\(94\)00037-F](https://doi.org/10.1016/0378-2166(94)00037-F)
- . 2004. *Gesture: Visible Action as Utterance*. Cambridge, UK: Cambridge University Press.
- Kosmala, L. 2021. *A multimodal contrastive study of (dis)fluency across languages and settings: towards a multidimensional scale of inter-(dis)fluency*. PhD thesis. Université Sorbonne Nouvelle, France.
- Levis, J. M. 2005. *Changing Contexts and Shifting Paradigms in Pronunciation Teaching*. *TESOL Quarterly* 39 (3): 369–77. <https://doi.org/10.2307/3588485>
- . 2018. *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge, UK: Cambridge University Press.
- Long, M. 1991. *Focus on Form: A Design Feature in Language Teaching Methodology*. In K. de Bot, R.B. Ginsberg, and C. Kramsch (eds) *Foreign Language Research in Cross-Cultural Perspective*. Amsterdam: John Benjamins: 39–52. <https://doi.org/10.1075/sibil.2.07lon>

- Manrique, E. and N. J. Enfield. 2015. *Suspending the next Turn as a Form of Repair Initiation: Evidence from Argentine Sign Language*. *Frontiers in Psychology* 6 (1326): 1–21. <https://doi.org/10.3389/fpsyg.2015.01326>
- Mauranen, A. 2006. *Signaling and Preventing Misunderstanding in English as Lingua Franca Communication*. *International Journal of the Sociology of Language* 177: 123–50. <https://doi.org/10.1515/IJSL.2006.008>
- Mondada, L. 2007. *Multimodal Resources for Turn-Taking: Pointing and the Emergence of Possible next Speakers*. *Discourse Studies* 9 (2): 194–225. <https://doi.org/10.1177/1461445607075346>
- . 2011. *Understanding as an Embodied, Situated and Sequential Achievement in Interaction*. *Journal of Pragmatics* 43: 542–522. <https://doi.org/10.1016/j.pragma.2010.08.019>
- . 2018. *Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality*. *Research on Language and Social Interaction* 51 (1): 85–106. <https://doi.org/10.1080/08351813.2018.1413878>
- . 2019. *Contemporary Issues in Conversation Analysis: Embodiment and Materiality, Multimodality and Multisensoriality in Social Interaction*. *Journal of Pragmatics* 145: 47–62. <https://doi.org/10.1016/j.pragma.2019.01.016>
- Munro, M. J., and T.M. Derwing. 1995. *Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners*. *Language Learning* 45 (1): 73–97.
- Nakahama, Y., Andrea T., and L. Van Lier. 2001. *Negotiation of Meaning in Conversational and Information Gap Activities: A Comparative Discourse Analysis*. *TESOL Quarterly* 35 (3): 377–405. <https://doi.org/10.2307/3588028>
- Nota, N., J. P. Trujillo, and J. Holler. 2021. *Facial Signals and Social Actions in Multimodal Face-to-Face Interaction*. *Brain Sciences* 11 (8): 1000: 1017. <https://doi.org/10.3390/brainsci11081017>
- Rossano, F. 2013. *Gaze in Conversation*. In J. Sidnell and T. Stivers (eds) *The Handbook of Conversation Analysis*, Malden, MA: Blackwell Publishing: 308-329. <https://doi.org/10.1002/9781118325001.ch15>
- Saito, K., P. Trofimovich, and T. Isaacs. 2017. *Using Listener Judgments to Investigate Linguistic Influences on L2 Comprehensibility and Accentedness: A Validation and Generalization Study*. *Applied Linguistics* 38 (4): 439–62.
- Scheuer, S., and Horgues, C. 2021. *Corrective feedback and unintelligibility: Do they work in tandem during tandem interactions?*. In A. Kirkova-Naskova, A. Henderson & J. Fouz-González (eds), *English Pronunciation Instruction: Research-based Insights*, Amsterdam, The Netherlands: John Benjamins. 223-252, <https://doi.org/10.1075/aals.19.10sch>
- Stam, G. 2001. *Lexical Failure and Gesture in Second Language Development*. In C. Cavé, I. Guaïtella, and S. Santi (eds), *Oralité et Gestualité: Interactions et Comportements Multimodaux Dans La Communication*, Paris: L’Harmattan: 271–75.
- Stivers, T. and Jack Sidnell. 2005. *Introduction: Multimodal Interaction*. *Semiotica* 2005: 1–20. <https://doi.org/10.1515/semi.2005.2005.156.1>
- Strawbridge, T. 2021. *Modern Language: Interaction in Conversational NS-NNS Video SCMC ETandem Exchanges*. *Language Learning & Technology*, 2: 94–110.
- Swain, M. and S. Lapkin. 1998. *Interaction and Second Language Learning: Two Adolescent French Immersion Students Working Together*. *The Modern Language Journal* 82 (3): 320–37. <https://doi.org/10.1111/j.1540-4781.1998.tb01209.x>
- Trofimovich, P. and T. Isaacs. 2012. *Disentangling Accent from Comprehensibility*. *Bilingualism: Language and Cognition* 15 (4): 905–16. <https://doi.org/10.1017/S1366728912000168>
- Tsunemoto, A., R. Lindberg, P. Trofimovich, and K. McDonough. 2022. *Visual Cues and Rater Perceptions of Second Language Comprehensibility, Accentedness, and Fluency*. *Studies in Second Language Acquisition* 44 (3): 659–84. <https://doi.org/10.1017/S0272263121000425>
- Varonis, E. M. and S. Gass. 1985a. *Non-Native/Non-Native Conversations: A Model for Negotiation of Meaning*. *Applied Linguistics* 6 (1): 71–90. <https://doi.org/10.1093/applin/6.1.71>

-
- Varonis, E. M. and S. Gass. 1985b. *Miscommunication in Native/Nonnative Conversation*. *Language in Society* 14 (3): 327–43. <https://doi.org/10.1017/S0047404500011295>
- Vassallo, M. L., and J. A. Telles. 2006. *Foreign Language Learning In-Tandem: Theoretical Principles and Research Perspectives*. *The Specialist* 27 (1).
- Wheeler, P. and Kazuya S. 2022. *Second Language Speech Intelligibility Revisited: Differential Roles of Phonological Accuracy, Visual Speech, and Iconic Gesture*. *The Modern Language Journal* 106 (2): 429–48. <https://doi.org/10.1111/modl.12779>