





## Sequentiality, Mutual Visibility, and Behavioral Matching: Body Sway and Pitch Register During Joint Decision Making

Melisa Stevanovic, Tommi Himberg, Maija Niinisalo, Mikko Kahri, Anssi Peräkylä, Mikko Sams & Riitta Hari


To cite this article: Melisa Stevanovic, Tommi Himberg, Maija Niinisalo, Mikko Kahri, Anssi Peräkylä, Mikko Sams & Riitta Hari (2017) Sequentiality, Mutual Visibility, and Behavioral Matching: Body Sway and Pitch Register During Joint Decision Making, Research on Language and Social Interaction, 50:1, 33-53, DOI: [10.1080/08351813.2017.1262130](https://doi.org/10.1080/08351813.2017.1262130)

To link to this article: <http://dx.doi.org/10.1080/08351813.2017.1262130>

 View supplementary material [↗](#)

 Published online: 03 Jan 2017.

 Submit your article to this journal [↗](#)

 Article views: 141

 View related articles [↗](#)

 View Crossmark data [↗](#)

 Citing articles: 1 View citing articles [↗](#)

## Sequentiality, Mutual Visibility, and Behavioral Matching: Body Sway and Pitch Register During Joint Decision Making

Melisa Stevanovic<sup>a</sup>, Tommi Himberg<sup>b</sup>, Maija Niinistö<sup>b</sup>, Mikko Kahri<sup>a</sup>, Anssi Peräkylä<sup>a</sup>,  
Mikko Sams<sup>b</sup>, and Riitta Hari<sup>c</sup>

<sup>a</sup>Finnish Centre of Excellence in Research on Intersubjectivity in Interaction, University of Helsinki, Finland;

<sup>b</sup>Department of Neuroscience and Biomedical Engineering, Aalto University, Finland; <sup>c</sup>Department of Art, Aalto University, Finland

### ABSTRACT

We studied behavioral matching during joint decision making. Drawing on motion-capture and voice data from 12 dyads, we analyzed *body-sway* and *pitch-register* matching during sequential transitions and continuations, with and without mutual visibility. Body sway was matched most strongly during sequential transitions in the conditions of mutual visibility. Pitch-register matching was higher during sequential transitions than continuations only when the participants could *not* see each other. These results suggest that both body sway and pitch register are used to manage sequential transitions, while mutual visibility influences the relative weights of these two resources. The conversational data are in Finnish with English translation.

Social interaction consists of cooperative activities that require a great amount of interpersonal coordination (Shockley, Santana, & Fowler, 2003, p. 326). In empirical behavioral research on social interaction, at least two significantly different approaches have been used to study such coordination. On one hand, experimental psychologists and cognitive neuroscientists seek to understand the exact behavioral and brain mechanisms underlying successful communication and the feelings of connection. On the other, researchers within the domain of conversation analysis aim to describe how participants in naturally occurring interactions achieve coordination at the level of turn-by-turn unfolding of sequences of action. In this article, we aim to bring these two approaches into a dialogue in the hope of contributing to a better understanding of interpersonal coordination. We report a study that was motivated both by prior experimental work on communication and by conversation-analytic findings on sequential structures of social interaction. Focusing on joint decision-making sequences, we sought to find out whether the amount of behavioral similarity in two domains of behavior—*body sway* and *pitch register*—is sensitive to where in the decision-making sequence the participants at each moment are, and whether the results would depend on whether the participants can see each other or not. More specifically, we address the following two research questions:

RQ1: Is there a difference between sequential continuations and sequential transitions regarding the degree of similarity in participants' body sway and use of pitch register?

**CONTACT** Melisa Stevanovic ✉ [melisa.stevanovic@helsinki.fi](mailto:melisa.stevanovic@helsinki.fi) 📧 University of Helsinki, P. O. Box 4 (Vuorikatu 3 A), 00014 Finland.

📄 Supplemental data for this article can be accessed on the [publisher's website](#).

The first and second authors contributed equally to this work. All authors wish to thank the Department of Computer Science, Aalto University, for the use of their Motion Lab and motion capture system.

Color versions of one or more of the figures in this article can be found online at [www.tandfonline.com/HRLS](http://www.tandfonline.com/HRLS).

RQ2: Does mutual visibility of the participants influence the similarity of body sway and pitch register, and if so, does the effect of mutual visibility interact with that of sequentiality?

## Behavioral matching and mismatching

For cooperative activities—especially those containing speech—one apparent index of coordination is the similarity of behavior. People match body movements at various levels (Chartrand & Bargh, 1999; Hari, Himberg, Nummenmaa, Hämäläinen, & Parkkonen, 2013; Ramseyer & Tschacher, 2008; Shockley, Richardson, & Dale, 2009), starting from copying of each other's lexical choices (Garrod & Anderson, 1987), accents (Giles & Powesland, 1975), and syntax (Branigan, Pickering, & Cleland, 2000), and extending to mimicking each other's gestures (Kimbara, 2006) and facial expressions (Lundquist & Dimberg, 1995), synchronizing postures (LaFrance, 1982) and gaze behavior (Richardson & Dale, 2005), converging in speaking rate (Street, 1984), vocal intensity (Natale, 1975), and pausing frequency (Cappella & Planalp, 1981), as well as entraining into the melodic and rhythmic features of each other's speech (De Looze, Oerte, Rauzy, & Campbell, 2011; Himberg, Hirvenkari, Mandel, & Hari, 2015). Although all these phenomena are strongly related, the terminology used varies, mostly depending on the exact temporal coupling between the two (similar) behaviors of interest. The term *synchrony*, for instance, refers to temporal coupling of independent oscillators that enter into a phase relationship (e.g., Miles, Griffiths, Richardson, & Macrae, 2010). In contrast, the term *convergence* denotes the ways in which participants' behaviors become more similar over time (Paxton & Dale, 2013), and the term *imitation* refers to similarities of behavior occurring at time lags short enough for an observer to be able to recognize the original behavior and the one that represents a copy of it (Louwerse, Dale, Bard, & Jeuniaux, 2012). In this article, we use a term *behavioral matching*, as it is neutral with respect to the aforementioned issues of temporality, as well as with regard to the causal mechanisms underlying the coupling—that is, whether participants specifically intend to make their behaviors similar to those of their coparticipants or they end up doing that as a result of other factors that drive their behavior.

People engaged in joint actions also coordinate their dissimilar behaviors—for example, in tasks with mutually understood structure (Sebanz, Bekkering, & Knoblich, 2006). For example, lifting a heavy box or conducting a financial transaction often requires that two people perform complementary actions in a time-aligned way—otherwise the box falls or the financial transaction ends in confusion. These examples point to a more general principle that it may be precisely the alternation between behavioral matching and mismatching that drives the interaction and makes it interesting (Beebe & Lachman, 2002; Fuchs & De Jaegher, 2009)—a suggestion that this article seeks to elaborate.

Extensive literature suggests that behavioral matching has functions that make it more likely to occur in certain social situations and less likely in others (for a review, see Beňuš, 2014). In particular, Howard Giles et al. have developed this idea over the last two decades in their communication–accommodation theory (see e.g., Giles, Coupland, & Coupland, 1991). The core insight behind the theory is that the degree of behavioral matching may be used as a means for achieving a desired social distance between a participant and the coparticipant, more similarity generally leading to a smaller social distance and vice versa. Accordingly, empirical studies have linked the similarity of speech rate to more positive ratings of competence (Street, 1984), attractiveness (Putnam & Street, 1984), and supportiveness (Giles, Mulac, Bradac, & Johnson, 1987). Such effects have also been shown to have actual future consequences in dyadic interaction: Participants whose behaviors were similar during the initial interaction displayed increased trust and prosociality during subsequent collaborative tasks (Hove & Risen, 2009; Manson, Bryant, Gervais, & Kline, 2013; Valdesolo, Ouyang, & DeSteno, 2010). Overall, people are motivated to match their behavior depending on the interaction partner. Such selectivity of behavioral matching to social contextual affordances is in line with the general idea that social interaction consists of alternations between behavioral matching and mismatching. But we still do not know much about those social contextual affordances that vary from moment to moment in interaction, even if the interaction partner remains the same.

The idea of alternations between behavioral matching and mismatching gets somewhat more specific when we know how the degree of similarity between the participants' behaviors is influenced by task

requirements. Paxton and Dale (2013) found significantly less bodily synchrony within a dyad during argumentative than affiliative conversational settings. Fusaroli and Tylén (2012) found that when dyads were making joint decisions in a psychophysical task, the degree to which the participants matched each other's *task-relevant* expressions correlated positively with their task performance, whereas the indiscriminate matching of *all* expressions had the opposite effect on task performance. Hence, people seem to be sensitive to what to match and when.

In this article, we aim to take those social contextual affordances that drive the alternations between behavioral matching and mismatching one crucial step forward. In the aforementioned literature, the participants' behaviors have mostly been examined with reference to the qualities of the entire (and relatively long) interactional episodes (however, Manson et al. (2013) divided 10-minute conversations in three phases for the sake of a more detailed analysis). In this study, in contrast, we investigated behavioral matching with reference to significantly smaller units of social interaction: successive phases of the sequential unfolding of interaction. Informed by the conversation-analytic perspective, we asked whether behavioral matching has interactional functions with relevancies that would vary with respect to the sequential context.

### Coordination between sequences

The question how participants coordinate their behaviors in sequences of action is of fundamental interest to conversation analysis. In contrast to the studies in experimental psychology and cognitive science previously reviewed, conversation-analytic research has focused on a detailed description of the moment-by-moment dynamics of interactional events. Any particular behavior is analyzed with respect to the very action that it implements in its immediate sequential environment. Conversation-analytic studies have shown that the interactional import of two identical behaviors can differ drastically depending on what has been said and done in the interaction before. From this point of view, it is apparent that a deeper understanding of the interactional functions of phenomena such as behavioral matching and mismatching requires that the sequential context is taken into consideration.

Action sequences are of many different types (e.g., a question–answer adjacency pair, an invitation followed by an acceptance) and of different lengths (e.g., an exchange of *Hello* vs. a storytelling). However, common to all of them is that their composite elements are bound together by a relation of conditional relevance, whereby the first action invokes an expectation of a more or less narrow range of possible next actions (Schegloff, 2007). While some sequences have easily identifiable and projectable start and end points (e.g., exchange of *Hello*), in longer or less canonical sequences (see Stivers & Rossano, 2010), the *transitions* between individual sequences may constitute significant challenges and triggers of coordination with the other participant(s). Coordinated body movements may have an important role in the successful accomplishment of such transitions. For example, in Mandarin Chinese conversations, the instances of the so-called recipient intervening question–answer sequences involved the recipient leaning toward the speaker and holding that posture until the response was provided, while a postural change indicated closure of the sequence (Li, 2014). Other conversation-analytic studies have emphasized the importance of prosodic cues (Couper-Kuhlen, 2004; Goldberg, 2004; Szczepek Reed, 2006, 2009) and gaze behavior (Rossano, 2012) in marking the transitions between sequences (for a detailed analysis of multimodal practices in the accomplishment of sequential transitions, see Mondada, 2006).

Coordinated sequential transitions are particularly important in the context of joint decision making. Establishing a joint decision (whatever the decision is about) must start with a proposal, which subsequently gets accepted (Houtkoop, 1987; Stevanovic, 2012a, 2012b). However, what counts as an acceptance is not always straightforward. Even if the adjacency pair consisting of a proposal and an acceptance may represent the core of a decision-making sequence, it frequently gets expanded by other adjacency-pair-like substructures, as the participants fail to come up with proposals that their coparticipants could immediately accept (e.g., on “insert expansions,” see Schegloff, 2007, pp. 97–114). As a result, a typical decision-making sequence may best be characterized as what Schegloff (2007, p. 252) has referred to as the “sequence of sequences.” But how exactly do the participants then signal that, at a

particular moment of interaction, the decision-making sequence is (not) ripe for closure? This is an empirical question that we aim to address in this article.

Extract 1, drawn from a planning meeting between two church officials, gives a grasp of the emergence of new decisions at sequential transitions (for a description of the data set, see Stevanovic, 2013). The participants—a pastor (P) and a cantor (C)—are selecting hymns for the next Sunday’s mass and have previously considered the aptness of a certain hymn as the opening hymn of the mass. At the beginning of the extract, the participants appear to come to a joint conclusion (for the meaning of the glossing abbreviations, see the appendix A).

(1) (HTM\_0:59)  
 01 C: *tää* >on kyl iha hieno,<  
       DEM be PRT quite fine  
       *this* >is quite fine, <  
 02 P: se on ihan hieno.  
       DEM be quite fine  
       *it is quite fine.*  
 03 C: *tää* on hyvä alkuvirsi.  
       DEM be good opening.hymn  
       *this is a good opening hymn.*  
 04 (0.9)  
 05 P: *sitten*, hhhhhh  
       PRT  
       *then*, hhhhhh  
 06 (0.4)  
 07 C: joo-o?  
       PRT  
       o-kay?  
 08 (0.8)  
 09 P: *kolminaisuusvirteen*  
       trinitarian.hymn-ILL  
       *to the trinitarian hymn*

After the pastor and the cantor have both provided their general positive evaluations of the hymn (lines 1–2), the cantor makes a more specific statement, asserting that the hymn would indeed be apt to the specific purpose at hand (line 3). In this sequential environment, where the participants have been trying to find a suitable opening hymn for the mass and have ultimately displayed agreement about the virtues of a certain hymn, the cantor’s statement is treated as an indication that a decision has been established. This becomes apparent by what happens next: The pastor launches a new sequence—one about the next hymn in the mass (line 9).

In the context of joint decision making, the management of sequential transitions is not only about the convenience and communicative efficacy associated with smoothly unfolding interaction. It is also about the *jointness* of the emerging decisions. For sure, to launch a new sequence at a point at which the coparticipant has *not* yet received access to what the decision will be about, expressed agreement with his/her coparticipant’s views, or displayed commitment to the proposed action, is to establish a unilateral decision (Stevanovic, 2012a). In other words, in this context, the jointness of a decision is essentially a matter of interpersonal coordination, which involves the participants constantly displaying a shared understanding of where in a sequence they are. A lack of such coordination is sometimes the same as to impose one’s views and choices on others.

How do the participants involved in joint decision making then coordinate their arrival at a common conclusion that now all the participants have expressed their appreciation of a proposed idea strongly

enough so that a joint decision can be said to have emerged? Knowledge about the precise moment when a new decision emerges is important for two reasons. First, since the initial proposal might have gone through several modifications, the participants need to have a common understanding on what is the last and binding version of the idea—the one to which they commit themselves. Second, participants also need to manage their current interaction: They need to know when to stop pursuing the same proposal anymore and when it is acceptable to initiate a new topic. Hence, in organizational meetings of a large scale, it is usually the task of the chairperson, often by using the hammer, to mark the exact moment of the emergence of a new decision and to manage the flow of interaction. In informal everyday interactions, other resources are needed for that purpose, and in this article, we will examine the role of *behavioral matching* in this regard.

## Foci of the study

In this article, we will focus our investigation on two domains of behavior, where interacting participants may or may not exhibit similar behaviors: the phenomenon of *body sway* and the use of *pitch register*.

While, within the field of conversation analysis, body sway is a novel research topic, it has been studied quite extensively in experimental psychology and cognitive science. It has turned out that standing upright, while apparently simple, is actually a very complex sensorimotor task. The body sways continuously within a range of approximately 4 cm (Richardson, Dale, & Shockley, 2008, p. 80), and behaviors such as looking (Stoffregen, Pagulayan, Bardy, & Hettinger, 2000), reaching (Belen'kii, Gurfinkel, & Pal'tsev, 1967; Feldman, 1966), speaking, and breathing have an influence on the dynamics of the body sway (Conrad & Schönle, 1979; Dault, Yardley, & Frank, 2003; Jeong, 1991; Rimmer, Ford, & Whitelaw, 1995; Yardley, Gardner, Leadbetter, & Lavie, 1999). Even without any specific behavioral tasks, the configuration of the body must be constantly adjusted to keep the balance during changing mass distribution (Richardson et al., 2008, p. 80).

In a seminal study, Shockley et al. (2003) showed that two participants synchronized their body sways, maintaining similar postural trajectories longer, when they worked together on a puzzle task, compared with a condition where the participants were still in the same space but each performed the same task with a confederate. Quite surprisingly, the body-sway patterns were similar irrespective of whether or not the two participants could see each other during the task. On the basis of this finding, the researchers concluded that, at least to some extent, body sway matching could be an epiphenomenon of the convergence of the participants' speaking patterns (see also Shockley, Baker, Richardson, & Fowler, 2007). In this article, we seek to dig deeper in these ideas. We ask whether there are specific interactional circumstances, such as sequential transitions, where the similarity of body sway could serve as a resource of coordination. From this point of view, it will be crucial to learn (a) whether the degree of similarity in participants' body-sway patterns is sensitive to different sequential phases and (b) whether such sensitivity would be enhanced by the participants' mutual visibility. If we obtain a positive answer to both of these questions, we may conclude that the participants' body-sway patterns function as a *resource* for sequential coordination, instead of being (merely) epiphenomena of such coordination.

Unlike body sway, participants' use of pitch register has already been discussed within the field of conversation analysis. Similarity, in the context of pitch register, means that two speakers use pitch levels that are similar in relation to their own voice ranges (Szczepek Reed, 2006, p. 42). In her pioneering study on a riddle-guessing game on a radio phone-in, Couper-Kuhlen (1996) showed that similarity in the use of pitch register (accompanying lexically highly repetitive utterances) is a way for the second speakers to mark their utterances as quotations of prior speakers' utterances (while the matching of absolute pitch levels was associated with the second speakers mimicking the prior speakers in a disaffiliative way). Later studies by Couper-Kuhlen (2004) and Szczepek Reed (2006, 2009) have shed light on the use of pitch register as a resource in the management of sequences: Sequential transitions typically coincide with an extreme upwards shift in pitch register, while sequential continuations co-occur with a continuation of the pitch register of the previous speaker.

In this article, we aim to find out how pitch register<sup>1</sup> configures transitions between joint decision-making sequences and whether participants' mutual visibility plays any role in this regard.

On the basis of prior conversation-analytic research on the management of sequences and the particular challenges of coordination associated with sequential transitions, we expected that the between-participant similarity of body sway and pitch register would be different during sequential transitions, compared with sequential continuations. However, given the lack of previous literature on the topic, we refrained from taking a strong position on where in a sequence we would find the most (or least) behavioral matching, but we thought that a sequential transition would possibly be an environment that particularly calls for the participants to exhibit similar behaviors. Also regarding the question about the influence of participants' mutual visibility on behavioral matching, the literature did not allow us to form clear hypotheses. While the previous research on behavioral matching suggests that at least some similarities of behavior may simply be epiphenomena of other cues such as rhythm of conversation, the conversation-analytic view on interaction suggests that even the most subtle, and apparently automatic, behaviors may serve as resources of coordination of joint action. We expected that, by considering the influence of mutual visibility on behavioral matching and on the ways in which it may be sensitive to sequential context, we could find evidence that would support one view or the other.

## Method

Two Finnish-speaking participants at a time engaged in joint decision-making tasks, while we recorded their body movements with an optical motion-capture system and their voices with portable head-worn microphones. The dyads needed to discuss, negotiate, and decide on descriptions of fictional characters, while either facing or not facing each other. These immersive and engaging tasks were developed to afford naturalistic dynamics of joint decision-making interaction, while the task structure provided repeatability and comparability of the data across all dyads.

### Participants

Altogether 24 healthy participants (seven female–female and five male–male dyads; mean  $\pm$  *SD* age  $27.0 \pm 6.6$  years) were recruited via e-mail lists. Four dyads knew each other very well, four not at all, and four were somewhere in between. Their identity was revealed only to the members of the research group, who needed to sign a confidentiality contract. The study had prior approval by the Aalto University Research Ethics Committee. All participants were informed about the use of the data and signed a consent form.

### Apparatus

#### Optical motion capture

We collected body-movement data with a 20-camera OptiTrack motion-capture system, which measures at 10-millisecond intervals the locations of 37 optical reflectors attached to each participant's suit. From the resulting three-dimensional representation of the participants' body movements, we then examined the positions, movement velocities, and accelerations of different body parts.

#### Audio recordings

Each participant's voice was recorded using a DPA d:fine™ portable head-worn condenser microphone that has a frequency response from 20 Hz to 20 kHz.

---

<sup>1</sup>In the studies by Couper-Kuhlen (2004) and Szczepek Reed (2006, 2009) the similarities and differences in speakers' usage of pitch register frequently co-occurred with analogous similarities and differences in the realm of intonation. In this study, we focus solely on speakers' usage of pitch register, since the role of intonation in Finnish is generally less central than in Indo-European languages, such as French, English, and German (Iivonen, 1998, p. 319).

### Additional Recordings

HD video recordings of the trials were used as a reference and to illustrate the patterns identified in the quantitative statistical data analysis. We also measured the participants' gaze direction using portable eye-tracking glasses, but these data will not be reported here.

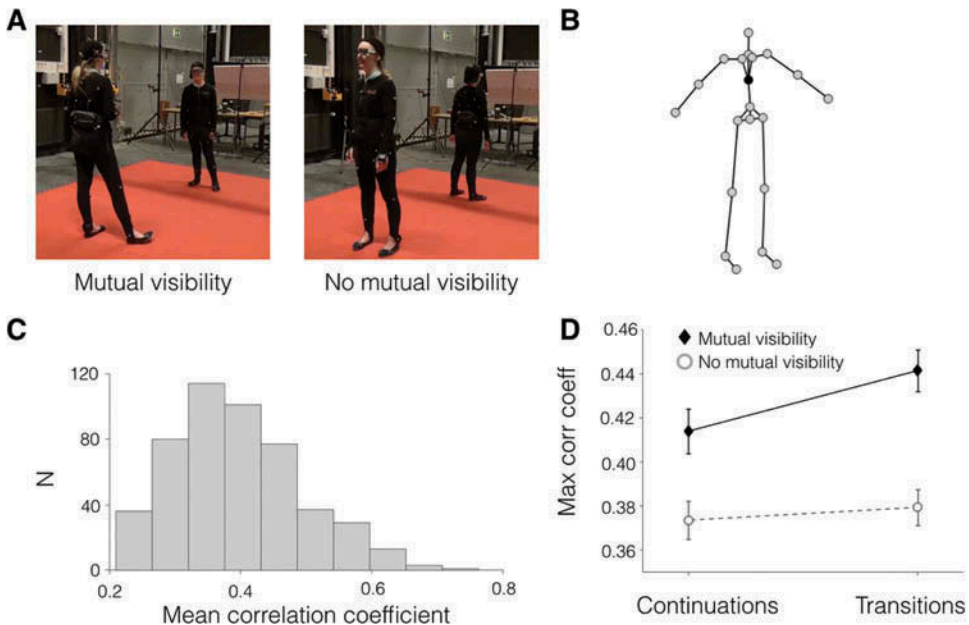
### Procedure

A single dyad was studied at a time. At the beginning of each session, the two participants put on the motion-capture suits, head-worn microphones, and eye-tracking glasses, which were then calibrated. The participants carried out three warm-up tasks, aimed to help the participants to forget the measuring equipment, relax, and get acquainted with interacting with each other.

The participants were asked to choose together an adjective that best describes a fictional target. The adjective needed to start with a given letter, and once a decision was reached, the dyad had to move to the next letter in the alphabet, deciding altogether on eight adjectives. As a motivation for the task, the participants were told to imagine being editors of a children's book, teaching the alphabet to kids by featuring the target character, and they'd need to choose suitable adjectives for that purpose.

The task was performed twice. In one trial (consisting of eight decisions), the adjective target was Donald Duck, and in the other, Scrooge McDuck, while the letters were either  $[H, I, J, K, L, M, N, O]$  or  $[N, O, P, R, S, T, U, V]$ . In one condition, the participants could see each other, in the other they could not (Figure 1A).

The order of the two visibility conditions (mutual visibility, no mutual visibility), the type of the target (Donald Duck, Scrooge McDuck) and the alphabet list, as well as the order of this task in relation to another task not reported here, were counterbalanced across pairs. At the end of each session, the participants filled in a questionnaire about their experiences with the task requirements and their collaboration partners.



**Figure 1.** A: The two conditions of the experiment. B: Body-joint marker locations (black marker is the chest point that was used to calculate the body sway). C: Histogram of mean correlation coefficients of participants' body sways. D: Body sway synchrony: Interaction of visibility and sequence phase. The error bars indicate the standard error of the mean.



## Measured data and data processing

Out of the 24 trials carried out (12 dyads  $\times$  2 visibility conditions), we obtained audio, video, and motion data from 22 trials. Two trials were discarded due to data corruption (camera running out of memory or the motion data being accidentally overwritten). We then had 10 successful trials for mutual visibility and 12 for no mutual visibility. As the tasks were self-paced, the durations of the trials varied a lot (from 2 minutes to more than 7 minutes, most trials lasting about 3–4 minutes). The audio analyses were primarily carried out using recordings from the head-worn microphones. However, in four trials, the data from the microphone were so noisy that the audios recorded by the eye-tracking glasses were used instead.

## Annotations

All trials were annotated for decision-making sequences and their phases, whereby the boundaries between *sequential transitions* and *sequential continuations* were determined. In the initial annotations, each decision-making sequence was broken down to their beginning, middle, and end phases. In the beginning phase, the participants established their current task (i.e., the letter with which the next adjective should begin with). In the middle phase, the participants made proposals and discussed their merits; this phase started at the point at which one of the participants made her/his first suggestion. Finally, in the end phase, the participants displayed commitment to their joint decisions (e.g., *otetaan se* “let’s take it”). Extract 2 illustrates our ways of annotating the three sequential phases.

```
(2) (P6_selin_I)
01 A: I, _____
      LETT BEGINNING
      I,
02 B: I,
      LETT
      I,
03 A: eiks se oo aika innokas, _____
      NEG-Q DEM be quite eager MIDDLE
      isn't he quite eager,
04 (0.5)
05 A: vaikka toisaalta se on kyllä aika laiska.
      PRT on.the.other.hand DEM be PRT quite lazy
      but on the other hand he is certainly quite lazy.
06 (.)
07 B: se o kyl laiska enemmän ehkä ku innokas.=
      DEM be PRT lazy more perhaps PRT eager
      certainly he is lazy more than eager.=
08 =se ei oo kyl,=
      DEM NEG be PRT
      =he is certainly not,=
09 =ja se ei oo ilkeä.
      AND DEM NEG be mean
      =and he is not mean.
```

- 10 A: ei ookkaan.  
NEG be-CLI  
*no he isn't.*
- 11 (9.0)
- 12 A: mm,  
PRT  
*mm,*
- 13 (1.6)
- 14 A: joo.  
PRT  
*yea.*
- 15 (0.4)
- 16 A: on se aika ilonen kyllä mutta,  
be DEM quite happy PRT but  
*he is certainly quite happy but,*
- 17 (3.2)
- 18 A: välillä sitten kuitenkin ei.  
sometimes PRT however NEG  
*sometimes he isn't however.*
- 19 (3.4)
- 20 A: itseppäine,  
stubborn  
*stubborn,*
- 21 (0.4)
- 22 B: ↑itseppäinen on kyl todella hyvä ninku, (0.5) END  
stubborn be PRT really good PRT  
*↑stubborn is really a very good (word), (0.5)*
- 23 ööö kuvaamaan Aku Ankkaa,  
describe-INF-ILL nameFictionalCharacter  
*err to describe Donald Duck,*
- 24 (1.6)
- 25 B: otetaan se.  
take-PASS DEM  
*let's take it.*
- 26 A: otetaa vaan.  
take-PASS PRT  
*let's take (it) then.*
- 27 (1.6)
- 28 B: sitte J.  
PRT LETT  
*then J.*

---

 BEGINNING

In the beginning phase, the participants establish their joint task by mentioning the letter to be talked about next (lines 1–2). The middle phase starts when one of the participants (A) makes a task-relevant proposal *innokas* (“eager,” line 3). After having discussed several less-than-optimal decision alternatives, A comes up with a new adjective: *itseppäinen* (“stubborn,” line 20). Then, in what we annotated as the end

phase of the sequence, B expresses a strong agreement with this proposal (lines 22–23) and suggests that the participants would select it (line 25), which is followed by A displaying his commitment to the decision (line 26). After a silence (line 27), B initiates a new sequence (line 28).

On the basis of these initial annotations, we formed two clearly distinct sequential contexts for the analysis: The time windows containing the end and beginning phases formed contexts that we called *sequential transitions*, while the middle phases formed *sequential continuations* (for the technical details of the annotation procedure, see the Supplementary Materials).

### **Movement data**

The movement data from the optical motion capture system were processed to enable the analysis of *body sway*. From the raw position data of the 37 markers in the motion-capture suit, 21 body-joint positions were calculated. One of these, the chest point, was then chosen for the analysis of body sway (black circle in [Figure 1B](#)). Thereafter, the second time derivative of position (acceleration) was calculated. Finally, from the three spatial dimensions, acceleration along the trajectory of the movement was calculated by taking the Euclidian norm. This gave us a time series of the chest's acceleration without regard to its direction, making the data comparable across conditions. This norm acceleration of the chest point then represented the participant's body sway. To calculate the within-dyad synchrony in body-sway, these chest-acceleration time series for the two participants were cross-correlated with each other. We calculated the cross-correlations for time lags up to 3 seconds, so as to be able to detect matches between the time series even if they would occur at a delay. To consider the changes in body-sway synchrony over time, the cross-correlation analysis was carried out in a 6-second moving window (600 samples) through each entire trial. From the resulting time series, we picked the maximum correlation coefficients for each time point and averaged across the length of each sequential phase to obtain the average maximum correlations used in the following statistical analysis ([Figures 1C](#) and [1D](#)). The two-dimensional time series of correlations and their lags were also visualized as cross-correlograms ([Figures 2B](#) and [3B](#); for more details on these measures, see the Supplementary Materials).

### **Audio data**

For the analysis of *pitch register*, the fundamental frequency data were extracted from the audio recordings at 10-millisecond intervals in Praat. The participants' use of *pitch register* was analyzed by first calculating the "global" pitch mode in semitones (re 1 Hz; noted as  $st_1$  Hz) for each participant. For each turn ( $N = 2,040$ , stretch of speech separated by at least 0.3 seconds of silence), we extracted the mean fundamental frequency in semitones and then, to obtain the relative pitch level for each turn, subtracted the speaker's global pitch mode from the absolute mean pitch level of each turn. It was these relative pitch levels that we then considered with reference to the notion of pitch register. Thus, to measure how similarly the speakers used their pitch registers, we subtracted the relative pitch levels in each turn pair (adjacent turns by speakers A and B). In the statistical analysis, we then compared the degree of similarity in pitch register use during sequential transitions against the degree of similarity during sequential continuations (for a more detailed description of the pitch analysis, see the Supplementary Materials).

## **Results**

### **Matching of body sway**

The average maximum correlation coefficients of the chest-point accelerations between the two participants of a dyad were 0.39 ([Figure 1C](#)). These relatively low coefficients were to be expected, as participants were not engaged in rhythmic activity such as dancing (cf. Himberg & Thompson, 2011). Then we conducted a two-way ANOVA (2 x 2) on these correlation data, with sequential context (sequential transition, sequential continuation) and visibility (mutual visibility, no mutual visibility) as the two independent factors (see [Table 1](#)). There was a statistically nonsignificant trend

**Table 1.** Results of the Two-Way ANOVA's for Mean Body Sway Correlation, Individual Speakers' Pitch Register Use, and Their Turn-By-Turn Pitch Register Matching, With Factors *Visibility* (Mutual Visibility, No Mutual Visibility), and *Sequential Phase* (Sequential Transition, Sequential Continuation).

<i>Dependent Variable</i>	<i>Effect</i>	F	df, df error	p	<i>part. η<sup>2</sup></i>	<i>sig.</i>
Mean body sway correlation	Visibility	31.6	1, 351	< .001	0.083	***
	Sequential phase	3.35	1, 351	.068	0.009	n.s.
	Visibility*Seq. phase	1.41	1, 351	.235	0.004	n.s.
Individual pitch register use	Visibility	7.164	1, 2036	.007	0.004	**
	Sequential phase	4.150	1, 1296	.042	0.002	*
	Visibility*Seq. phase	0.035	2, 1296	.851	0.01	n.s.
Turn-by-turn pitch register matching	Visibility	0.00	1, 1282	.996	0.000	n.s.
	Sequential phase	0.131	1, 1282	.717	0.000	n.s.
	Visibility*Seq. phase	6.191	1, 1282	.013	0.013	*

Note. Partial eta squared (*part. η<sup>2</sup>*) was used as an estimate of effect size.

\*\*\* $p < .001$ ; \*\* $p < .01$ ; \* $p < .05$ ; n.s. = not significant.

for the participants to synchronize their body sways more during sequential transitions than sequential continuations ( $p = .068$ , solid line in Figure 1D). The participants' mutual visibility demonstrated a statistically significant effect on the synchronization patterns of their body sways: Body sways were more similar when the participants could see each other. In addition, there was a statistically nonsignificant trend in an interaction effect, indicating that the difference between the two sequential contexts was smaller when the participants could not see each other (see Table 1).

To get a grasp of the concrete interactional behaviors that underlie the trend of the participants to synchronize their body sways more during sequential transitions than during sequential continuations, we inspected in our data instances of high body sway correlations in more detail. Extracts 3 and 4 represent cases where the sensitivity of body sway to sequentiality is particularly clear. But of course, as the statistical nonsignificance of our aggregate result suggests, not all instances of high body-sway synchrony perfectly overlap with sequential transitions. Still, we believe that the scrutiny of these single cases illustrates the kinds of behaviors that underlie our trending result.

Extract 3 is drawn from a condition where the participants try to agree on adjectives describing Scrooge McDuck and have been previously trying to find an adjective starting with the letter *N*. The extract starts by B suggesting that the adjective for the letter *N* could be *nilkki* ("jerk," line 1). This is followed by A repeating the adjective and producing a "compliance token" (Stevanovic & Peräkylä, 2012) *okei* ("okay," line 3) in overlap with A's repetition of the adjective (line 4). The sequence is brought to a close with A's subsequent particle *joo* ("yeah," line 5), which is delivered with the kind of final rise that is typical for closings of routine-like subactivities within larger projects (see Sorjonen, 2001, p. 150). The participants' mutual understanding that a new decision has emerged is manifested in the participants' subsequent conduct: After a silence (line 6), B produces the particle *sitten* ("then," line 7), thus marking a shift to a new sequence, which is accompanied by A summoning the next letter in the alphabet (line 8).

```
(3) (P3_kasvokkain_N)
01 B: nilkki.
      jerk
      jerk.

02      (0.3)

03 A: nilkki. *(.) [okei. ]
      jerk          PRT
      jerk.        (.) [okay. ]
                  * FRAME 1
```

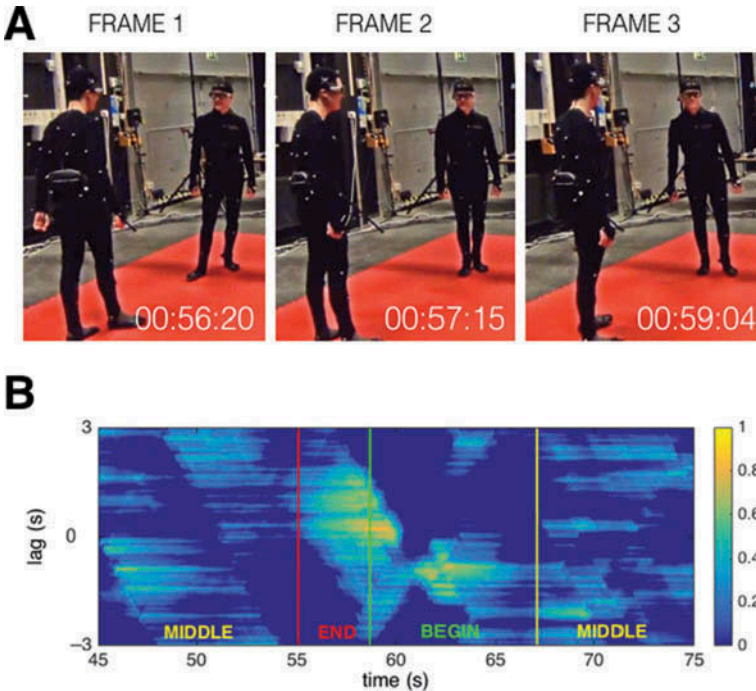
04 B: [nilkki.]  
 jerk  
 [jerk. ]

05 A: joo?  
 PRT  
 yeah?

06 \* (0.7)  
 \* FRAME 2

07 B: sitten,  
 PRT  
 then,

08 A: O,\*  
 LETT  
 O,  
 \* FRAME 3



**Figure 2.** A: Frames for Extract 3. Participant A on the left. B: Cross-correlogram for a 30-second timespan (from 45 to 75 seconds) containing Extract 3. Vertical lines show the boundaries between sequential phases, and in the graph, bright colors show time points and lags with high correlation between participants' body sways.

During the sequential transition, the participants engaged in a series of postural changes, which exhibited remarkable similarity. Frame 1 in [Figure 2A](#) shows the participants' body postures when the joint decision is about to be reached (line 3). Immediately thereafter, during A's particle *joo* ("yeah," line 5), both participants change their postures. As shown in Frame 2, at the beginning of an ensuing silence (line 6), both participants have taken a step to the side to bring their legs together. But then, during the subsequent utterances that, in this context, convey the participants' readiness to move to the next sequence (lines 7–8), the participants once more change their postures, this time by moving their legs further away from each other. The result of these postural changes can be seen in Frame 3. Notably, all these synchronous movements happened within a timeframe of less than 3 seconds.

These simultaneous postural changes also give rise to body-sway synchrony, and they accentuate the differences of body-sway synchrony in different sequential contexts (see Figure 2B). In this cross-correlogram, color represents the strength of the correlation, time runs along the *x*-axis, while *y*-axis represents lags. Bright yellow indicates strong correlation, and the nearer such color is to the middle of the vertical axis (lag 0), the more simultaneous the participants' movements are. In Figure 2B, there is no clear synchrony during either of the middle phases (sequential continuations), whereas the bright color during the end and beginning phases (sequential transition) indicates high body-sway synchrony with only a very small lag.

Extract 3 exemplifies the types of synchronous postural changes that may co-occur at sequential transitions and that are reflected in the similarity of the acceleration patterns of the two participants' chest joints. Notably, however, not all instances of synchronized body sway that we found from our data involved such easily noticeable postural changes. This is demonstrated in Extract 4, where only one participant (A) engaged in visible postural changes, while the other participant (B) was just standing still.

In this extract, the participants are trying to find an *M*-initial adjective to describe Scrooge McDuck. Previously, A has stated that, in her view, Scrooge likes to contemplate about the good old days. In response to that, B suggests that the participants would select the adjective *mietiskelevä* ("contemplative," line 1) for the letter *M*. This is followed by A repeating the adjective (line 2)—something that is treated by B as a request for confirmation (line 3). After a silence (line 4), B makes a humorous reference to the task instruction (line 5), which involved a prompt for the participants to imagine that they are editing a children's book. B's utterance is produced with a couple of laughter tokens, and A responds to this by producing a laughter token of her own (line 6). Thereafter, A displays commitment to the decision (line 7), B agrees with the affirmative response token *joo* ("yeah," line 8), and the participants' orientation to the decision as having been established becomes clear in that, immediately after this, they start a new sequence (lines 9–10).

```
(4) (P7_kasvokkain_M)
01 B: otetaanko      mietiskelevä
      take-PASS-Q  contemplate-ADV
      shall we take contemplative

02 A: mietiskelevä
      contemplate-ADV
      contemplative

03 B: .joo
      PRT
      .yeah

04   *(0.8)
      *FRAME 1

05 B: tämmönen      opettav(h)ainen [ tari]na laps(h)ille *
      this.kind.of  educative      story   child-PL-ALL
      this kind of an educative [stor]y to the children

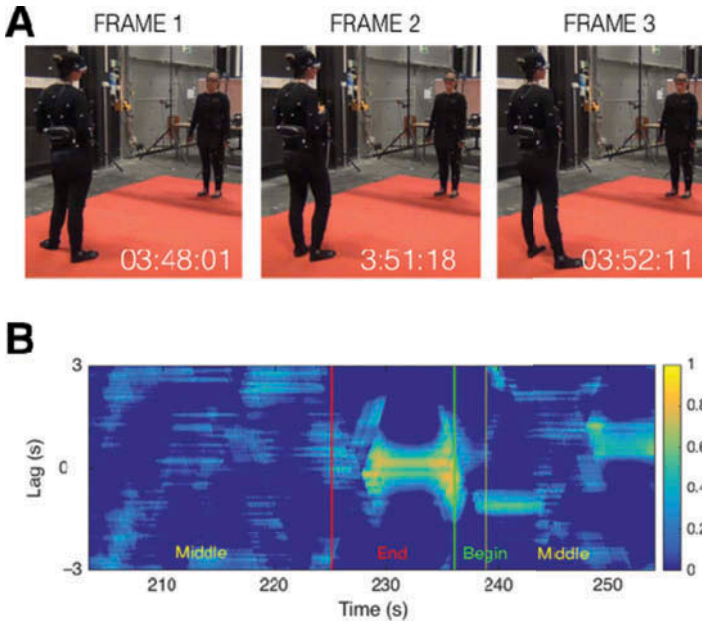
06 A:                                     [heh]
                                           *FRAME 2

07 A: .hhh * joo (.) otetaan      se.=
      PRT   take-PASS DEM=
      .hhh  yeah (.) let's take it.=
      * FRAME 3

08 B: =joo (.) °joo,°
      PRT   PRT
      =yeah (.) °yeah,°
```

09 A: sitten N  
 PRT LETT  
 then N

10 B: N  
 LETT  
 N



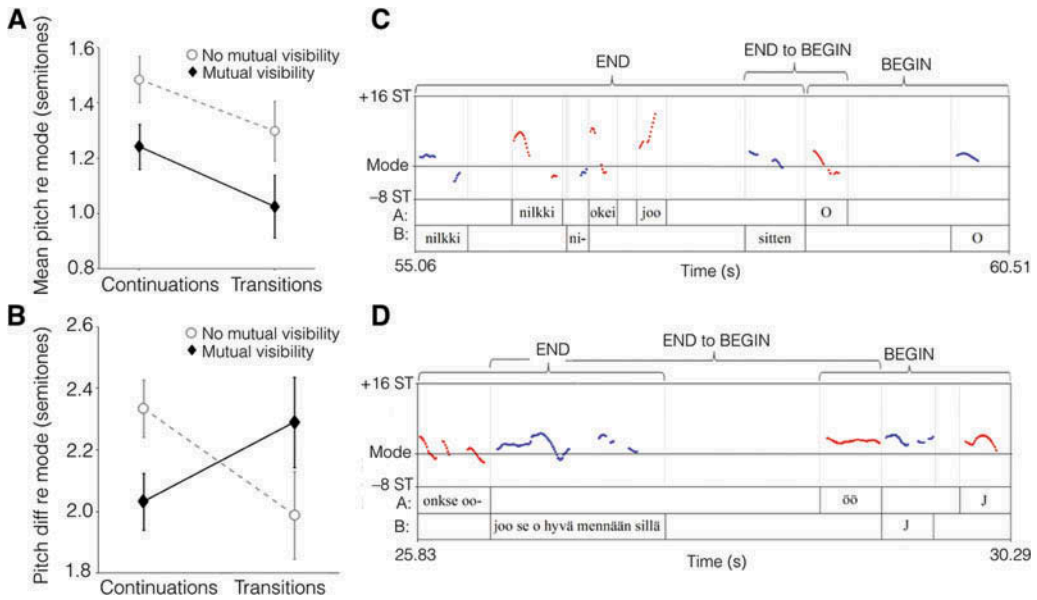
**Figure 3.** A: Frames for Extract 4. Participant A on the left. B: Cross-correlogram for the 60-second timespan containing Extract 4.

At the beginning of Extract 4, both participants appear to stand still (line 4, see Frame 1 in [Figure 3A](#)). Later, after B’s first “serious” move toward a decision, A engages in postural changes. First, during B’s reference to the children’s book (line 5), she brings her legs together (see Frame 2). Then, during her own audible in-breath preparing the subsequent announcement of a decision (line 7), she takes with her right leg a step to the right (see Frame 3). Her coparticipant, however, appears to stand still; only a careful observation of the moving video will show the slight right-left swaying movement of B’s upper body. Still, the cross-correlogram of the two participants’ body sways during this extract (see [Figure 3B](#)) demonstrated a particularly high synchronicity ( $r > 0.7$ ), which suggests that not all interpersonal coordination in terms of body sway can be reduced to synchronous postural changes.

As can be seen in the body-sway cross-correlograms for Extracts 3 and 4 (see [Figures 2B](#) and [3B](#)), in both cases, the segment containing the highest zero-lag correlation (yellow color in the middle of the vertical axis) starts at the end phase of a prior sequence extending to the beginning phase of the new sequence. Thus, at least at some level, what we refer to as “sequential transitions” may be oriented to by the participants as somewhat coherent units of joint action—despite the variety of more specific actions accomplished in and through the individual participants’ turns during the transitions.

### Matching of pitch register

To obtain an overall grasp of the participants’ use of pitch register in the two visibility conditions (mutual visibility, no mutual visibility) and the two different sequential contexts (sequential transitions vs. sequential continuations), we first carried out a two-way ANOVA (2 x 2) on these relative pitch-level data (see [Table 1](#)). Visibility had a statistically significant effect on the



**Figure 4.** A: The participants' use of pitch register in the two visibility conditions (mutual visibility, no mutual visibility) and the two different sequential contexts (sequential transitions, sequential continuations) of our study. B: The degree of difference/similarity in the participants' use of pitch register in the aforementioned conditions. In A and B, the error bars represent standard error of the mean. C: Pitch contours of the two speakers' utterances in Extract 3. D: Pitch contours of the two speakers' utterances in Extract 4.

participants' use of pitch register: On average, the participants used higher relative pitch levels when they could *not* see each other, compared with when they could. Also the effect of sequential context reached statistical significance: The speakers' relative pitch levels were generally lower for sequential transitions than for sequential continuations. There was no interaction between visibility and sequential context (see [Figure 4A](#)).

To study the degree of similarity in the participants' use of pitch register, we organized all the turns produced by two different speakers in a series of turn pairs, and within each of such pair, we calculated the distance between the relative pitch levels of the two turns on the semitone scale, thus obtaining a value for the degree of similarity in the participants' use of pitch register. The smaller the semitone value, the more similar the participants' use of pitch register and vice versa. Then, to assess whether the similarity in the participants' use of pitch register across speaker changes was sensitive to sequential context and to whether the participants could see each other or not, we carried out a two-way ANOVA (2 x 2) on the pitch-register-difference data, with independent factors as before (see [Table 1](#)). In contrast to the previous results, here, neither visibility nor sequential context alone had an influence on the similarity in the participants' use of pitch register across speaker changes. However, there was a statistically significant interaction effect of visibility and sequential context. When the participants could see each other, their use of pitch register was more similar during sequential continuations and less so during sequential transitions. But when the participants could *not* see each other, the pattern was reversed: Their use of pitch register was most similar during sequential transitions (see [Figure 4](#)).

To illustrate our results on pitch register, let us turn to examples from the interaction data. We were particularly interested in the patterns of pitch-register matching during sequential transitions—that is, whether the participants moving from one sequence to the next would match each other's pitch levels less when seeing and more so when *not* seeing each other.

First let us return to Extract 3, where the participants stood face-to-face, discussing the characteristics of Scrooge McDuck. [Figure 4C](#) shows the pitch contours of the two participants' speech during the extract, plotted with reference to each speaker's global pitch mode (see



Lennes, Stevanovic, Aalto, & Palo, 2015) and using the semitone scale. The contours thus represent the speakers' relative pitch levels, not absolute ones.

This sequential transition starts with the both participants uttering the adjective *nilkki* ("jerk"), but they do it by using quite different relative pitch levels. While B's turn (*nilkki*) is produced around his mode, A's turn (*nilkki*) starts from a significantly higher relative pitch level than that of B's. Thereafter, while B continues to use the same register as he did before (*ni-*), A produces two other turns (*okei, joo*) in an even higher pitch register than A's first turn. Arguably, as a response to a proposal, the mismatching of pitch register (along with the affirmative content of the given utterances) conveys the type of high agency that allows the recipient to substantiate his commitment to what is about to be decided (Stevanovic & Kahri, 2011; Stevanovic, 2012b; cf. the work on agreements by Heritage & Raymond, 2005; Ogden, 2006). Subsequently, A indeed treats the decision as established by closing the previous sequence and by initiating a shift to a new sequence (*sitten* "then"). Compared with the previous turn, this turn again exhibits a great downward shift in the pitch register used. During the following two speaker changes, the participants nevertheless match each other's pitch levels relatively well. Overall, the sequential transition involved several speaker changes where the participants' pitch registers essentially differed from each other.

Let us then consider Extract 5, where the same dyad as in Extract 3 is engaged in joint decision making about Donald Duck, this time not facing each other (Figure 4D). Previously, the participants have been discussing the letter *I*, and—similarly as one of the participants in Extract 2 discussed earlier—A has suggested that the adjective *innokas* ("eager") would describe the target. His coparticipant has then stated that the adjective indeed describes Donald Duck. Thus, so as to call for the participants' joint commitment to the choice, A asks whether B thinks that this option would be "okay" (*onkse ookoo* "is it okay," line 1).

```
(5) (P3_selin_I)
01 A: onkse oo[koo.]
      be-Q.DEM PRT
      is it o[kay. ]

02 B: joo se o hyvä mennään [sillä. ]
      PRT DEM be good go-PASS DEM-ADE
      yes that's good let's go [with it.]
03 A: [joo. ]
      PRT
      [yeah. ]

04 (1.2)
05 A: öö
      PRT
      err
06 B: J
      LETT
      J
07 A: J
      LETT
      J
```

As a response to A's question (line 1), B expresses both an agreement with the view that the adjective is "good" for the purpose at hand (*joo se o hyvä* "yes that's good," line 2) and a display of commitment to the action of jointly selecting that adjective (*mennään sillä* "let's go with it," line 2; see Stevanovic, 2012a). As for the verbal content of the turn, it makes the recipient's approval of the first speaker's proposal more explicit than was the case in Extract 3, where the recipient repeated the proposal and produced the tokens *okay* ("okay") and *joo* ("yeah"). But then again, once more in contrast to Extract 3, the recipient does not make use of a high-pitch register but instead a register similar to that used in A's prior turn. The same similarity in the participants' use of pitch register is carried on across all the subsequent speaker changes within the extract. In sum, unlike in Extract 3, where the participants could see each other, in Extract 5, the participants matched each other's pitch registers relatively closely throughout the sequential transition.

Extracts 3 and 5 serve as illustrations of the concrete prosodic patterns reflected in the quantitative results reported in this section. As shown in these examples, the first turn of a new sequence is typically spoken using a pitch register that is near to the speaker's global mode. In these types of joint decision-making sequences, where the participants' interaction follows a mutually known and predetermined structure, downward register shifts at the beginning of new sequences seem to be frequent, while upward register shifts at the same sequential location are rare (cf. Couper-Kuhlen, 2004; Szczeprek Reed, 2006, 2009). Indeed, as shown in Extract 5, a transition from one sequence to a next can also be accomplished without any significant shift in pitch register—and, intriguingly, this seems to be what participants are more inclined to do when *not* seeing each other, compared with when they do.

What then can account for those moments during sequential transitions when speaker changes *do* involve great differences in the two speakers' use of pitch register? As exemplified in Extract 3, the differences in pitch register are frequently associated with the end of the prior sequence and with one speaker producing short exclamatory utterances using a high pitch register, while the other speaker stays at his/her most comfortable pitch level. This view is in line with the previously suggested idea that establishing a new decision often happens by the recipient of a proposal engaging in affective prosodic displays of approval that accompany utterances whose lexical content may be more or less vague (see Stevanovic, 2012b). Our quantitative results suggest this strategy to be more likely in the condition where the two participants can see each other, while other ways of establishing new decisions—those involving more pitch-register matching—may be more adequate in the conditions of no mutual visibility.

## Discussion

As our first research question (RQ1), we asked whether sequential continuations and sequential transitions would differ with respect to the degree of similarity in participants' body sway and the use of pitch register. We found that the instances of highest body-sway synchrony occurred in sequential transitions. This finding is in line with those conversation-analytic studies that associate postural change with sequence closure (e.g., Li, 2014). However, instead of highlighting the importance of such behaviors per se, our study highlights between-participant *similarity* in these behaviors as the critical resource in the management of sequential transitions.

We found that participants use higher mean pitch levels during sequential continuations and lower levels during sequential transitions. Such a result suggests regular alternations in the overall levels of the participants' affective engagement in the task at hand, high mean pitches being associated with a high level of such engagement and low mean pitches with a low level of such engagement (Gupta, Bone, Lee, & Narayanan, 2016; Scherer, 2003; Waaramaa, Laukkanen, Airas, & Alku, 2010). Notably, however, sequential context alone had *no* statistically significant effect on the *similarity* in the participants' use of pitch register across speaker changes.

In our second research question (RQ2), we asked whether the participants' mutual visibility would influence the similarity of body sway and pitch register. We found the participants' body sways to be much more synchronized when the participants saw each other, compared with when they could not. Here, our results differ from those of some previous studies (e.g., Shockley et al., 2007, 2003), where the body-sway patterns of two interacting participants were similar independently of whether the participants could or could not see each other. We were also interested in whether the participants' mutual visibility would be more apt to increase the similarity of body sway during sequential transitions, compared with sequential continuations. This turned out to be the case, even if the interaction effect was only a trend. Still, this finding suggests that, instead of being just an automatic reaction to the visible cues provided by the coparticipant, the similarity in the participants' body-sway patterns may well function as an interactional resource at those moments of interaction when a close coordination is particularly challenging, yet ever more needed—that is, during sequential transitions.

Similarly to sequential context, mutual visibility also had an effect on the participants' use of pitch register: The participants used higher relative pitch when they could *not* see each other, compared with when they could. This difference may be simply due to the matter of securing hearing in the condition where

the participants cannot see each other; an increase in subglottal pressure during speech is likely to increase the fundamental frequency (see, e.g., Ladefoged & McKinney, 1963), while already a high pitch itself facilitates hearing (see, e.g., Neuhoff, Wayand, & Kramer, 2002). However, just like the sequential context, the mutual visibility or its absence also had no effect on the similarity of the participants' use of pitch register across speaker changes. Pitch-register matching nevertheless exhibited a statistically significant interaction effect of visibility and sequential context. While in the conditions of mutual visibility the participants' use of pitch register was more similar during sequential continuations than during sequential transitions, in the conditions of no mutual visibility, the use of pitch register was more similar during sequential transitions than during sequential continuations. What factors could account for these findings?

In our qualitative analysis of data examples on sequential transitions, we discussed why participants who do *not* see each other match each other's pitch registers more closely than those who can see each other. In our analysis, we paid particular attention to the two essentially different ways of establishing joint decisions identified in previous research: (a) recipients' explicit verbal displays of agreement and commitment in response to their coparticipants' proposals and (b) recipients' affective prosodic displays of approval accompanying short exclamatory utterances whose lexical content may be more or less vague (Stevanovic, 2012a, 2012b). According to our qualitative analysis, the first types of instances are more closely associated with pitch-register matching than the second types of instances, where the pitch register used in the recipient's approving response may deviate greatly from the pitch register used in the first speaker's proposal.

From this point of view, our quantitative results now suggest that there is something in the condition of the two participants being able to see each other that may favor the second strategy over the first and vice versa. Further research is needed to learn whether this difference is related to the facilitation of mutual visibility of spontaneous emotional response strategies over the more deliberate verbal ones. Another possibility is that it is the specific challenges associated with sequential coordination when not being able to see the coparticipant that call for the participants to match each other's pitch registers more closely (as it were, to compensate for the lack of synchronized body sway). In any case, our results suggest that participants use both body sway and pitch register in the management of sequential transitions and that the conditions of mutual visibility influence the relative weight given to these two resources.

We acknowledge some shortcomings in our study. First, we analyzed the matching of body movements only with reference to body sway. Obviously also movement in other body parts may be sensitive to sequentiality and mutual visibility—possibly in ways different from body sway. Second, we addressed the prosodic features of speech only with reference to the participants' use of pitch register. Again, a study on other prosodic parameters could reveal patterns of prosodic matching that would be a result of sequentiality and mutual visibility interacting in ways not described here. Besides, as hinted at in our qualitative analysis of single data extracts, visible bodily cues and prosodic patterns might be intricately intertwined with the lexical content of the participants' spoken utterances. Another limitation of our study is its focus on a specific type of sequence—decision-making sequence—in a dyadic setting. Future research should thus test the extent to which our results apply to other types of sequences—in both dyadic and multiparty interactions.

Unlike the traditional conversation-analytic studies, which scrutinize participants' behaviors at the level of turn-by-turn unfolding of sequences of action, in our quantitative analysis, we have summarized sequential phenomena in general metrics of sequential phases that represent behavior in wider units of interaction. From the point of view of conversation analysis, this level of granularity of the analysis may thus come across as relatively coarse. At the same time, however, compared with the previous studies on behavioral matching where participants' behaviors in an entire interaction episode have been summarized as a single data point (see, e.g., Giles et al., 1987; Hove & Risen, 2009; Manson et al., 2013; Putnam & Street, 1984; Street, 1984; Valdesolo et al., 2010), our approach is actually relatively detailed. Most importantly, if it is the alternation of matches and mismatches that drives social interaction (see Beebe & Lachman, 2002; Fuchs & De Jaegher, 2009), our study is among the very few attempts to address these alternations directly.

Human interpersonal coordination happens at multiple time scales ranging from fast automatic reactions to interpersonal cues, across behavioral and gestural coordination of action within sequences of action during single encounters, to the long-term interactional patterns associated with interaction histories and personal relationships (De Jaegher, Peräkylä, & Stevanovic, 2016). From this point of view, we are still far away from being able to outline anything but small parts of the big picture of the phenomenon. We therefore need interdisciplinary collaboration between different kinds of “local” and “global” approaches to better understand the mechanisms, resources, and consequences of human interpersonal coordination. We have here tried to build a bridge between two such approaches, traditionally associated with two separate research fields. Still, to unravel the mysteries of human connectedness, many more such bridges need to be built.

## Funding

This work was supported by the Academy of Finland (#274735 and #131483), the European Research Council (Advanced Grant #232946), and the Louis-Jeantet Prize for Medicine.

## References

- Beebe, B., & Lachman, F. (2002). *Infant research and adult treatment: Co-constructing interactions*. Hillsdale, NJ: Analytic Press.
- Belen'kii, V. Y., Gurfinkel, V. S., & Pal'tsev, Y. I. (1967). Elements of control of voluntary movement. *Biophysics*, *12*, 154–161.
- Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation*, *6*(4), 802–813. doi:10.1007/s12559-014-9261-4
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, *75*(2), B13–B25. doi:10.1016/S0010-0277(99)00081-5
- Cappella, J., & Planalp, S. (1981). Talk and silence sequences in informal conversations III: Interspeaker influence. *Human Communications Research*, *7*(2), 117–132. doi:10.1111/j.1468-2958.1981.tb00564.x
- Chartrand, T., & Bargh, J. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910. doi:10.1037/0022-3514.76.6.893
- Conrad, B., & Schönle, P. (1979). Speech and respiration. *Archiv für Psychiatrie und Nervenkrankheiten*, *226*(4), 251–268. doi:10.1007/BF00342238
- Couper-Kuhlen, E. (1996). The prosody of repetition. On quoting and mimicry. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation* (pp. 366–405). Cambridge, England: Cambridge University Press.
- Couper-Kuhlen, E. (2004). Prosody and sequence organization: The case of new beginnings. In E. Couper-Kuhlen & C. E. Ford (Eds.), *Sound patterns in interaction: Cross-linguistic studies from conversation* (pp. 335–376). Amsterdam, The Netherlands: John Benjamins.
- Dault, M. C., Yardley, L., & Frank, J. S. (2003). Does articulation contribute to modifications of postural control during dual-task performance? *Cognitive Brain Research*, *16*, 434–440. doi:10.1016/S0926-6410(03)00058-2
- De Jaegher, H., Peräkylä, A., & Stevanovic, M. (2016). The co-creation of meaningful action: Bridging enaction and interactional sociology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150378. doi:10.1098/rstb.2015.0378
- De Looze, C., Oerte, C., Rauzy, S., & Campbell, N. (2011). Measuring dynamics of mimicry by means of prosodic cues in conversational speech. In W. S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 17–21 August, 2011, Hong Kong (pp. 1294–1297). Hong Kong: City University of Hong Kong.
- Feldman, A. G. (1966). Functional tuning of the nervous system during control of movement or maintenance of a steady posture: III. Mechanographic analysis of the execution by man of the simplest motor tasks. *Biophysics*, *11*, 766–775.
- Fuchs, T., & De Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, *8*(4), 465–486. doi:10.1007/s11097-009-9136-4
- Fusaroli, R., & Tylén, K. (2012). Carving language for social coordination: A dynamical approach. *Interaction Studies*, *13*(1), 103–124. doi:10.1075/is.13.1.07fus
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition*, *27*(2), 181–218. doi:10.1016/0010-0277(87)90018-7
- Giles, H., Coupland, N., & Coupland, J. (Eds.). (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge, England: Cambridge University Press.
- Giles, H., Mulac, A., Bradac, J. J., & Johnson, P. (1987). Speech accommodation theory: The first decade and beyond. In M. McLaughlin (Ed.), *Communication yearbook 10* (pp. 13–48). Newbury Park, CA: Sage.
- Giles, H., & Powesland, P. F. (1975). *Speech styles and social evaluation*. New York, NY: Academic Press.

- Goldberg, J. A. (2004). The amplitude shift mechanism in conversational closing sequences. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 257–297). Amsterdam, The Netherlands: John Benjamins.
- Gupta, R., Bone, D., Lee, S., & Narayanan, S. (2016). Analysis of engagement behavior in children during dyadic interactions using prosodic cues. *Computer Speech & Language*, 37, 47–66. doi:10.1016/j.csl.2015.09.003
- Hari, R., Himberg, T., Nummenmaa, L., Hämäläinen, M., & Parkkonen, L. (2013). Synchrony of brains and bodies during implicit interpersonal interaction. *Trends in Cognitive Sciences*, 17(3), 105–106. doi:10.1016/j.tics.2013.01.003
- Heritage, J., & Raymond, G. (2005). The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly*, 68(1), 15–38. doi:10.1177/019027250506800103
- Himberg, T., Hirvenkari, L., Mandel, A., & Hari, R. (2015). Word-by-word entrainment of speech rhythm during joint story building. *Frontiers in Psychology*, 6, 797. doi:10.3389/fpsyg.2015.00797
- Himberg, T., & Thompson, M. R. (2011). Learning and synchronising dance movements in South African songs—Cross-cultural motion-capture study. *Dance Research*, 29(2), 303–328.
- Houtkoop, H. (1987). *Establishing agreement: An analysis of proposal-acceptance sequences*. Dordrecht, The Netherlands: Foris.
- Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, 27(6), 949–960. doi:10.1521/soco.2009.27.6.949
- Iivonen, A. (1998). Intonation in Finnish. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 322–338). Cambridge, England: Cambridge University Press.
- Jeong, B. Y. (1991). Respiration effect on standing balance. *Archives of Physical Medicine and Rehabilitation*, 72(9), 642–645.
- Kimbara, I. (2006). On gestural mimicry. *Gesture*, 6(1), 39–61. doi:10.1075/gest.6.1
- Ladefoged, P., & McKinney, N. P. (1963). Loudness, sound pressure, and subglottal pressure in speech. *The Journal of the Acoustical Society of America*, 35(4), 454–460. doi:10.1121/1.1918503
- LaFrance, M. (1982). Posture mirroring and rapport. In M. Davis (Ed.), *Interaction rhythms: Periodicity in communicative behavior* (pp. 279–298). New York, NY: Human Sciences Press.
- Lennes, M., Stevanovic, M., Aalto, D., & Palo, P. (2015). Comparing pitch distributions using Praat and R. *Phonetician*, 111–112, 35–53.
- Li, X. (2014). Leaning and recipient intervening questions in Mandarin conversation. *Journal of Pragmatics*, 67, 34–60. doi:10.1016/j.pragma.2014.03.011
- Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, 36(8), 1404–1426. doi:10.1111/cogs.2012.36.issue-8
- Lundquist, L. O., & Dimberg, A. (1995). Facial expressions are contagious. *Journal of Psychophysiology*, 9, 203–211.
- Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, 34(6), 419–426. doi:10.1016/j.evolhumbehav.2013.08.001
- Miles, L. K., Griffiths, J. L., Richardson, M. J., & Macrae, C. N. (2010). Too late to coordinate: Contextual influences on behavioral synchrony. *European Journal of Social Psychology*, 40(1), 52–60.
- Mondada, L. (2006). Participants' online analysis and multimodal practices: Projecting the end of the turn and the closing of the sequence. *Discourse Studies*, 8(1), 117–129. doi:10.1177/1461445606059561
- Natale, M. (1975). Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills*, 40(3), 827–830. doi:10.2466/pms.1975.40.3.827
- Neuhoff, J. G., Wayand, J., & Kramer, G. (2002). Pitch and loudness interact in auditory displays: Can the data get lost in the map? *Journal of Experimental Psychology: Applied*, 8(1), 17–25.
- Ogden, R. (2006). Phonetics and social action in agreements and disagreements. *Journal of Pragmatics*, 38(10), 1752–1775. doi:10.1016/j.pragma.2005.04.011
- Paxton, A., & Dale, R. (2013). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods*, 45(2), 329–343. doi:10.3758/s13428-012-0249-2
- Putnam, W., & Street, R. L. (1984). The conception and perception of noncontent speech performance: Implications for speech accommodation theory. *International Journal of the Sociology of Language*, 46, 97–114.
- Ramseyer, F., & Tschacher, W. (2008). Synchrony in dyadic psychotherapy sessions. In S. Vrobel, O. E. Roessler, & T. Marks-Tarlow (Eds.), *Simultaneity: Temporal structures and observer perspectives* (pp. 329–347). Singapore: World Scientific.
- Richardson, D., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29(6), 1045–1060. doi:10.1207/s15516709cog0000\_29
- Richardson, D., Dale, R., & Shockley, K. (2008). Synchrony and swing in conversation: Coordination, temporal dynamics and communication. In I. Wachsmuth, M. Lenzen, & G. Knoblich (Eds.), *Embodied communication in humans and machines* (pp. 75–93). Oxford, England: Oxford University Press.
- Rimmer, K. P., Ford, G. T., & Whitelaw, W. A. (1995). Interaction between postural and respiratory control of human intercostals muscles. *Journal of Applied Physiology*, 79(5), 1556–1561.
- Rossano, F. (2012). *Gaze behavior in face-to-face interaction* (PhD dissertation). Radboud University, Nijmegen, The Netherlands.
- Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge, England: Cambridge University Press.

- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1), 227–256. doi:10.1016/S0167-6393(02)00084-5
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–76. doi:10.1016/j.tics.2005.12.009
- Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), 201–208.
- Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, 1(2), 305–319. doi:10.1111/tops.2009.1.issue-2
- Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 326–332.
- Sorjonen, M.-L. (2001). *Responding in conversation: A study of response particles in Finnish*. Amsterdam, The Netherlands: John Benjamins.
- Stevanovic, M. (2012a). Establishing joint decisions in a dyad. *Discourse Studies*, 14(6), 779–803. doi:10.1177/1461445612456654
- Stevanovic, M. (2012b). Prosodic salience and the emergence of new decisions: On the prosody of approval in Finnish workplace interaction. *Journal of Pragmatics*, 44(6), 843–862. doi:10.1016/j.pragma.2012.03.007
- Stevanovic, M. (2013). Deontic rights in interaction: A conversation analytic study on authority and cooperation (Doctoral dissertation). Department of Social Research, University of Helsinki, Helsinki, Finland.
- Stevanovic, M., & Kahri, M. (2011). Puheäänen musiikilliset piirteet ja sosiaalinen toiminta [Social action and the musical aspects of speech]. *Sosiologia*, 48, 1–24.
- Stevanovic, M., & Peräkylä, A. (2012). Deontic authority in interaction: The right to announce, propose and decide. *Research on Language & Social Interaction*, 45(3), 297–321. doi:10.1080/08351813.2012.699260
- Stivers, T., & Rossano, F. (2010). Mobilizing response. *Research on Language & Social Interaction*, 43(1), 3–31. doi:10.1080/08351810903471258
- Stoffregen, T. A., Pagulayan, R. J., Bardy, B. G., & Hettinger, L. J. (2000). Modulating postural control to facilitate visual performance. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2), 203–220.
- Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, 11(2), 139–169. doi:10.1111/hcre.1984.11.issue-2
- Szczepek Reed, B. (2006). *Prosodic orientation in English conversations*. Basingstoke, England: Palgrave MacMillan.
- Szczepek Reed, B. (2009). Prosodic orientation: A practice for sequence organization in broadcast telephone openings. *Journal of Pragmatics*, 41(6), 1223–1247. doi:10.1016/j.pragma.2008.08.009
- Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology*, 46(4), 693–695. doi:10.1016/j.jesp.2010.03.004
- Waaramaa, T., Laukkanen, A.-M., Airas, M., & Alku, P. (2010). Perception of emotional valences and activity levels from vowel segments of continuous speech. *Journal of Voice*, 24(1), 30–38. doi:10.1016/j.jvoice.2008.04.004
- Yardley, L., Gardner, M., Leadbetter, A., & Lavie, N. (1999). Effect of articulation and mental tasks on postural control. *Neuroreport*, 10(2), 215–219. doi:10.1097/00001756-199902050-00003

## Appendix

### Glossing abbreviations

PL	plural
ILL	illative (“into”)
ALL	allative (“onto”)
INF	infinitive
Q	question clitic
CLI	particle clitic
NEG	negation
PASS	passive
ADV	adverb
DEM	demonstrative pronoun
PRT	particle
LETT	name of a letter in the alphabet

Singular, third person, nominative, active, and present tense are forms that have been considered unmarked. These have not been glossed.