Grounding Functions of Instrument Plays in Dialogue: a Case-Study of Piano Duos in Joint Practice

Chika Oshima[1] and Atsushi Shimojima[1,2]

[1] School of Knowledge Science
Japan Advanced Institute of Science and Technology
1-1 Asahi-dai, Tatsunokuchi, Nomi-gun, Ishikawa 923-1292, Japan

[2] ATR Media Information Science Labs
2-2 Hikari-dai, Seika, Soraku-gun, Kyoto 619-0288, Japan

{cooshi,ashimoji}@jaist.ac.jp

Abstract

This paper reports an exploratory study of the grounding functions of instrument plays in dialogue. Grounding theories of dialogue (Clark & Schaefer 1989, Clark 1996, Traum 1994) have been mainly modeled on exchanges in spoken language, and the contributions of non-verbal media to the grounding process are a largely untouched area. We used the Clark-Schaefer contribution model as a core theory, and analyzed six dialogues conducted in joint practice session of a piano duo play. We found that the two general functions of external representations, namely, a cognitive function for personal manipulation of information and a communicative function for interpersonal presentation of information, were co-present and integrated in piano plays in our data.

A series of studies by Clark and others (Clark & Schaefer 1989, Clark 1996, Traum 1994) has persuasively shown that many features of human dialogue can be understood as measures and endeavors for *grounding*, the process to increase a set of beliefs that are agreed upon by all participants of dialogue. The set of beliefs thus established at a given stage of dialogue are called the *common ground* of that stage, and any beliefs in the common ground can be presupposed in subsequent stages. Participants are engaged in grounding as a joint project, and they ensure effective grounding by taking various measures, including efficient presentation of information and compact exchanges of understanding signals.

Although grounding can involve both verbal and non-verbal presentation of information in principle, theories of grounding has been mainly modeled on face-to-face dialogue conducted by spoken language. The grounding process involving a substantive use of non-verbal representations such as gestures, drawings, and instrument plays, is largely an open question. How do people establish a common ground when a non-verbal representation system is available in addition to spoken language? Is the grounding process significantly different from the one involving only spoken language? If so, what exactly are the impacts of the presence of a non-verbal system on the grounding process?

In addition to a long tradition of semiotic research, there has been a renewed interest in the cognitive function of external non-verbal representations, especially of drawings (e.g., Zhang & Norman 1994, Shimojima 1995, Stenning & Oberlander 1995, Suwa &Tversky 1997).

The focus of research, however, has been on the way external non-verbal representations facilitate a *personal* process of designing and problem-solving, rather than on their functions in communicative settings such as dialogue.

Some recent research has focused on the impacts of non-verbal representations on spoken dialogue. For example, McNeil (1995) was concerned with gestures in face-to-face conversation; Neilson & Lee (1994) with drawings in collaborative design, and Umata et al. (2000) with drawings in collaborative problem solving. These studies, however, have been rather sporadic, examining different aspects of dialogue on which non-verbal representations extend their impacts. None have studied their impacts on the grounding process particularly.

Clark & Brennan (1991) offers a conceptual comparison of different communicative settings (e.g., face-to-face dialogue, telephone dialogue, letter correspondence, video conferencing) in terms of their impacts on the grounding process. Communicative settings generally restrict the available types of representations. For example, telephone conversation allows only spoken language but an addition of a shared whiteboard introduces written text and various types of drawings. Thus, Clark & Brennan's research should have some direct implications on the impacts of certain non-verbal media on the grounding process. The exact implications are yet to be explored however.

In this paper, we are interested in the impacts of instrumental plays on grounding processes. Instrumental plays are non-verbal representations, just as drawings and gestures are. They are however *non-persist* representations that occur and diminish in a limited time span. In this respect, they differ from drawings whose traces usually remain on drawing surfaces until somebody erases them. Instrument plays differ from written text in the same sense. They also differ from gestures in being *auditory* media as opposed to visual media. Instrument plays are actually closer to spoken language in these respects.

Their distance from more typical non-verbal representations and their closeness to spoken language make instrument plays a particularly interesting subject of study. By comparing the use of instrument plays with that of gestures and drawings, we can abstract away special features of particular non-verbal media and extract more general functional differences between non-verbal repre-

sentations and spoken language in grounding processes.

A study of instrumental plays from the grounding perspective has practical implications too. Ideas about music, concerning how to play it, how it should sound, and how it sounded, are notoriously difficult information to convey in spoken language. Players, composers, instructors, and students however manage to exchange such ideas in everyday conversation. Detailed studies of how people use instrumental plays in dialogue should lead to an analysis of their communicative expertise, suggesting an efficient methodology of presenting and grounding musical ideas.

Method

Subjects

Our subjects consisted of one male pair and one female pair of graduate students, in the age range from 23 to 32. The members of each pair were unfamiliar with each other. One member of each pair had beginning piano-playing skill, having finished a Beyer primer: the other member had intermediate skill, having practiced for about nine years.

Task

The music used in our observation was F. Chopin's *Prelude Op.28–15*, commonly known as *Rain Drop*. We used a shorter version rearranged for duo play, consisting of 28 bars. The part for Primo was significantly easier than the part for Secondo, and was assigned to the beginner member of each pair. The Secondo part was assigned to the intermediate member. Each part was estimated to be easy enough for the respective players to learn after a reasonable amount of practice.

Each pair was asked to jointly practise duo play of *Rain Drop* in three joint practice sessions. The joint sessions for each pair were held over a period of approximately 10 days, with each session being 40 minutes in length. Each subject was instructed to practice her part in advance of the joint practice series, at least until she could play through her part without mistakes. The subjects received limited monetary compensation for both their personal practice and their participation in the joint practice sessions.

Observation setting

The joint practice sessions were held in a sound-proof room, using the piano model YAMAHA C5. We videotaped the subjects' hand movements on the music scores on the piano keys, the subjects' faces, and their foot movements on the pedals with separate video cameras. The sound and speech were recorded on compact discs through four microphones.

Play units

We transcribed the speech of the six dialogues that we recorded, with special attention to the timing of piano plays relative to speech. For our purpose, a *play unit* is a series of piano sounds, segmeted by (1) the occurrence of a silence followed by speech and (2) a change of players.

Using this definition, we identified a total of 404 play units in the six dialogues we recorded. Table 1 shows the distribution of different categories of play units and thier mean durations.

Of the play units we identified, 273 were partial plays of the music by single players, accounting for 67.5% of all play units; 74 were partial plays by pairs; 55 were full plays of the music by pairs; only 2 were full plays by single players. The average duration of play units was 4.2 seconds in the solo/partial case, 17.4 seconds in the duo/partial case, 92.4 seconds in the solo/full case, and 93.0 seconds in the duo/full case.

Contribution Tagging

According to Clark & Schaefer (1989), grounding proceeds through basic units of exchanges, called *contributions*. A contribution consists of the following two stages:

Presentation phase Speaker A presents utterance u for partner B to consider. A does so on the assumption that, if B gives evidence e or stronger, A can believe that B understands what A means by u.

Acceptance phase B accepts utterance u by giving evidence e' that B believes B understands what A means by u. B does so on the assumption that, once A registers evidence e', A will also believe that B understands.

Our immediate goal was to analyze the roles of piano plays in the grounding process, so a natural first step was to apply the Clark-Schaefer model to our data. Thus we attempted to classify utterances and play units in terms of which play units they belong to, and which phase of the contribution they belong to.

The endeavor, however, immediately turned disastrous. Numerous instances of play units in our data resisted straightforward classification of this scheme. The following is an excerpt from our dialogue data:[1]

Secondo: dono kuraikana
          (How fast should it be?)
          <u>Play bar1 [Secondo]: 3.78</u>
                    (2.94)[
                    kono kuraide
                    (Let us play at this tempo
                     or so.)

This sequence occurred at the beginning of the second joint practice session by the female pair. The pair had agreed to play through the music as a starter for this

---

[1]We generally follow the Jefferson convention of transcription. The Japanese trasnscription of each utterance is followed by an English gloss in parentheses. An underlined text represents a play unit, with an indication of the part of music and the duration of play.

Table 1: Frequencies and mean durations of play units, classified to solo/partial plays, duo/partial plays, solo/full plays, and duo/full plays.

categories	frequency	mean duration (seconds)
Solo/Partial	273 (67.6%)	4.2
Duo/Partial	74 (18.3%)	17.4
Solo/Full	2 (0.5%)	93
Duo/Full	55 (13.6%)	92.4
Total	404 (100.0%)	

day's session, and Secondo was apparently trying to find the appropriate tempo for the joint play. In this context, Secondo's question "dono kuraikana (How fast should it be?)" was directed more to herself than to the partner. In fact, without waiting for the partner's response, Secondo went on to play a part of his part, as though she was experimenting on his play. This play, therefore, apparently started as a personal experiment rather than a presentation to the partner. It sounded as a "monologue play", so to speak.

Our data contains many examples of such monologue plays. According the Clark-Schaefer model, such a play cannot be classified neither as a presentation nor an acceptance, since it is not what the player presents for the partner to consider or to register as evidence of understanding.

However, we cannot dismiss such monologue plays as irrelevant to the grounding process. For in some cases, a play that started as a monologue play was used as a presentation to the partner. For example, towards the end of the monologue play in the above excerpt, the speaker made a proposal clearly directed to the partner, saying "kono kuraide (Let us play at this tempo or so)". Here, the speaker used "kono (this)" to refer to the tempo of the play that she just performed. Thus, the play was used as a presentation after all. It was a "monologue play turned to be a public exhibition", so to speak.

As we will see later, our data contained several other types of play units that are "irregular" to the standard Clark-Schafer model of contributions. The grounding functions of piano plays in our data appears more complex, and we had to significantly extend the standard model to do justice to this complexity. The rest of this section describes our strategy.

We focused our attention on the case of solo/partial plays, the dominant case accounting for nearly 70% of all play units in our data. For each play unit in this category, we first judged whether it belongs to a "presentation" or an "acceptance".[2] For our purpose, a *presentation* is a set of verbal or non-verbal signals (such as piano plays) presented for *some* participant of the dialogue to consider, and an *acceptance* is a set of verbal or non-verbal sig-

nals for *some* participant of the dialogue to consider as an evidence of understanding of the prior presentation.

This conception of presentation and acceptance extends the original Clark-Schaefer definition, in that it permits a presentation of signals to the presenter her/himself and an acceptance from the presenter her/himself. Thus, when a play unit was judged to serve to a presentation, we further judged (1) the target of presentation, namely, to whom the presentation was directed, and (2) the acceptor of presentation, namely, from whom the presentation received an acceptance. For the target of presentation, we employed the following criteria:

- A play unit p belongs to a *to-other* presentation if either p itself is a Clark-Schaefer presentation, or the player issues a positive signal that p is a part of a Clark-Schaefer presentation.

- p belongs to a *to-self* presentation if p itself is not a Clark-Schaefer presentation, and before or during p, the player issues no positive signal that p is a part of a Clark-Schaefer presentation.

Note that the to-other case and the to-self case are not mutually exclusive. Both definitions apply if the player issues a relevant signal *only after* p is finished, while p itself is not a Clark-Shaefer presentation and no relevant signal is issued before or during p. We labeled such a play p as belonging to a *to-both* presentations.

Our criteria for the acceptor of presentation was the following:

- A presentation X receives a *from-other* acceptance if X receives a Clark-Schaefer acceptance from the partner.

- X receives a *from-self* acceptance if the presenter her/himself issues a signal of understanding of X as though X were presented by the partner.

A presentation was labeled as *from-both* if it receives both types of acceptances; a presentation was labeled as *from-none* if it receives none of these types of acceptances.

In view of the complexity of information required for tagging, we employed the method of consensus labeling. An inter-labeler reliability test will be administered in future research.

In the following, a presentation is called a *playing presentation* if one or more play units belong to it. Since a single presentation may contain more than one play unit,

---

[2]On the Clark-Schaefer model, a presentation in one contribution usually works as an acceptance in another contribution, and even an entire contribution can work as an acceptance in another contribution. Here, we simplify the terminology by reserving the term "presentation" for those presentations in contributions that do not work as acceptances in other contributions. Any other presentations are called "acceptances".

the number of plays units are generally larger than the number of playing presentations in a given dialogue.

Results

Overall, 260 play units belonged to presentations and 13 to acceptances, making up the total of 225 presentations and 13 acceptances involving piano plays. Thus, most piano plays served as presentation components rather than as acceptance components; most acceptances were done in speech only.

Table 2 shows an analysis of the 225 presentations involving play units in terms of their targets and acceptors. As shown in at the right-most column, a majority of presentations involving plays are directed to the dialogue partner (134 cases, 59.6%), whereas a significant number are directed to the presenter her/himself (86 cases, 38.2%). Some presentations are directed to both presenter and partner (5 cases, 2.2%). Thus, the presenters' behaviors already suggest multiple functions of piano plays in the dialogues we observed.

Turning to the bottom row, we see that about half of the presentations obtain acceptances from the dialogue partner (112 cases, 49.8%), whereas again a significant number of presentations are accepted by none (54 cases, 24.0%) or by the presenter her/himself (46 cases, 20.4%). Some are even accepted by both presenter and partner (13 cases, 5.8%). Thus, the accepters' behaviors also suggested a deviation from the standard grounding exchange consisting of a presentation to and an acceptance from the partner. Therefore, it's no wonder our data looked chaotic at first glance in terms of this standard.

Discussions

What, more exactly, does this result imply about the functions of piano plays in spoken dialogue? Let us discuss its implications by looking more closely at Table 2 and examining actual examples of playing presentations taken from our data.

Two functions of playing presentations

We start with "other/other" cases, namely, playing presentations classified as "to-other" and "from-other". (We use abbreviations such as "other/self" and "self/both" in the same way.)

Other/other case Although Table 2 looks chaotic at first glance, it shows that the most frequent use of piano plays are still for standard grounding exchanges. There are 97 cases of "other/other" presentations (namely, presentations classified as "to-other" and "from-other"), which account for 43.1% of all playing presentations. These are cases where the presentation is directed to the partner and accepted by the partner, constituting a paradigmatic interpersonal grounding exchange. The following is a typical example taken from our data:

Secondo: nanka kono amadare yakara
      (Well, because these... are rain drops,)

Play a G tone several times [secondo]: 2.04
[[
korega renzoku suru janaidesuka=
(this is repeated, isn't it?)
Primo: =hai
      (Yes.)

Here, Secondo makes a confirmation question ("this is repeated, isn't it?") as he plays a G tone repeatedly. Thus, his play clearly works as informational augmentation of his utterance: at minimum, it provides a referent for "kore (this)" by issuing a G tone and emphasizes repetitiveness through repeated issues of the G tone. This reinforced presentation is immiediately accepted by Primo's confirmation ("Yes").

The high frequency of this type of playing presentation indicates that piano plays, just as linguistic utterances, can be integral components of a presentation in standard interpersonal grounding exchanges.[3]

Self/none case Although other/other presentations were most frequent, they are by no means dominant. Accounting for only 43.1% of all playing presentations, they do not even constitute the majority of cases.

We observed the self/none case as the second most frequent type of case (43 cases, 19.1%). In a self/none case, the playing presentation is directed to the presenter herself, and it receives no signals for acceptance. The following is an example from our data:

Primo: chotto matte kudasaine
      (Wait a second, please.)
            [[
            Play bar 1 [primo]:1.39
(Pointing to a part of the music) ja kono-
-henmade zenbuha hikenainode
(Well, shall we play until around here?
 Because I can't play all.)

The first utterance requests a momentary withdrawal from the joint process, and the speaker immediately goes on to a short play of her part. This play, therefore, is a typical to-self presentation. In fact, there is no signal for acceptance from the partner, and the speaker immediately goes on to her next utterance, whose content—a proposal on the next practice playing—indicates her return to the joint process.

Clearly, the piano play in this case performs a different function than the interpersonal grounding function in the other/other case. The play is more personal than communicative, as evidenced by the prior request for withdrawal from the joint process. The point of playing appears to be a personal experiment, for the purpose of checking or

---

[3] As indicated before, we also found several instances of piano plays used as signals of acceptance to partner presentations. This provides further evidence of standard grounding functions of piano plays.

Table 2: Distribution of presentations involving solo/partial play units, classified by the targets and the acceptors of presentation.

	From-self	From-other	From-both	From-none	total
To-self	28(12.4%)	11(4.9%)	4(1.8%)	43(19.1%)	86(38.2%)
To-other	18(8.0%)	97(43.1%)	9(4.0%)	10(4.4%)	134(59.6%)
To-both	0(0.0%)	4(1.8%)	0(0.0%)	1(0.4%)	5(2.2%)
total	46(20.4%)	112(49.8%)	13(5.8%)	54(24.0%)	225(100.0%)

learning how it would sound when a particular sequence of notes is played in a certain way or under a certain condition. The piano is an external site for an informational experiment by the player, just as sketching paper is for architects (Suwa & Tversky 1997), artists (Grabska 2001), and problem-solvers (Zhang & Norman 1994, Stenning & Oberlander 1995).

For case of reference, we may call the function of such piano plays the *cognitive function*, as opposed to the *communicative function* typically performed by piano plays in other/other presentations.

Self/self case The cognitive function is also manifested in self/self presentations. The following is an example from our data:

Secondo: nanyaro
  (What shall I do?)
  Play a C tone many times:3.43
  e::to=
  (Hmm....)
  =Play bar10 [Secondo]: 2.74
  are
  (What?)

In this excerpt, the speaker makes three consecutive utterances, inserting two piano plays in between. He issues no explicit signal that directs his plays to the partner, and in fact no signals of acceptance are issued from the partner. It is the presenter himself who issues acceptance signals in the utterances "e::to (Hmm)" and "are (What?)". Thus, again the play is more personal experimentation than interpersonal exhibition, and its main function is cognitive rather than communicative.

Our data contains 28 cases of self/self presentation, which, together with self/none cases, account for as much as 31.6% of all playing presentations. Consequently, the piano plays contained in dialogue processes are not homogeneous. There is a large class of presentations performing the communicative function but also a comparatively large class performing the cognitive function. Plays with radically different functions can co-exist in dialogue processes.

Continuity of the two functions

This, however, does not mean that the communicative and the cognitive functions are "discontinuous", always distributed to separate classes of presentations. The two functions are actually continuous, as shown by the following three types of playing presentations.

Self/other case The following is a typical self/other presentation, where the presentation is directed to the presenter her/himself but accepted by the partner:

Secondo: mattene
  (Give me a moment.)
  Play a part [primo]: 7.60
        (5.46) [
Primo: aa kimochi motto tameteirutte
        kannji desukane
        (Oh, you're holding a bit longer,
        aren't you?)

Secondo's utterance requests a momentary withdrawal from the joint process between her and the partner, and she goes on to a 7.60 second play. This play, therefore, is a to-self presentation, and its primary function should be cognitive, an experimentation on how it sounds when she plays the partner's part in a certain way. Primo the partner, however, issues an acceptance to this presentation, making a comment on Second's performance. This indicates that certain information about Secondo's preferred way of playing is exhibited to Primo, and Primo in turn signals her comprehension of the information. This completes a grounding process. Thus, Secondo's playing presentation clearly has a communicative effect as well as cognitive effect.

We observe 11 presentations of this type. We also observed 4 cases of self/both presentations, where the playing presentation is accepted by the partner and the presenter her/himself, although the play is directed to the presenter only. These cases together account for 6.7% of all playing presentations. They exemplify one particular way in which the cognitive and communicative functions of piano plays are integrated: a piano play done for personal experimentation publicly exhibits certain information to be accepted by the partner.

Other/self Still another integration pattern is exhibited by other/self cases, where the presentation directed to the partner is accepted by the presenter herself. For example:

Secondo: a ja kon kurai kana
  (Oh, okay, how about something like this?)
  Play a part [secondo]: 1.22=
  =n

(Eh?)

The piano play in this example clearly started out as a demonstration of the content of "kon (this)" in the preceding utterance. Yet, the presentation process is suspended by the presenter's own utterance "n (Eh?)", which signals that the presenter has caught something probably unsatisfactory in his performance. Thus, this play ends up functioning as a personal experiment, returning somewhat negative information.

Our data contains 18 presentations of this type. We also observed 9 cases of other/both presentations where the presentation is accepted by the presenter as well as the partner, although it starts out as a presentation to the partner. The cases of other/self and other/both together make up 12% of all playing presentations. They exemplify a second type of cognitive and communicative integration in playing presentations: a play done for public exhibition functions as personal experimentation to be accepted by the presenter herself.

Both/other case   Although small in number, both/other presentations exemplify the third variety of the cognitive and communicative integration: a play starting with an experimentation function acquires an exdhibition function during the play. The following is an example from our data:

Primo: chotto kyokutan sugitaka
          (Well, a bit too extreme.)
          Play bar 6–7 [primo]: 10.77
          ttekanjide doudesuka
          (...is sounding all right?)
Secondo: hai iidesune
          (Yes, it is good.)

Before and during the play, there is no public signal that it is a part of any Clark-Schaefer presentation. Thus to this point, its main function is personal experimentation. Nevertheless, the play is then embedded in the next utterance that asks for the partner's evaluation of the play. At this point, therefore, the play acquires an additional, interpersonal function of augmenting the content of a question to the partner. The communicative function is attached to the cognitive function during the very piano play.

Conclusion

We extended the Clark-Schaefer model of contribution to analyze the grounding functions of piano plays found in our data. We identified two distinct functions of piano plays, as an expression tool to present complex ideas about a music, and as an experimentation tool to acquire a great amount of information about a performance that would be hard to predict otherwise. These two functions were also continuous in that plays with one function can acquire the other function during or after the performance. Whichever function a play may be intended for, it is necessarily a public event when done in dialogue, and it comes with certain unpredictability even for experienced performers. Thus, a play with one function always has a chance to acquire the other. This plausibly explains particularly rich structures of grounding afforded by instrument plays.

Our study was exploratory in nature, and the above picture must be further tested on larger samples involving a greater number of subject pairs. The playing skills, social relationship, and personal styles of communication of subject pairs may well affect the target/acceptor combinations of playing presentations, generating different distributions of the cognitive and communicative functions of instrument plays. It would be particularly interesting to study how these different distributions in turn affect the efficacy of grounding process, as measured by the types and amount of information successfully grounded in dialogues.

Clark, H. H., & Schaefer, E. E. (1989). Contributing to discourse, *Cognitive science*, *13*, 259–294.

Clark H. H., & Brennan, S. (1991). Grounding in communication. In L.B. Resnick, J. Levine & S. Teasley, (Eds.), *Perspectives on socially shared cognition*. Washington DC: APA Press.

Clark, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Grabska, E. (2001). Emergent shapes in graphical design. In D.M. Dubois, (Ed.), Computing Anticipatory Sytems, *Proceedings of American Institute of Physics*.

McNeil, D. (1995). *Hand and Mind*. Chicago: University of Chicago Press.

Neilson, I., & Lee, J. (1994). Conversations with Graphics: Implications for the Design of Natural Language/Graphics Interfaces. *International Journal of Human-Computer Studies*, *40*, 509–541.

Shimojima, A. (1995). Operational Constraints in Diagrammatic Reasoning. In J. Barwise & G. Allwein (Eds.), *Logical Reasoning with Diagrams* (pp. 27–48). Oxford: Oxford University Press.

Stenning, K., & Oberlander, J. (1995). A Cognitive Theory of Graphical and Linguistic Reasoning: Logic and Implementation. *Cognitive Science*, *191*, 97–140.

Suwa, M., & Tversky, B. (1997). How Do Designers Shift Their Focus of Attention in Their Own Sketches? In M. Anderson (Ed.), *Reasoning with Diagrammatic Representations: Papers from the 1997 AAAI Spring Symposium* (pp. 102–108). Cambridge, MA: AAAI Press.

Traum, D. R. (1994). *A computational theory of grounding in natural language conversation*.University of Rochester.

Umata,I., Shimojima, A., & Katagiri, Y. (2000). Talking through Graphics: An Empirical Study of the Sequential Integration of Modalities. *The Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 529–534).

Zhang, J., & Norman, D. A. (1994). Representations in Distributed Cognitive Tasks. *Cognitive Science*, *18*, 87–122.