

Multimodality and the emergence of language: an experimental study of spontaneous non-linguistic
communication

Vinicius Macuch Silva

Max Planck Institute for Psycholinguistics

Abstract

In face-to-face interaction people communicate by means of both speech sounds and visible body gestures. Situated linguistic communication is thus characterized by the employment of both the vocal-auditory and the visuospatial modality. Yet, despite extensive investigation of how speech and visible bodily behavior are combined in modern human language, little is known about how vocal and visual signals might have supported the emergence and early evolution of language. In this article, I report on a laboratory experiment which was used to investigate how improvised multimodal signaling can bootstrap communication in the absence of conventionalized language. Contrary to previous literature, the results of the present study show that multimodality has advantages over unimodal gestural signaling in certain scenarios. Ultimately, the findings demonstrate that both the visual and the auditory modality can be fruitfully exploited in scenarios where people communicate devoid of both verbal language and conventionalized non-linguistic signs.

Keywords: improvised multimodal communication, acoustic and visual signaling, non-linguistic communication, language emergence, language evolution

1. Background

Theories of language origins and evolution have for the most part been polarized in terms of speech or gesture-first accounts of early human communication. More recently, multimodal accounts placing emphasis on both speech and gesture have entered the theoretical landscape of linguistic evolutionary research. Support for such multimodal accounts of language evolution stems primarily from the understanding of modern linguistic behavior: speech and gesture are integrated across different linguistic timescales, playing complementary roles in the acquisition, processing, and situated use of language. However, despite extensive investigation of how vocal and visible bodily behavior are combined in modern language, one dimension relating to multimodal communicative behavior remains little explored: how is it that vocal and visual signals might come together in establishing communication from the ground up and creating language anew? This is a relevant question to ask when thinking of how new languages emerge, and it can be particularly informative when thinking of how humans first started communicating. Focusing on such an evolutionary dimension of language, I report on a laboratory experiment testing the extent to which acoustic and visual signals can be used to bootstrap communication in the absence of prior communicative conventions. The results show differences between unimodal and multimodal signaling in regard to participants' accuracy and efficiency in describing auditory and visual stimuli. Ultimately, such findings demonstrate how the natural affordances of the visual and auditory modality are best exploited in interactive scenarios where people communicate devoid of verbal language.

1.1 Modern multimodal communicative behavior

When communicating in face-to-face scenarios people draw not only on their voice but also on visible signals produced with bodily articulators such as the face and the hands. Human communication can thus be characterized as a multimodal form of signaling which rests on the employment of both acoustic and visual signals. Indeed, human communication as a whole, and the situated use of language in particular, are perhaps best understood as phenomena involving the co-articulation of two distinct semiotic modalities: the vocal-auditory modality on the one hand, and the visuospatial modality on the other.

Many accounts of language use in spontaneous interaction underscore the multimodal character of linguistic communication. Naturalistic research conducted in the field has shown that people of different cultural and linguistic backgrounds employ verbal as well as non-verbal signals when communicating to each other in everyday contexts (e.g., Enfield, 2009; Kendon, 2004). In these contexts, since interaction is, above all, proximal and direct, language users attend not only to the vocal signals produced by their fellow interactants but also to the various sorts of visual behavior which are produced by using one's body and which anchor communication in the immediate surrounding environment (e.g., Seyfeddinipur & Gullberg, 2014; Stivers & Sidnell, 2005).

Evidence of how the vocal and visual modality are combined in language also stems from the laboratory: researchers interested in the interaction between verbal language and visual behavior have been showing that vocal and visual signals are integrated both at a cognitive and at a neural

level (e.g., Holler & Beattie, 2003a; Ozyurek, Willems, Kita, Hagoort, 2007). Indeed, experimental research conducted in the lab has demonstrated that visual behavior is fundamental to the processing of language; both when producing speech and when comprehending it people combine verbal information with that derived from visual modes of communication, including eye gaze (e.g., Hanna & Brennan, 2007), manual gestures (e.g., Kelly, Ozyurek, & Marris, 2009), as well as other visible bodily behavior such as head movements (e.g., Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004).

Regarding the types of visual behavior one encounters coupled with language, perhaps the most well studied are movements produced with the hands and the arms. Manual gestures are known to assume multiple forms and to serve various communicative functions (Bavelas, Chovil, Lawrie, & Wade, 1992), iconic hand gestures in particular being a powerful non-verbal means of expression and representation. Representational gestures are typically produced in tight temporal and semantic integration with accompanying speech, aiding language users both in comprehending others (e.g., Kelly, Barr, Church, Lynch, 1999) and in expressing themselves more effectively (e.g., Holler & Beattie, 2003b). Co-speech gestures in general are also helpful in managing turns at conversation (e.g., Mondada, 2007) and in indexing relevant information both in the proximate and distal physical environment (e.g., Goodwin, 2003). Co-occurring with hand gestures are the gestures produced with the face, which together with more pronounced movements of the head are essential in signaling alignment and providing online feedback in face-to-face dialogue (e.g., Chovil, 1991; McClave, 2000). Eye gaze, hand movements, and body posture can also be combined to create multi-layered displays which convey not only general

understanding but also one's level of engagement in a conversation (e.g., Bull, 2016; Shockley, Richardson, & Dale, 2009).

Aside from its fully-developed and integrated use in adult language, visual behavior also plays a fundamental role in the acquisition of language by children. Visible gestures are known to precede, and to some extent predict, the ontogenetic development of language (Iverson & Goldin-Meadow, 2005), gestures serving as a scaffold to the very emergence of speech. The gestures children produce encompass a wide spectrum of forms and functions, varying from context-dependent deictic pointing to the symbolic use of conventionalized and iconic gestures (Acredolo & Goodwyn, 1988; Guidetti, 2002). As their first words emerge, children begin coupling their gestures with speech, thus communicating increasingly more multimodally (e.g., Capirci, Iverson, Pizzuto, & Volterra, 1996; Ozcaliskan & Goldin-Meadow, 2005), early communication relying on the support of not only manual gestures but also eye gaze and other facial cues (Brooks & Meltzoff, 2005; Flom, Lee, & Muir, 2007; Walker-Andrews, 1997; see also Emery, 2000).

1.2 Multimodality and the origins of human language

What both naturalistic and experimental research shows is that speech is complemented by visual behavior not only in regard to the content of what's being said, but also in regard to the pragmatic and interactional information which accompanies the more referential or semantic aspects of language. Together, verbal and visual behavior constitute unified semiotic packages

which language users attend to when communicating to one another in interactive settings. Indeed, research has helped sediment the view that language is a multimodal phenomenon of communication. However, a holistic and truly integrated theory of language assumes not only the understanding of linguistic phenomena *per se* (whether in terms of usage or structure), but also the understanding of the origins and evolution of what is considered to be modern human language. Such an object of inquiry is at the center of research on language evolution, a field which has seen a rapid increase in empirical investigations ever since its resurfacing as a respectable scientific enterprise (Christiansen & Kirby, 2003).

The bulk of theorizing concerning the role of modality in the emergence of human language can be divided into two poles, one defending the centrality of speech and vocal communication, the other stressing the importance of manual action and visible gesture. The received view is that language must have been realized in the vocal modality from its very onset, as vocal signaling is the main form of linguistic communication across modern human populations. In opposition to speech-first theories of language evolution stand those which suggest that the manual modality has played a major role in the early stages of human communication (e.g., Arbib, Liebal, & Pika, 2008; Gentilucci & Corballis, 2006). So-called gesture-first theories of language evolution have gained mounting popularity in recent years, mostly due to advances in comparative and neurobiological research which highlight the ties between manual praxis and visual-gestural signaling in both humans and non-human primates (for a short summary of this research in the context of language evolution, see Kanero, 2014).

Further evidence that visual-gestural communication can bootstrap the emergence of language is found in modern-day sign systems. Indeed, homesigns, impromptu communication systems created by deaf children born to hearing parents, exemplify how gestural communication, mostly iconic in nature, can give rise to systematically structured communication systems. Goldin-Meadow, Mylander, & Franklin (2007), for instance, showed that deaf children can establish new communicative conventions by spontaneously communicating with their hearing parents. Interestingly, not only are children able to create new inventories of form-meaning associations out of hand gestures, these inventories also develop combinatorial structure both at the lexical and the syntactic level. In other words, sets of motivated gestural displays created by deaf children can adopt systematic structure akin to that of language over repeated use in situated communication. In fact, emerging sign languages are living proof that, given sufficient use and transmission, unconventionalized sign systems can evolve and acquire increasingly more complex linguistic structure (e.g., Sandler, Meir, Padden, & Aronoff, 2005; Senghas, Kita, & Ozyurek, 2004).

Despite evidence that visible gesture can lead to the emergence of systematic systems of signed communication, the question of how early gestural communication might have evolved into multimodal language remains unclear. Indeed, gesture-first approaches to language evolution are criticized for failing to explain the discontinuity between the primarily vocal communication of modern human beings and the gestural communication of apes and primates it is associated with.

Combining evidence from different fronts and bridging the gap between the two opposing sides is another approach to the emergence of language, one which considers that human communication is and has, for the most part, always been multimodal. So-called multimodal or gesture-plus-speech theories of language evolution have as their central premise the fact that both vocal and visual modes of communication are central to human communication (e.g., Kendon, 2009; Levinson & Holler, 2014; McNeill, 2012). On the one hand, evidence in support of such theories is drawn from modern human communicative behavior (see section 1.1), while on the other it rests on neurobiological findings which point to the interrelatedness between vocal and visual behavior not only in modern humans but possibly too in other species, both extant and extinct, in the phylogenetic line leading to hominins (e.g., Taglialatela, Russell, Schaeffer, & Hopkins, 2011). Indeed, studies have started showing that, in modern humans, vocal and manual activity have deep neurological ties, such findings suggesting an old evolutionary coupling between the two (for a discussion of the neurological evidence in the context of language evolution, see Kendon, 2009). Moreover, as highlighted by ethological and comparative research, humans are not alone in being multimodal communicators: non-human apes as well as other primates seem to combine vocal and visual signals in their communication too (Liebal, Waller, Slocombe, & Burrows, 2013; see also Slocombe, Waller, & Liebal, 2011).

1.3 Studying the emergence and evolution of human language in the wild and in the lab

Since traces of early forms of human language cannot be unearthed, researchers interested in understanding its origins and early evolution often rely on indirect sources of evidence. One such

source is the study of modern language emergence: pidgins and creoles, for instance, are studied in order to understand how impromptu communication systems evolve into fully-fledged natural languages (e.g., Arends, Muysken, & Smith, 1994; Roberge, 2008). Similarly, research on homesigns and emerging sign languages provides insightful clues about how communication can be bootstrapped in the visual modality and about how it can acquire systematic linguistic structure over time (e.g., Aronoff, Meir, Padden, & Sandler, 2008; Senghas, Kita, & Ozyurek, 2004). Together with naturalistic approaches, experimental methods have also made their way into linguistic evolutionary research. Indeed, studying how artificial languages are created in the lab has become a serious method for testing hypotheses about language emergence and evolution. To that end, experimental semioticians look at how human participants bootstrap new communication systems altogether and at how unstructured semiotic inventories adapt to become more language-like in nature (Galantucci & Garrod, 2011; Galantucci, Garrod, & Roberts, 2012; for similar work conducted with non-human agents, see Skyrms, 2010; Steels, 1999).

Existing experimental paradigms make use of communication games which allow experimenters to manipulate not only what participants communicate about, but also, crucially, how they go about in doing so. Previous studies have explored, for instance, the cognitive preconditions necessary for the emergence of language, from how people signal communicative intent (e.g., de Ruiter et al., 2010), to how they establish common ground (e.g., Scott-Phillips, Kirby, & Ritchie, 2009). Recently, studies have started exploring how different semiotic modalities can be used to communicate in the absence of verbal language, including how the use of combined modalities might affect improvised communication. The findings yielded by these studies stress the power

of visual-gestural signaling, as well as the tight link between visible gesture and the origins of human communication.

One such study is of particular relevance to the experiment here reported. In Fay, Arbib, & Garrod (2013), participants were faced with the task of communicating concepts to one another using only the natural affordances provided to them by the vocal and visual modalities. Participants were therefore allocated to experimental conditions in which they could communicate either by producing visible body gestures, non-linguistic vocalizations, or both non-linguistic vocalizations and visible gesture. Participants' communicative efforts were measured both in terms of effectiveness, i.e., how accurate they were at communicating items, and in terms of efficiency, i.e., how fast it took them in guessing the items being communicated (which was done by selecting an option out of a fixed list of options).

The results of the study showed that participants in the gestural condition were much more effective in communicating different categories of items, namely, actions, objects, and emotions, than their counterparts in the vocal condition. Moreover, combining gesture and vocalization did not result in more effective descriptions than those produced with gestures alone, although only 25% of trials in the combined condition included both gesture and vocalization. As for participants' efficiency in communicating using different modalities, objects and actions were communicated faster in the visual conditions than in the vocal one, again there being no differences between gestural and multimodal communication. The authors concluded that participants succeed at the task by basing their descriptions on motivated properties of the

concepts they had to describe. Fay and colleagues argued that this strategy was especially effective when people communicated using visible gestures, as these lend themselves more naturally to motivated representations.

In a subsequent study, Fay et al. (2014) again tested how people communicate in a referential communication task using nothing but bodily gestures and non-linguistic vocalizations. Unlike the previous experiment, participants changed roles after a given number of trials, a twist which allowed the experimenters to measure how much agreement there was between pairs of players and the communicative conventions they established. The results not only replicated Fay et al.'s (2013) previous findings, but they also showed that gestural descriptions were more aligned between players in comparison to both vocal descriptions and multimodal descriptions (i.e., descriptions composed of both sounds and gestures). Thus, as participants described the same items several times, those doing so visually produced descriptions which became more conventionalized than those of participants who communicated vocally or even multimodally. Much like in their previous study, the authors suggested that, due to motivated relationships between the gestural descriptions and their semantic referents, the visual modality was particularly suitable for representing concrete concepts tied to participants' bodily experience (as in the case where the concept 'fighting' was described by producing a pantomimic display simulating a fist fight).

Both studies by Fay and colleagues have shown that visible gestures are particularly useful in communicating in the absence of linguistic conventions. Furthermore, they have shown that

multimodality has no advantage over unimodality in improvised communication. In these experiments, however, participants were asked to describe items which referred to conventional semantic meanings, such as the act of sleeping, or the emotion of pain. Thus, in many cases, non-linguistic yet otherwise still conventionalized signs proved appropriate solutions to participants' lack of proper communicative conventions. For example, in order to depict the act of sleeping, one might have simulated a snore, or in order to communicate the feeling of pain one might have screamed or even simulated a cry. Provided enough context, conventional representations of the sort are powerful non-verbal means of expression. As such, in the studies discussed above, the actual role of modality in motivating spontaneous non-linguistic descriptions might have been obscured by the use of conventional signs. For the same reason, the combined use of modalities might have proven rather unnecessary given conventionalized signs which were powerful enough on their own. In short, the differences found in how participants communicated using different sensorimotor modalities might be due to the availability of conventional signs in a given modality rather than to the actual representational affordances it provides.

Taking into account such considerations, it remains unclear what role modality might play in getting communication off the ground. As such, next I report on an experiment which addresses the following question: how do people communicate in the absence of semiotic conventions using only the natural affordances of the vocal and visual modalities? Crucially, in order to avoid eliciting conventional non-linguistic signs, I introduce novel stimuli which do not refer to conventionalized entities, actions, or qualities.

2. The present study

In order to investigate how people communicate on the basis of spontaneously created signals, pairs of participants were asked to describe novel meanings to one another using non-linguistic vocalizations and visible gestures. Consisting of sounds and images (see section 3.2.1 for more details), the stimuli were designed so that both the vocal-auditory modality and the visuospatial modality could be used to generate perceptually motivated descriptions of those same stimuli. As such, by the very design of the experiment, auditory stimuli should allow for more motivated acoustic descriptions, whereas visual stimuli should allow for more motivated visual descriptions. The motivated potential of each signaling modality is summarized in Table 1.

		Signaling modality	
		Acoustic	Visual
Stimuli	Auditory (sounds)	Motivated	Not motivated
	Visual (images)	Not motivated	Motivated

Table 1. Summary of the motivated potential of each signaling modality (acoustic vs. visual) in relation to each stimulus type (auditory vs. visual).

By comparing unimodal acoustic and visual signaling to multimodal (i.e., acoustic + visual) signaling in an interactive communication game, I explore how the use of combined modalities might affect improvised communication. Thus, the study is aimed at understanding:

- (i) how players communicate different perceptual stimuli by representing those stimuli using visual and/ or acoustic signals;
- (ii) whether players are successful in mapping their partners' visual and/ or acoustic representations onto pre-specified stimuli items;
- (iii) whether multimodal signaling grants players any advantages over unimodal signaling in the context of improvised communication.

To this end, participants are asked to perform a referential communication task (Clark & Wilkes-Gibbs, 1986; Yule, 1997). Their performance is measured both in terms of the (i) efficiency with which they communicate items to one another, and in terms of their (ii) accuracy in guessing the communicated items correctly.

As discussed earlier, due to a direct linkage between the perceptual modality in which stimuli are presented and the modality in which people represent those stimuli, participants are expected to perform generally better at the task when describing stimuli which match the modality they are communicating in. Thus, auditory stimuli should be communicated better by players communicating acoustically, whereas visual stimuli should be communicated better by players communicating visually. However, as demonstrated by the studies by Fay et al. (2013, 2014), visible gesture is a particularly powerful means of communication on its own. As such, despite the modality mismatch between the stimuli and the semiotic means by which these stimuli are

represented, participants relying on visible gesture are expected to describe sounds successfully, given the expressive power of gesture.

Differently from participants who can only communicate acoustically or visually, participants who can communicate both acoustically and visually are likely to adjust their signaling behavior according to the situation at hand. As such, when presented with sounds, these participants have the chance to imitate those sounds directly. Similarly, when presented with images, they can represent the stimuli by relying on visual descriptions. Further yet, these participants can communicate by combining both acoustic and visual signals, which might result in improved communication compared to that of their counterparts who communicate strictly unimodally. Having in mind these considerations, we hypothesize that:

- (1) participants who communicate in a signaling modality which allows for motivated representations of the stimuli will perform better than those who communicate in a signaling modality which does not allow for motivated representations;
- (2) participants who communicate in a signaling modality which allows for the use of visible gesture will perform better than those who communicate in a signaling modality which does not allow for the use of visible gesture;
- (3) participants who can communicate multimodally will perform better than those who can only communicate unimodally.

3. Method

3.1 Participants

The experiment was conducted with 15 dyads (30 participants in total, 20 females). All participants were recruited using the participant database of the hosting institution, the Max Planck Institute for Psycholinguistics, and none had any command of any sign language.

3.2 Apparatus and material

The experiment was conducted at the Max Planck Institute for Psycholinguistics, in a facility dedicated to research on gestures and sign language. The equipment included tables and chairs as well as 4 Canon XF 205 HD genlocked cameras (frame rate 29.97P), 3 of which were used to record participants in both video and audio. One camera captured both players, while the remaining 2 cameras were set one at each player (frontal view). Participants performed the task using HP Probook 470 laptops (screen resolution 1600 x 900) which were positioned at a table in front of them. Aside from laptops and other related hardware (i.e., mice, power cables, etc.), participants were provided with headphones which were used to interact with the auditory items.

3.2.1 Stimuli

The stimuli consisted of non-linguistic items, namely sounds and images. Auditory stimuli, described in Table 2, consisted of 8 sounds resembling both generic natural sounds (e.g., animal wings flapping, tree falling) and man-made/ artificial sounds (e.g., paper being crumbled, balloon bursting). Visual stimuli, presented in Fig. 1, consisted of 8 images of circles filled with different patterns and shapes (e.g., lines, triangles, etc.).

Item	Sound description ¹
1	Air leak
2	Paper being crumpled
3	Wings flapping
4	Pages of a book being turned very quickly
5	Boing (onomatopoeic-like bouncing)
6	Door creaking
7	Tree falling
8	Balloon bursting

Table 2. Auditory stimuli set.

¹ It should be noted that the descriptions here provided are based on entirely subjective perceptual judgments. Readers are directed to the supplementary materials for the audio files which were used as stimuli.

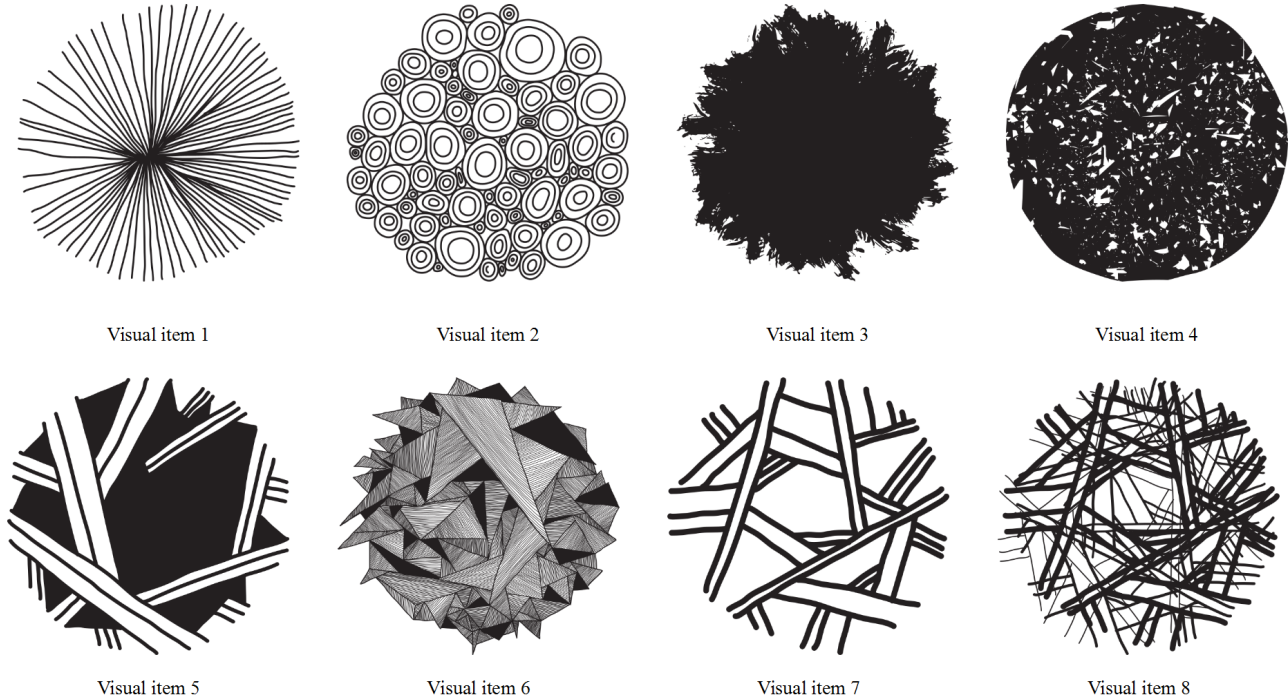


Figure 1. Visual stimuli set.

The stimuli were designed and pre-tested so as to avoid eliciting conventionalized non-linguistic signs such as certain types of visible gesture (e.g., belly rub indicating ‘hunger’) or vocalizations (e.g., mooing indicating ‘cow’). They were initially pre-tested with individual participants, in a non-interactive setting. Participants were asked to perform a routine similar to that of a referential communication task: they had to describe a stimulus presented to them as if a partner had to guess what that stimulus was. Participants described both auditory and visual stimuli, one set at a time, doing so vocally, by means of visible gestures, and by combining both vocalizations and body movements.

In the pretest, individual participants were recorded in all 3 scenarios, their descriptions being subsequently analyzed. The analysis was aimed at understanding whether participants were responding well to the stimuli, or in other words, whether the stimuli were eliciting adequate and sufficient descriptions from participants. Given the inspection of the recorded descriptions, it was decided that all participants had been able to properly describe the stimuli items, having done so in a timely and spontaneous manner, and without relying on pre-established semiotic conventions.

After pre-testing the stimuli with individual participants, the interactive setting in which communication games usually take place was also piloted. Pairs of participants engaged with the stimuli in an improvised communication game. The improvised version of the game followed the same procedure as that of the experiment task: one participant had to describe a target stimulus, while a second participant had to pick a candidate option from an array of options, based on the first participant's description. Participants were recorded playing the game, and their descriptions were subsequently analyzed. As in the non-interactive pilot, players responded well to the stimuli. An inspection of the recorded sessions showed that participants engaged with the task successfully, having faced no apparent problems when describing the stimuli presented to them. Much like individual participants before them, pairs in the interactive game produced timely and spontaneous descriptions, often engaging in conversation-like exchanges of requests and (dis)confirmations. Overall, their descriptions did not take long to produce and did not seem to involve extensive generative effort (as indicated by self-corrections, extended periods of preparation, hesitations, etc.).

Both in the pilot version and in the full-blown experiment, each trial matchers saw two distractors alongside the target item they had to identify. These distractors were selected so as to introduce the need to distinguish between specific perceptual/ semantic dimensions of the stimuli (e.g., specific visual patterns found inside a circle), as opposed to more general dimensions (e.g., circular shape), which meant that each trial one of distractors was perceptually similar to either the other distractor or the target item.

3.3 Design

Given that the aim of the study was to investigate how the affordances of the auditory and visual modality could be exploited in a scenario of restricted non-linguistic communication, the experiment was designed so that the use of both acoustic and visual communicative signals could be compared. As such, separate conditions in which participants could communicate either acoustically or visually were introduced. In order to clarify whether unimodal signaling – both acoustic and visual – differed from multimodal signaling, a third condition was added in which participants could communicate both acoustically and visually.

Dyads were randomly allocated to one of the 3 experimental conditions (n=10 per condition), all dyads alike describing both auditory and visual stimuli. The order in which dyads encountered visual and auditory items was counterbalanced, which is to say that in each condition some dyads described images first, while others described sounds first. Since there were five dyads in each

condition, there was an uneven number of dyads assigned to each starting order. The design of the study is summarized in Table 3.

	Acoustic condition	Visual condition	Multimodal condition
Auditory stimuli	Pair 1 – 5	Pair 6 – 10	Pair 11 – 15
Visual stimuli	Pair 1 – 5	Pair 6 – 10	Pair 11 – 15

Table 3. Study design.

3.4 Task and procedure

Prior to performing the experiment task, participants were submitted to a short training session. In it, they were taught how to navigate the game, including how to interact with the screen. The game itself consisted of a referential communication task in which one of the players had to describe an item while the other player had to correctly guess what that item was, given a fixed set of options. The game was played in alternating trials: each trial players switched between the roles of director – the one describing an item – and matcher – the one guessing what item was being described. Trials were administered in two blocks. In one of the blocks, players had to describe visual stimuli (i.e., images), while in the other block they had to describe acoustic stimuli (i.e., sounds). Each individual stimuli set consisted of 8 items, which were presented once to each player as director (see ‘Playing as a director’), and once as matcher (see ‘Playing as a matcher’). The role of director and matcher alternated with each trial, and the order of the stimuli was randomized. A game thus consisted of 16 trials, and in total participants in each condition

played 4 consecutive games for each stimuli set block, which means there were, overall, 8 games and 128 individual trials.

The experimental manipulation consisted in varying the affordances players had access to in order to communicate to one another. Players were allowed to communicate using only the natural affordances of their body, which were restricted in the following way:

Condition 1 (acoustic) – players could only produce sounds to communicate to one another. These sounds could be produced with the mouth or any other bodily articulator, and could not consist of spoken words or any other form of verbal language (e.g., whistling, clapping hands, hitting the desk). In this condition, players sat back to back and had no visual contact whatsoever. Each player faced one laptop placed on a table opposite to them (see Fig. 2);

Condition 2 (visual) – players could only produce visible body movements to communicate to one another. These visible movements could be produced with the hands or any other bodily articulator, including the face and the torso (e.g., waving arms in the air, tracing shapes on the table, nodding). In this condition, players faced one another, their individual laptops being placed on a single table located between both players, which allowed for direct line of sight and full visual contact (see Fig. 2);

Condition 3 (multimodal) – players could produce both sounds and visible body movements to communicate to one another, following the same restrictions as in Conditions 1 and 2. The setup was identical to that of Condition 2, which meant that players sat facing one another and had full visual contact.



Figure 2. Experimental setup. Panel A illustrates the setup of the acoustic condition, whereas panel B illustrates the setup of the visual and multimodal conditions.

Playing as a director

As directors, players were presented with a target item, which consisted of an image in the case of visual stimuli and of a sound in the case of auditory stimuli. Images were directly available on screen, while sounds could be heard by pressing a ‘play’ button located in the middle of the

screen. Directors had to describe the target item to their partner using nothing but the affordances available to them according to the experimental conditions described above. No explicit instruction was given concerning the nature of the descriptions which were to be produced. Directors were instructed to move to the next trial once their partner had guessed what the target item was, and that was done by pressing a button labeled as ‘next round’. In addition to attending to their partner’s description and making a choice, matchers were instructed to signal their understanding once they knew what was being described, which was meant not only to provide the director with awareness of the matcher’s understanding, but mainly to ensure a streamlined transition between trials. Players were not given feedback regarding whether they had guessed correctly or not.

Playing as a matcher

As matchers, players had to select the correct item based on their partner’s description of that item. Matchers were presented with 3 individual buttons on screen. Buttons were labeled ‘a’, ‘b’, and ‘c’, and by clicking on them matchers were able to activate one of 3 different options, which included, aside from the target item, 2 distractors. In the case of auditory items, by pressing each button, players were able to hear different sounds, whereas with visual items they were able to see different images. Importantly, in order to match the availability of both auditory and visual items, after clicking on a button and activating an image, the image remained visible on screen for 1s, which was meant to mimic the transitoriness of auditory items. Players could move freely

between the 3 options and could hear/ see items as many times as wished. Items were selected by clicking on a button labeled ‘choose’, which was located underneath each option button.

Unrestricted gameplay

As previously explained, the goal of the game was to guess what item was being communicated based on a player’s description of that item. Similarly to previous studies (Fay et al., 2013, 2014), players were able to interact freely, which meant that they could seek clarification rather than simply describing an item (director) and guessing what that item was (matcher). Crucially, matchers could react to their partner’s descriptions, issuing requests of information or candidate descriptions of their own, which means that trials could be extended in case matchers were uncertain about their partners’ descriptions. Directors were free to respond to requests from the side of matchers, which implies a self-managed and open-ended trial time. A trial, therefore, could be as short as the director issuing a description followed by the matcher acknowledging understanding. However, it could also involve players taking various turns in communicating before the matcher finally makes a guess. Players were not instructed in any particular way as for what could be done if a description was not sufficiently informative, yet, they were made aware that clarifications could be requested at any point, these clarifications following the same restrictions as the rest of the game: players could only communicate using non-linguistic sounds and/ or visible body movements.

4. Coding

The data was coded using ELAN (version 4.9.4), an annotation software distributed freely by the Max Planck Institute for Psycholinguistics (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). The coding involved annotating the video recordings for acoustic and visual behavior produced by participants as intentional acts of communication. The communicative behavior participants produced was further annotated in respect to turns each player took in a trial. Each trial consisted of an attempt by a director to communicate a stimulus item to their partner, and for each trial director and matcher could communicate indefinitely until the matcher made a choice and the trial was brought to an end.

4.1 Communicative acts

For the purposes of the annotation, an acoustic act of communication was considered to be whatever sound/ sequence of sounds players produced either by vocalizing (e.g., whistling) or by making use of means other than the voice (e.g., knocking on the table). When players vocalized, false starts and hesitations markers were not annotated. However, together with natural pauses, these were used to delineate a break between one acoustic communicative act and another. A visual act of communication, on the other hand, was considered to be whatever visible movement/ sequence of movements players produced using their body, including the head and the face (e.g., waving arms in the air, tracing shapes on the surface of the table, blinking, etc.). Gesture preparation and retraction, as well as extended freezes both before an initial stroke and

after a final stroke, were not annotated. This means that visual displays were annotated according to a minimal temporal unfolding: annotations started at the first identifiable moment of a visual display (cf. two-hand movement vs. one-hand movement, outstretched fingers vs. closed hand, averted eye gaze vs. directed eye gaze, etc.) and ended immediately after the beginning of a halt or retraction. For instance, a player raises one of their hands and positions it in line of sight of their partner. The player then reviews the target item before moving the hand around in circles. This would have been annotated from the moment the hand moves from its mid-air resting position, and not from the first moment the hand was positioned in line of sight. For an example of how acts of communication were coded, see Fig. 3.

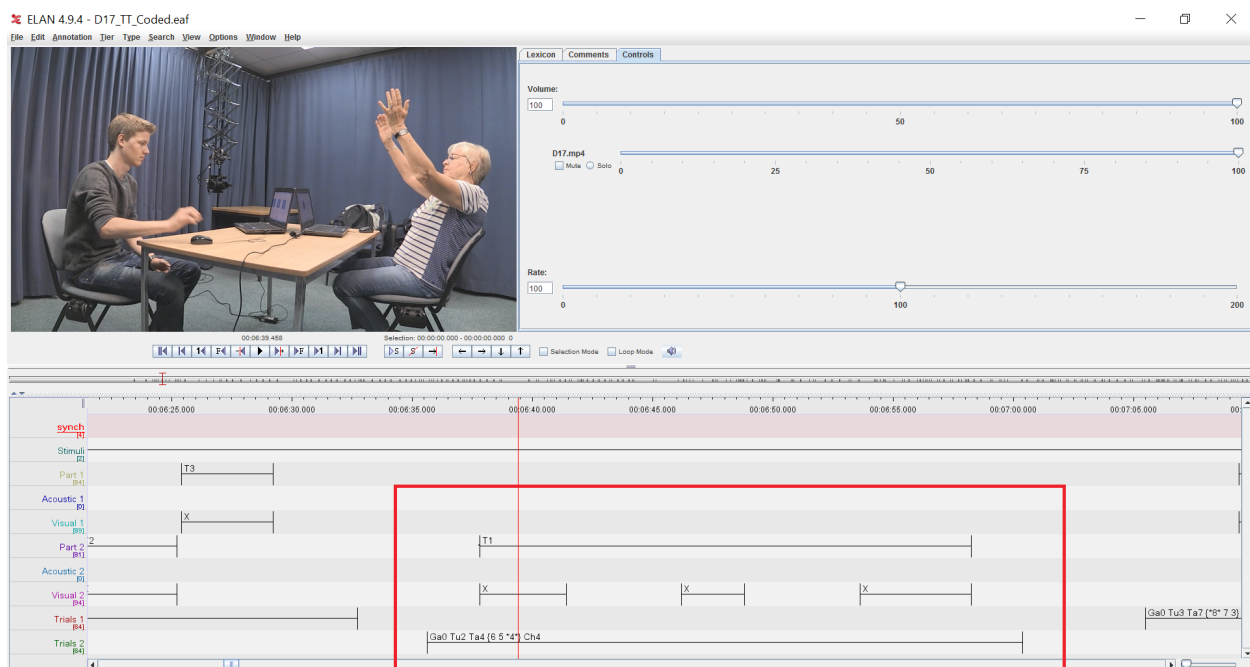


Figure 3. Communicative acts as coded on ELAN. Each annotation labeled with an 'X' constitutes a separate act of communication within a single trial.

In the multimodal condition, players communicated both acoustically and visually. Whether issued concurrently or non-concurrently, acoustic and visual signals were coded separately in all cases, according to the respective modality guidelines, as described above. Communicative acts which matchers produced to signal their understanding of a partner's description were not annotated, as this was a routine which players were explicitly told to follow and which itself followed a pre-specified format.

When coding the data on ELAN, the in-built ruler was used as a reference for establishing the relative beginning and end of annotations (i.e., annotation boundaries). The coding was conservative insofar as, in case of doubt, the start of an annotation would correspond to the indent immediately to the left of the potential starting point, whereas the end of an annotation would correspond to the indent immediately to the right of the potential ending point. Moreover, the ruler was set at different zoom values when coding data from the acoustic condition (100% zoom) and from the visual and multimodal conditions (75% zoom).

4.2 Turns

In previous experimental semiotic studies in which people had to communicate only on the basis of non-linguistic bodily behavior, players either remained in the role of director/ matcher throughout the entire experiment (Fay et al., 2013), or they exchanged roles at the end of a game (Fay et al., 2014). In Fay et al. (2014), aside from measuring how accurate players were in communicating, the authors measured the relative amount of alignment between players'

descriptions. Thus, by comparing their vocal and visual displays from one game to another, Fay and colleagues were able to assess how conventionalized these descriptions were but also how conventionalization was linked to accuracy.

In the present study, rather than analyzing communication in terms of formal alignment between isolated descriptions, we focus on dyads and their interactive efforts, quantifying communication not only in terms of accuracy and efficiency but also in terms of turns taken by players in communicating individual items during the game. Such analysis allows for a more fine-grained understanding of communicative performance, as it provides a window onto what happens within the trials themselves. Ultimately, complementing the main analyses with an in-depth view of turn-by-turn communicative interaction provides clues as for how players managed improvised communication in the first place.

In natural conversation, a turn corresponds to whenever a person holds the floor while speaking. For the purposes of the annotation, a turn in the game consisted of any communicative display produced by a player within a trial without interruption or intrusion from the side of their partner. For instance, a director might describe an item by producing a vocalization, then pause, and upon receiving no response from their partner vocalize once again. Despite the two vocalizations being separate acts of communication, which were to be annotated as such, the whole sequence would be regarded as a single turn, as the director communicates without interruption or intrusion from the side of the matcher. The beginning of a player's turn thus coincided with the beginning of the first communicative act produced by them. The end of a player's turn coincided with the end of a

communicative act produced by that player either before the end of the trial or before the beginning of an intervention by their partner (i.e., beginning of a partner's turn). For an example of how turns were coded, see Fig. 4.

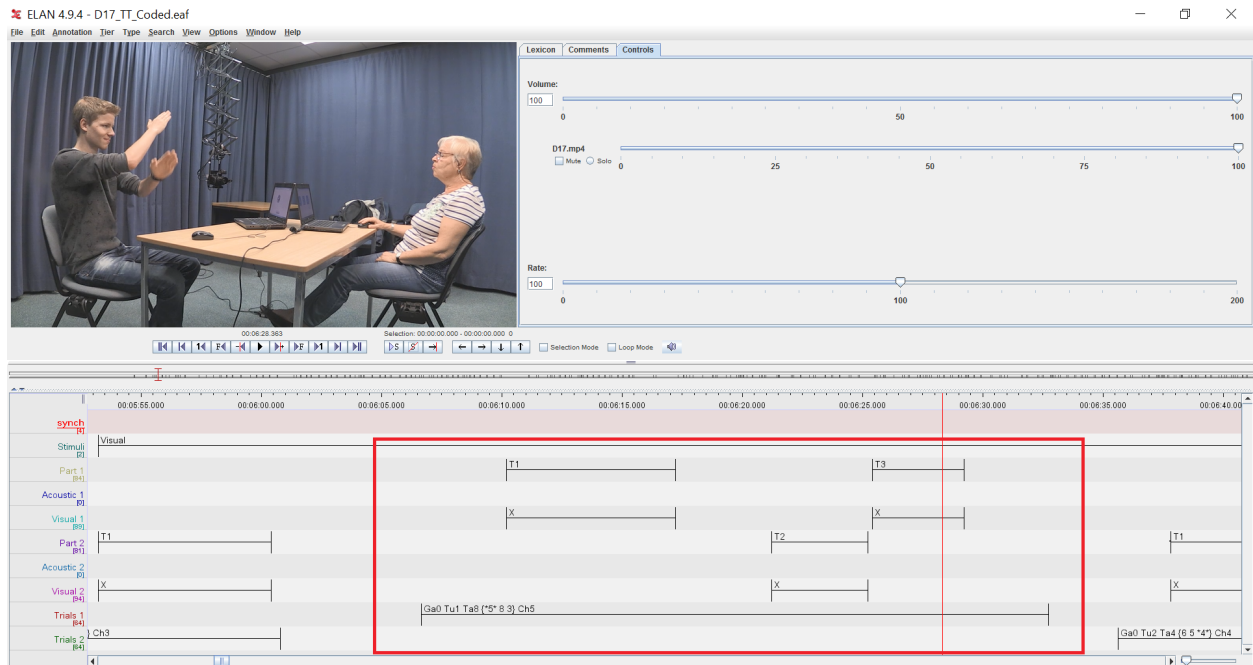


Figure 4. Turns as coded on ELAN. Each annotation labeled with a ‘T’ constitutes a turn. In the case highlighted, the trial is composed of 3 turns: an initial turn by the director (T1), a second turn by the matcher (T2), and a final turn by the director (T3). Each turn contains a single communicative act.

4.3 Post-coding – Accuracy and Efficiency

After coding the data for communicative acts and turns, the annotated ELAN files were processed using the `pympi` Python library, which automated the analysis of players’ communicative performance (see Lubbers & Torreira, 2014). All trials were parsed, and given both the beginning of the first annotation in a trial and the end of the last annotation in a trial, the

program automatically generated a trial annotation, which included information about the target item, the distractors, as well as the matcher's choice. Guessing accuracy and communicative efficiency were directly retrieved from the automated analysis of the data. The accuracy with which participants communicated to one another was measured based on their probability of guessing items correctly, whereas their efficiency in doing so was measured based on the trial length (calculated from the start of a director's first turn up until the moment their partner made a choice). In Fay et al. (2013, 2014), accuracy was measured based on the percentage of items correctly guessed, whereas efficiency (in Fay et al. (2013)) was measured as in the current study.

5. Results

The results of the analyses were inspected using linear mixed effects models, which were themselves scripted using the lme4 package in R (lme4: Bates, Maechler, Bolker, & Walker, 2015; R: R Core Team, 2016). The results of the linear mixed models are presented in the sections below. Readers are directed to the supplementary materials for full analyses and results.

5.1 Accuracy

Given that the experiment task consisted of communicating auditory and visual items acoustically and/ or visually, modality (acoustic vs. visual vs. multimodal), stimulus type (auditory vs. visual), as well as an interaction between the two were introduced as the primary

fixed effects in the model predicting accuracy (i.e., probability of guessing items correctly). As dyads performed the task over successive games, trial (1 - 64) was also added as a fixed effect. Additional fixed factors included two efficiency measures (i.e., total length of a trial, length of the first turn of a trial), the order of the stimuli sets (i.e., visual first vs. auditory first), as well as an interaction between trial length and modality, all of which were added as controls. Finally, a fixed factor predicting accuracy on the basis of multimodal signaling was introduced (see section 5.3 for an explanation). Given that players might have found certain items easier to communicate than others, but also given the fact that individual players, or specific dyads, might have been better communicators than others, item, director, and dyad (with director nested within dyad) were introduced as random effects.

The main results indicate that participants managed to communicate successfully on the basis of unconventionalized acoustic and visual displays (see Fig. 3). Indeed, participants in all three conditions performed the task above chance level, with no significant differences in accuracy across conditions (no main effect of modality) yet with a significant increase in accuracy across games (main effect of trial, $\chi^2(1)=51.60$, $p<.001^2$). There were no statistical differences between participants' performance in different conditions when it came to communicating about sounds. However, when communicating about images, participants in the acoustic condition performed worse compared to participants in the visual and multimodal conditions (interaction between modality and stimulus type, $\chi^2(2)=15.89$, $p<.001$). In addition to that, it was found that, across all conditions and for both stimuli types, the time players took in describing and choosing items

² All chi-square statistics are derived from the model comparison test, which indicates the amount of variance explained by each new predictor introduced in the analysis.

predicted their likelihood of making a correct guess: incorrect guesses had overall longer trial lengths (main effect of trial length, $\chi^2(1)=41.86$, $p<.001$).

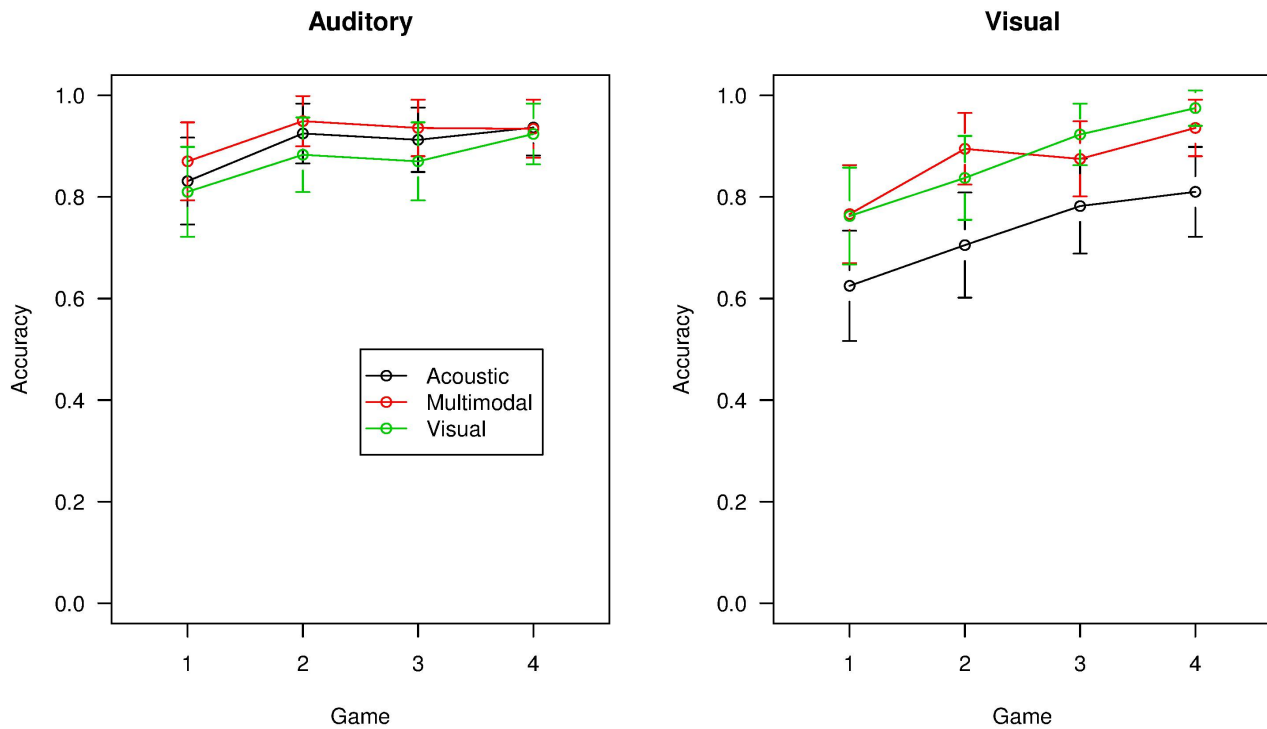


Figure 3. Overall probability of guessing items correctly, according to stimulus type (auditory vs. visual), game (1, 2, 3, & 4), and experimental condition (acoustic vs. visual vs. multimodal).

5.2 Efficiency

As in the analysis of accuracy, modality, stimulus type, as well as an interaction between the two were introduced as the primary fixed effects in the model predicting efficiency (i.e., trial length). Other predictors included game (measured at the trial level) and a non-linear expression of game (i.e., quadratic effect of game), as well as the following controls: multimodal signaling, a measure of communicative accuracy (i.e., whether an item was guessed correctly), a measure of communicative interaction (i.e., the total number of turns in a given trial), and the order of the stimuli sets. Several interactions between the fixed factors were introduced, mainly as additional control for possible co-influence between them: stimulus type x game, modality x game, modality x stimulus type x game, number of turns x modality, number of turns x stimulus type, number of turns x modality x stimulus type, modality x incorrectness, stimulus type x incorrectness, stimulus type x modality x incorrectness, multimodal signaling x stimulus type, modality x quadratic effect of game, stimuli set order x modality. The random effects were the same as in the accuracy analysis, namely item, director, and dyad (with director nested within dyad).

The main results point to an interaction between modality and stimulus type ($\chi^2(2)=14.23$, $p<.001$), which is explained by two separate trends (see Fig. 4). For auditory stimuli, participants in the multimodal condition communicated more efficiently than participants in both the acoustic and visual conditions. For visual stimuli, participants in the acoustic condition communicated more efficiently than participants in the visual and multimodal conditions. It should be noted that

although players in the acoustic condition were more efficient than their counterparts in the two other conditions in communicating images, their guessing accuracy was also considerably lower (interaction between modality and incorrectness, $\chi^2(2)=5.06$, $p<.01$), which is to say that trial times were shorter but at the same time players were less effective in guessing visual items correctly.

Further, the results show that participants in all conditions were more efficient in communicating items at later stages of the experiment than they were at the beginning of it (main effect of trial, $\chi^2(1)=429.07$, $p<.001$). Efficiency increased non-linearly (quadratic effect of game, $\chi^2(1)=51.65$, $p<.001$), which is to say that the rate at which participants became more efficient decreased as the experiment progressed. As players moved from one game to another, those in the multimodal condition became more efficient than those in the unimodal conditions (interaction between modality and game, $\chi^2(2)=13.37$, $p<.01$), the effect also holding for the non-linear expression of game (interaction between modality and quadratic effect of game, $\chi^2(2)=6.24$, $p<.05$).

Finally, as would have been expected, the more turns players took in communicating an item, the longer the trial itself took (main effect of turn number, $\chi^2(1)=500.12$, $p<.001$). Moreover, the longer a trial took, the more likely it was that the matcher's guess would be incorrect (main effect of incorrectness, $\chi^2(1)=26.55$, $p<.001$). Participants in the multimodal condition were found to take more turns to describe visual items than participants in the remaining two conditions (interaction between number of turns, modality, and stimulus type, $\chi^2(1)=4.13$,

$p < .05$), which means that their communication involved overall more negotiation between players.

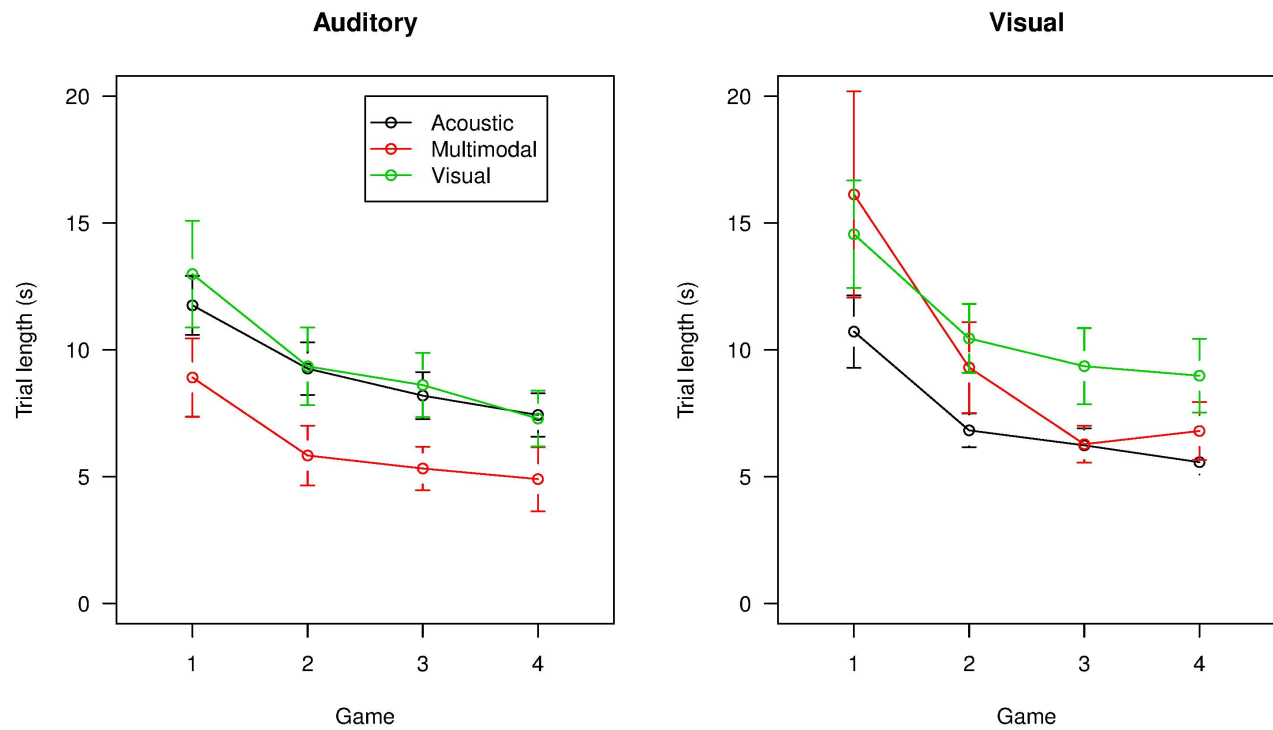


Figure 4. Trial length, as measured in seconds, according to stimulus type (auditory vs. visual), game (1, 2, 3, & 4), and experimental condition (acoustic vs. visual vs. multimodal).

5.3 Multimodal communication

In order to gain more insights about communication in the multimodal condition, players' communicative behavior in that condition was further analyzed in terms of the modalities in which it was produced (i.e., acoustic-only, visual-only, or both acoustic and visual). As such, it was possible to assess how multimodal signaling differed from acoustic and visual unimodal

signaling. Descriptions which included both acoustic and visual signals account for 43% of all descriptions produced by directors in the multimodal condition. Table 4 presents the descriptions issued in the multimodal condition in relation to the modality in which they were produced and the type of stimulus.

	Acoustic-only	Visual-only	Multimodal	Total
Auditory stimuli	22	68	238	328
Visual stimuli	0	314	29	343
Total	22	382	267	

Table 4. Director turns in the multimodal condition. Total number of director turns, according to stimulus type (auditory vs. visual) and turn type (acoustic vs. visual vs. multimodal).

As can be seen, participants in the multimodal condition employed the modalities at their disposal differently according to each stimulus type. In the case of visual stimuli, players relied heavily on visual signaling, producing no unimodal acoustic descriptions and only a few multimodal descriptions. In fact, multimodal descriptions account for only 8,5% of all descriptions of images, the remaining ones being visual-only in nature. In the case of auditory stimuli, although players did issue acoustic-only descriptions, the bulk of their descriptions was still visual in nature, varying between visual-only descriptions (unimodal) and visual-acoustic ones (multimodal), the latter accounting for 73% of all descriptions of sounds.

Within multimodal descriptions themselves, acoustic and visual components were used to different degrees. Figure 5 shows the proportional use of different types of signals in the

multimodal condition. The relative use of each modality is shown in terms of signals and their relative length within a turn.

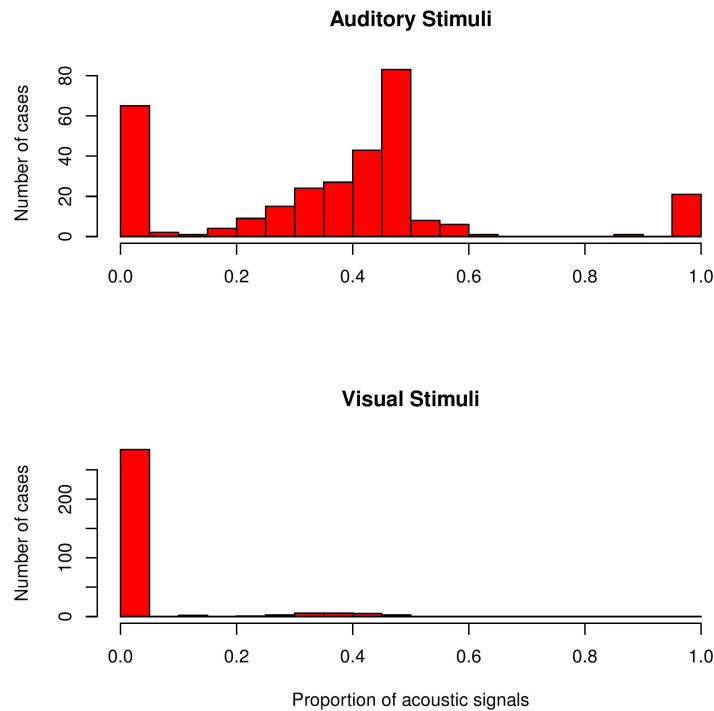


Figure 5. Distribution of signals produced by directors in the multimodal condition. The total time spent producing signals is shown in proportion to the time spent producing acoustic signals: a proportion of 1.0 stands for a turn composed only of unimodal acoustic signaling, whereas a proportion of 0.0 stands for a turn composed only of unimodal visual signaling. A proportion of 0.5 signifies a turn in which the same amount of time was dedicated to acoustic and visual signaling.

Since participants in the multimodal condition communicated both multimodally and unimodally, an in-depth analysis of communicative efficiency was conducted for trials in that

condition³. Figure 6 plots the efficiency with which players in the multimodal condition communicated in each signaling modality. The results show that players were as efficient when communicating unimodally, either vocally or visually, as when communicating multimodally. The trend was found for both types of stimuli, however, the graph on the right reflects the fact that players did not describe visual stimuli using acoustic signals only.

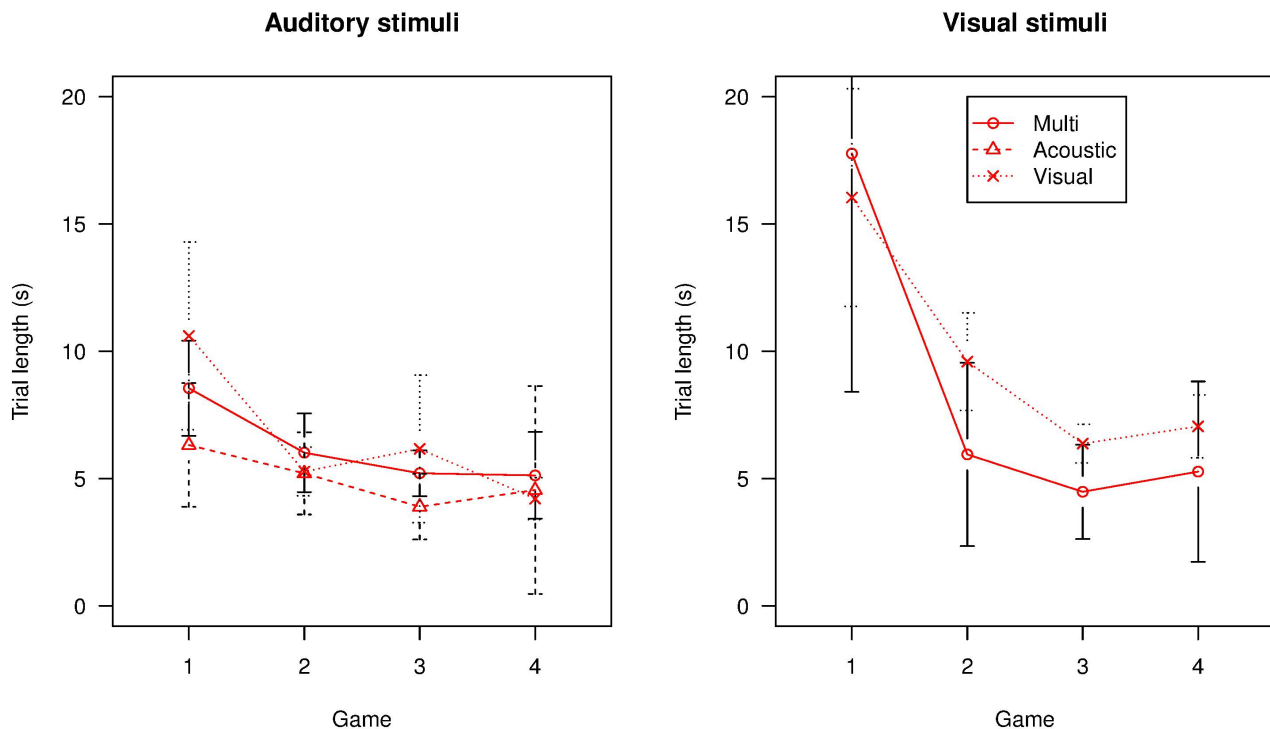


Figure 6. Trial length in the multimodal condition, as measured in seconds, according to stimulus type (auditory vs. visual), game (1, 2, 3, & 4), and signaling modality (acoustic vs. visual vs. multimodal).

In order to illustrate how players in the multimodal condition combined acoustic and visual signals, a step-by-step qualitative rendition of a trial containing multimodal turns is provided

³ The analysis was conducted using a linear mixed model predicting trial length for trials in the multimodal condition on the basis of signaling modality and stimulus type. No main effect of signaling modality was found.

below. Figure 7 illustrates how members of dyad 18 cooperate in specifying a reference for visual item #5. Director (Player 1) and matcher (Player 2) communicate extensively before the matcher finally makes a choice. Note should be taken that the matcher produces several vocalizations while communicating gesturally, these vocalizations drawing the attention of the director to the visual displays he is producing. Aside from pragmatic vocalizations, the matcher also produces a referential acoustic display (moment 7), which seems to complement the reference being established by its visible gestural counterpart.

P1: M1 {places fingers on table} [two hands] [index & middle finger] [coupled fingers]

P2: M2 {points at own shirt/ vocalizes} [index finger] [hə:/]

P1: M3 {taps chest} [one hand] [open hand]

P2: M4 {points at partner's shirt/ vocalizes} [index finger] [hə:/ + /ə:/]

P1: M4 {traces circle on table} [two hands] [index finger]

P2: M5 {points at partner's shirt/ vocalizes} [index finger] [hə:/]

M6 {places fingers on table/ vocalizes} [one hand] [all fingers] [fingers tips] [mu:ə:/]

M7 {places fingers on table/ vocalizes} [two hands] [index & middle finger]
[coupled fingers] [hə:/]

M8 {points at partner's shirt} [index finger]

P1: M9 {nods}

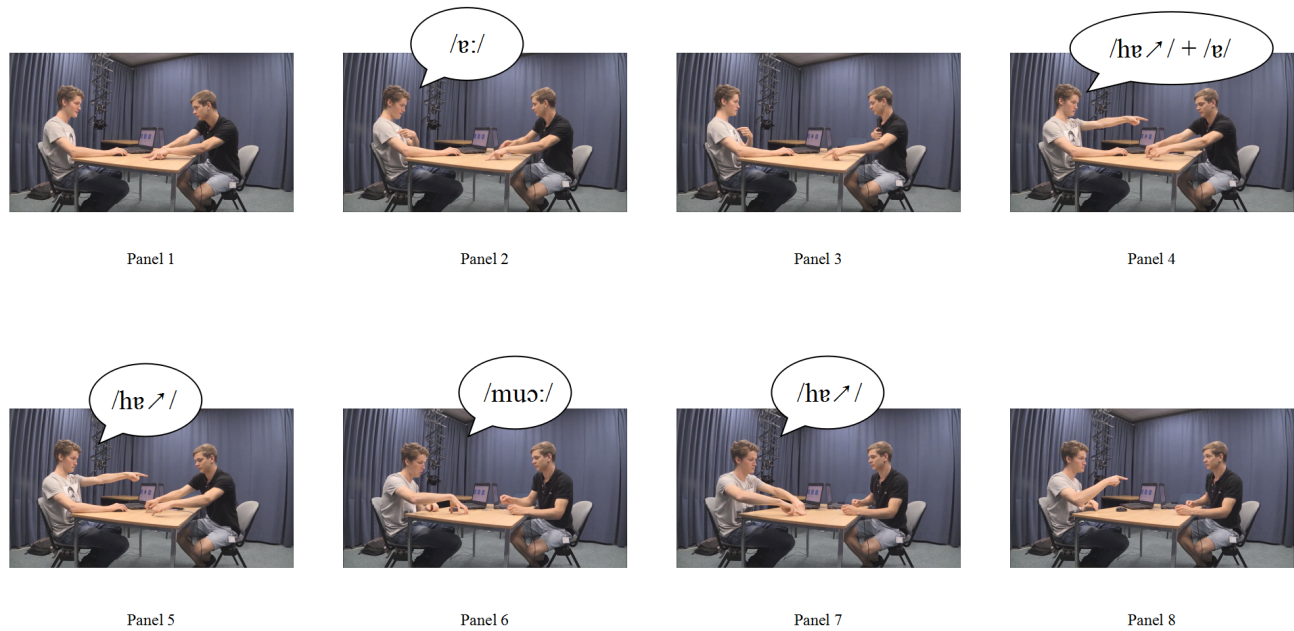


Figure 7. Step-by-step development of a trial in the multimodal condition. Each panel makes reference to a respective moment in interaction.

5.4 Extended communicative interaction

As explained in section 3.4, a basic trial in the experiment consisted of a director describing an item followed by a matcher making a choice. However, players could extend this basic interactive sequence whenever necessary, an extension meaning that members of a dyad addressed, and in principle resolved, some problem which was keeping the matcher from making their choice. These extensions of the basic game sequence impacted the efficiency with which players communicated, which is why it is relevant to consider the amount of extended interaction registered in each condition. Figure 8 shows the absolute number of turns taken by matchers,

these turns implying a follow up on a prior turn taken by a director and thus an extension of the basic trial.

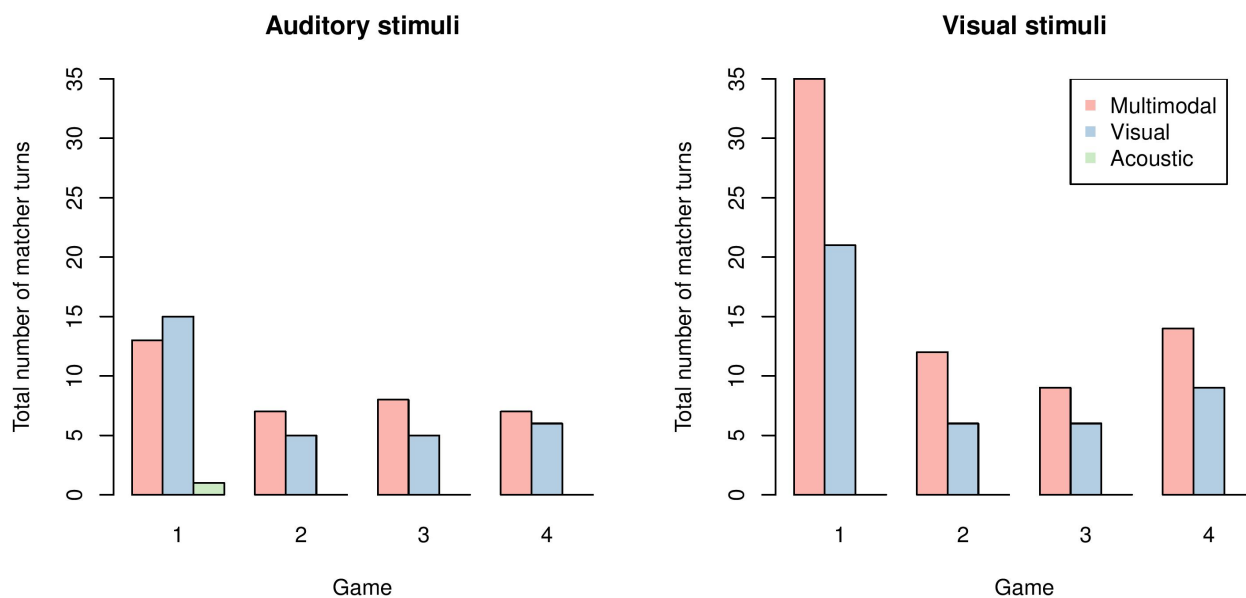


Figure 8. Absolute number of turns taken by matchers. Number of turns taken by matchers according to stimulus type (auditory vs. stimuli), game (1, 2, 3, & 4), and experimental condition (acoustic vs. visual vs. multimodal). For every turn taken by a matcher, a prior turn was taken by a director.

As indicated by the graphs, participants in the visual and multimodal conditions extended their interactions considerably more than their counterparts in the acoustic condition, participants in the multimodal condition doing so, overall, more than participants in the visual condition. Players extended their trials more when describing images than when describing sounds, a result which is suggestive of greater difficulty in describing visual stimuli but also, to some extent, greater willingness to negotiate descriptions of those stimuli in the first place.

6. Summary of the results

The main results show that participants in all conditions were equally accurate in communicating auditory stimuli, participants in the multimodal condition being overall the most efficient in doing so. As for visual stimuli, participants in the acoustic condition communicated less accurately and yet more efficiently than participants in both the visual and multimodal conditions. In regard to communication in the multimodal condition, it was found that players varied their signaling preferences according to the stimuli they were describing. To describe visual stimuli, they communicated mostly unimodally (91,5% of descriptions contained only visual signals), whereas to describe auditory stimuli they communicated mostly multimodally (73% of descriptions contained both acoustic and visual signals). All in all, descriptions issued in the multimodal condition were mostly visual in nature, either strictly visual in the case of descriptions of visual items, or multimodal but containing a larger visual component in the case of descriptions of auditory items.

The results also show that participants in all conditions communicated both more accurately and more efficiently as the experiment progressed. Moreover, the longer participants took in negotiating items, the more likely they were to make incorrect guesses, which suggests that interaction influenced players' accuracy in communicating items. Finally, the results reveal that, unlike participants in the acoustic condition, participants in the visual and multimodal conditions extended their trials regularly throughout the experiment.

6.1 Shortcomings and limitations of the analysis

In its current form, the analysis of efficiency depends on measuring the length of a whole trial, from the moment a director started communicating up until the moment their partner made a choice. However, given that trials are composed of turns, and turns composed of individual communicative acts, total trial efficiency fails to capture finer-grained dimensions such as the time elapsed between the end of a director's final turn and their partner's choice (i.e., time taken by matcher in making a guess), the length of each individual communicative act, or even the length of a director's first turn. These measures can be easily retrieved given total trial length, yet, even though combining different levels of measurement might provide more detailed information about players' performance, it also incurs complex results and convoluted data visualization.

7. Discussion

7.1 Communicating on the basis of spontaneous acoustic and visual signals

In the study here reported, participants performed a referential communication task interacting strictly on the basis of non-linguistic acoustic and visual behavior. Consisting of sounds and images, the stimuli used in the experiment allowed for both acoustically and visually motivated representations, given the respective use of the vocal-auditory and visuospatial modality in representing those stimuli. Thus, it was predicted that (1) participants communicating in a signaling modality which matched the motivated potential of the stimuli would perform better than those communicating in a modality which did not afford such motivatedness. However, given the findings of Fay et al. (2013, 2014), it was also predicted that (2) participants who could use gesture to communicate would nonetheless have an advantage over those who could not rely on gesture. Finally, it was predicted that (3) multimodality, that is, the combination of both acoustic and visual signaling, would grant communicators an inherent advantage over unimodal communication.

The results showed that throughout the experiment participants became both more accurate and more efficient in communicating only on the basis of acoustic and visual signals. Participants in all conditions were equally accurate in communicating sounds, those in the multimodal condition doing being more efficient than their counterparts in the acoustic and visual conditions. Multimodality seems to have had an advantage in terms of efficiency. As for accuracy, as

predicted by hypothesis (2), visible gesture did indeed provide participants in the visual and multimodal conditions with an inherently powerful means of expression, as indicated by the fact that they communicated as accurately as participants in the acoustic condition, who could produce perceptually motivated descriptions of auditory stimuli.

Unlike their counterparts in the visual condition, however, participants in the multimodal condition drew not only on visible gesture but also on the affordances of the vocal-auditory modality when communicating sounds. Indeed, their signaling flexibility enabled them to modulate their communicative behavior according to each situation, in some cases communicating acoustically, while in others communicating visually. Despite occasionally issuing unimodal descriptions, participants in the multimodal condition seem to have found a real advantage in communicating multimodally: by combining visual and acoustic signaling, they were able to describe sounds more efficiently than players who could communicate strictly unimodally (i.e., players in the acoustic and visual conditions), which confirms hypothesis (3).

Although participants in the acoustic condition could generate motivated descriptions of sounds, they were not as efficient as participants in the multimodal condition in communicating auditory items, which rejects hypothesis (1). Reduced efficiency in the acoustic condition might be explained by the circumstances under which participants communicated. On the one hand, players had no visual contact whatsoever, which deprived them of invaluable visual information, including eye gaze and facial expressions. These ostensive facial cues are known to act as online sources of feedback in face-to-face interaction (Chovil, 1991; Shockley, Richardson, & Dale,

2009), and so, possibly as a result of not being able to visually monitor one another, players in the acoustic condition might have changed their behavior in ways which affected their overall communicative efficiency. For instance, some players seem to have resorted to producing redundant descriptions of the stimuli, describing sounds by issuing reduplicated renditions of those sounds. Moreover, most players in the acoustic condition were simply unable to listen to a description and review the candidate options at the same time. Such a combination of factors could have resulted in a loss of efficiency compared to participants in the other conditions.

Regarding visual stimuli, indeed as predicted by hypothesis (1) and (2), participants who were able to communicate gesturally performed more accurately than those who could only communicate acoustically. Although there were no significant differences in accuracy between the visual and multimodal conditions, hypothesis (3) was neither confirmed nor rejected, since participants in the multimodal condition communicated almost exclusively gesturally when describing images. Communication was, therefore, unimodal in both the visual and the multimodal conditions.

As far as efficiency in communicating images is concerned, players in the acoustic condition were more efficient compared to players in the visual and multimodal conditions, which rejects all 3 predictions. Higher efficiency in the acoustic condition might be explained by the fact that acoustic descriptions of images seemed generally shorter than visual descriptions of those same images. In addition to that, visual stimuli elicited extensive negotiation between players in the visual and multimodal conditions. Indeed, as indicated by the analysis of communicative

interaction, participants in those conditions often extended their trials in the process of communicating visual items. Since this did not happen in the acoustic condition, even though players in the other two conditions might have been favored by the affordances of the visual modality, overall decrease in efficiency can be linked to the fact that they extended their trials.

7.3 Multimodality and its advantages

As explained in section 5.3, participants in the multimodal condition did not always communicate multimodally. Communication was multimodal for the most part when participants were describing auditory items, 73% of all descriptions of sounds being multimodal in nature. Otherwise, when describing visual items, and in fact for a considerable share of auditory items as well, communication in the multimodal condition operated mainly via the visual channel. What that means is that the displays produced by players were primarily visual, containing larger visual components in comparison to acoustic ones. Interestingly, there were no significant differences in efficiency when it came to communicating unimodally or multimodally in the multimodal condition, which suggests that participants in that condition resorted precisely to what was more appropriate in each scenario, whether that being visual signaling, acoustic signaling, or a combination of the two.

The findings of the study also point to the case that non-linguistic vocalizations might serve as optimal increments to visual-gestural signaling, acoustic signals having been used by participants not only to co-establish reference together with visible gesture but also to recruit interlocutors'

visual attention. Indeed, in the multimodal condition, acoustic signals seem to have served specialized purposes, vocalizations being used not only as motivated representations of auditory stimuli, as in the acoustic condition, but also as tools that helped communicators highlight non-referential dimensions related to the context in which their visual displays were being produced. Vocalizations seem to have been used, among other things, to contrast between competing descriptions and/or semantic features (see qualitative example, section 5.3), as well as to signal uncertainty and lack of understanding.

7.2 The effects of interactive communication

In most of the trials, matchers did not pursue any clarification as for what directors were describing, which is to say that, in most cases, matchers made their choices based only on the initial descriptions directors provided them with. However, in several cases players continued communicating until the indications given by the director corresponded to whatever expectations the matcher had in regard to what the target item was. An obvious corollary of such extended interaction between communicators is the decrease in their overall communicative efficiency, as shown by the association between the number of turns taken by players in a given trial and the total trial length.

Extended trial time was not only associated with reduced efficiency but also with a loss in accuracy. However, extended communication between players might have been ultimately beneficial, as it might have paved the way for increased form awareness and descriptive

systematicity. Indeed, by pursuing more information and requesting help from their peers, players might have sought to fine-tune their descriptions and thus to increase their chances of guessing items successfully. Additionally, players might have engaged in extended interaction whenever their descriptions deviated from certain expected formats or when the mapping between a description and the presented item seemed incorrect or unsatisfying. Evidence for why participants extended their trials could be gained from an in-depth analysis of their interactions, such an analysis lying beyond the scope of the present study. All in all, it should be noted that prolonged interaction did ultimately shape the course of the communicative encounters players took part in. Whether or not it facilitated the development of more refined communicative conventions, the fact that players spent time negotiating their descriptions meant that those descriptions were, to some extent or another, ingrained in their interactional history.

As explained in the previous section, the affordances of the vocal-auditory and the visuospatial modality seem to have been exploited by participants not only in signaling pure reference, but also in highlighting metareferential dimensions inherent to communication as a form of joint action. Indeed, communication was modulated not only by the affordances of a given modality, but also by players' reliance on ostensive signals which seem to have as pragmatic tools that supported referential communication. Arguably, whether participants were able to see each other or not was another factor which influenced their performance, even when players communicated solely via the acoustic channel. On the one hand, participants in the acoustic condition seem to have compensated for their lack of visual contact by reduplicating their descriptions of sounds. Participants in the visual and multimodal conditions, on the other hand, often extended their

trials when communicating, which we hypothesize to be connected to an increased disposition to interact given visual contact. Indeed, research has shown that not only is eye gaze linked to the initiation and maintenance of social interaction (Kendon, 1967; Goodwin, 1981; see also Rossano, 2013) but also that it embodies shared knowledge between communicators in face-to-face interaction (Shockley, Richardson, & Dale, 2009), which can help explicate the differences in the amount of interaction, and ultimately in performance, between participants in the acoustic condition and participants in the conditions in which visual contact was available.

7.4 Improvised multimodal signaling and the emergence of language

By investigating how people describe novel meanings to one another in the context of improvised, non-linguistic communication, the study here reported yields interesting clues about how the natural affordances of the human body might be used to bootstrap communication in the absence of verbal language. Indeed, the study helped map out the general representational affordances of both the vocal-auditory and the visuospatial modality. Previous empirical studies had provided evidence that visible gesture is a particularly expressive means of non-linguistic communication (Fay et al, 2013, 2014; Goldin-Meadow, So, Ozyurek, & Mylander, 2008; Schouwstra & de Swart, 2014). The studies by Fay et al. (2013, 2014) in particular had emphasized its inherent power, claiming that multimodality had no advantages over unimodal gestural communication.

The present study confirms the representational power of gesture in the context of improvised communication. However, it contradicts previous findings in showing that multimodal communication is, in certain cases, more powerful than gestural communication alone. Moreover, it demonstrates that, similar to gesture, the human voice is a powerful means in its own right, given its use in motivated acoustic representations (see Perlman, Dale, & Lupyan, 2015). In fact, the results suggest that sounds produced with the voice can to some extent increase the intrinsic power of gestural communication, allowing for more flexible and case-sensitive multimodal displays. Importantly, the results also point to the case that, notwithstanding its representational potential, acoustic signaling seems particularly suited to supporting visual signaling in improvised communication. Although the current analysis cannot support any claims, non-linguistic vocalizations seem to serve the function of drawing attention to visual displays, or otherwise of pragmatically marking their context of utterance.

The findings of the current study can be linked to theoretical proposals such as the one by Levinson & Holler (2014), who suggest that the multimodal system of communication modern humans possess emerged out of a stratified evolutionary process. As part of the incremental evolution which culminated in the gesture-plus-speech system in place today, they hypothesize the presence of an early capacity to produce ad hoc gestural displays, most likely on the basis of action sequences such as the ones produced by modern apes (Tomasello, 2008). Crucially, such gestural systems might have been enriched with simple vocalizations from a very early stage, vocalizations being used, first and foremost, to draw attention to whatever visual display one was producing. Given, then, a minimally powerful system of multimodal signaling, human

communication would be on the track to becoming more complex and expressive. Indeed, as suggested by Levinson & Holler, later adaptations would have ensured the development of vocal signaling into a much more powerful and complex medium, adding in flexibility and expressivity to the visual-gestural system that supported it. Ultimately, the two systems would have continued co-evolving, leading to the fully-integrated and dynamic multimodal system humans make use of today.

8. Conclusion

The present study explored how the natural affordances of the vocal-auditory and visuospatial modality can be exploited in the context of non-linguistic communication. Its findings extend previous literature in regard to the role of modality in language emergence. First, it was shown that gestural signaling is indeed a powerful means of improvised communication. Then, it was demonstrated that acoustic signaling has a representational potential of its own, that being especially evident in combination with visual signaling. Indeed, in contradiction to previous experimental literature, it was shown that visual-acoustic signaling is a flexible means of communication, in certain cases being more powerful than unimodal gestural communication. Finally, it has been shown that, despite the combination of acoustic and visual signaling, improvised multimodal behavior seems to depend more on gesture than on vocalizations.

Together with other visual modes of communication such as eye gaze and facial expressions, early visual-acoustic signaling might have launched human communication from its embodied

roots towards ever more complex and language-like forms of communication. Future research should address this issue more directly, exploring the role of improvised multimodal signaling not only in the emergence of referential communicative conventions but mainly in the development of interactive communication itself.

References

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child development*, 450-466.

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, 49(6), 1053-1076.

Arends, J., Muysken, P., & Smith, N. (Eds.). (1994). *Pidgins and creoles: An introduction* (Vol. 15). John Benjamins Publishing.

Aronoff, M., Meir, I., Padden, C., & Sandler, W. (2008). The roots of linguistic organization in a new language. *Interaction studies*, 9(1), 133.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Package 'lme4'.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse processes*, 15(4), 469-489.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental science*, 8(6), 535-543.

Bull, P. E. (2016). *Posture & Gesture* (Vol. 16). Elsevier.

Capirci, O., Iverson, J. M., Pizzuto, E., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child language*, 23(03), 645-673.

Chovil, N. (1991). Discourse-oriented facial displays in conversation. *Research on Language & Social Interaction*, 25(1-4), 163-194.

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in cognitive sciences*, 7(7), 300-307.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1-39.

De Ruiter, J. P., Noordzij, M. L., Newman-Norlund, S., Newman-Norlund, R., Hagoort, P., Levinson, S. C., & Toni, I. (2010). Exploring the cognitive infrastructure of communication. *Interaction Studies*, 11(1), 51-77.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, 19(10), 603-615.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6), 581-604.

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances* (Vol. 8). Cambridge University Press.

Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive science*, 37(7), 1356-1367.

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014). Creating a communication system from scratch: gesture beats vocalization hands down.

Flom, R. E., Lee, K. E., & Muir, D. E. (2007). *Gaze-following: Its development and significance*. Lawrence Erlbaum Associates Publishers.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in human neuroscience*, 5, 11.

Galantucci, B., Garrod, S., & Roberts, G. (2012). Experimental semiotics. *Language and Linguistics Compass*, 6(8), 477-493.

Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews*, 30(7), 949-960.

Givón, T. (1985). Iconicity, isomorphism and non-arbitrary coding in syntax. *Iconicity in syntax*, 187-219.

Givón, T. (1995). *Functionalism and grammar*. John Benjamins Publishing.

Goldin-Meadow, S., Mylander, C., & Franklin, A. (2007). How children make language out of gesture: Morphological structure in gesture systems developed by American and Chinese deaf children. *Cognitive Psychology*, 55(2), 87-135.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, 105(27), 9163-9168.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. Academic Press.

Goodwin, C. (2003). Pointing as situated practice. *Pointing: Where language, culture and cognition meet*, 217-241.

Guidetti, M. (2002). The emergence of pragmatics: forms and functions of conventional gestures in young French children. *First Language*, 22(3), 265-285.

Haiman, J. (1985). Natural syntax. Iconicity and erosion. *Cambridge Studies in Linguistics* London, (44), 1-285.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4), 596-615.

Holler, J., & Beattie, G. (2003a). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener?. *Gesture*, 3(2), 127-154.

Holler, J., & Beattie, G. (2003b). How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process?. *SEMIOTICA-LA HAYE THEN BERLIN-*, 146(1/4), 81-116.

IRVINE, E. (2014). WHAT ICONICITY CAN AND CANNOT DO FOR PROTO. In *The Evolution of Language: Proceedings of the 10th International Conference* (p. 122). World Scientific.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science*, 16(5), 367-371.

KANERO, J. (2014). The gesture theory of language origins: current issues and beyond. *The Past, Present and Future of Language Evolution Research*, 1.

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of memory and Language*, 40(4), 577-592.

Kelly, S. D., Özyürek, A., & Maris, E. (2009). Two sides of the same coin speech and gesture mutually interact to enhance comprehension. *Psychological Science*.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, 26, 22-63.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kendon, A. (2009). Language's matrix. *Gesture*, 9(3), 355.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Phil. Trans. R. Soc. B*, 369(1651), 20130302.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Phil. Trans. R. Soc. B*, 369(1651).

Liebal, K., Waller, B. M., Slocombe, K. E., & Burrows, A. M. (2013). *Primate Communication: a multimodal approach*. Cambridge University Press.

Lubbers M., Torreira F. (2014). Pympi Python Library Version 1.1.

McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of pragmatics*, 32(7), 855-878.

McNeill, D. (2012). *How language began: Gesture and speech in human evolution*. Cambridge University Press.

Mondada, L. (2007). Multimodal resources for turn-taking pointing and the emergence of possible next speakers. *Discourse studies*, 9(2), 194-225.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility head movement improves auditory speech perception. *Psychological science*, 15(2), 133-137.

Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, 96(3), B101-B113.

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of cognitive neuroscience*, 19(4), 605-616.

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Open Science*, 2(8), 150152.

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in psychology*, 1, 227.

Roberge, P. T. (2008). The creation of pidgins as a possible window on language evolution. LOT Occasional Series, 10, 101-138.

Rossano, F. (2013). Gaze in conversation. In T., Stivers, J., Sidnell (Eds.), *The handbook of conversation analysis* (pp. 308–29). Chichester, UK: Wiley-Blackwell.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7), 2661-2665.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, 131(3), 431-436.

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. (2009). Signalling signalhood and the emergence of communication. *Cognition*, 113(2), 226-233.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, 305(5691), 1779-1782.

Seyfeddinipur, M., & Gullberg, M. (2014). *From gesture in conversation to visible action as utterance*. Benjamins.

Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, 1(2), 305-319.

Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford University Press.

Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: the need for multimodality in primate communication research. *Animal Behaviour*, 81(5), 919-924.

Steels, L. (1999). *The talking heads experiment: Words and Meanings*. VUB Artificial Intelligence Laboratory.

Stivers, T., & Sidnell, J. (2005). Introduction: multimodal interaction. *Semiotica*, 2005(156), 1-20.

Tagliatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PloS one*, 6(4).

Tomasello, M. (2008). *Origins of human communication*. MIT press.

Verhoef, T., Roberts, S. G., & Dingemanse, M. (2015). Emergence of systematic iconicity: transmission, interaction and analogy. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2481-2486). Cognitive Science Society.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: differentiation of multimodal information. *Psychological bulletin*, 121(3), 437.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006, May). Elan: a professional framework for multimodality research. In *Proceedings of LREC* (Vol. 2006, p. 5th).

Yule, G. (1997). *Referential communication tasks*. Routledge.

Modality effects in a signalling game: Accuracy

Intro

This script uses data compiled by *analyseData.R*.

Load libraries

```
library(lme4)
library(sjPlot)
library(ggplot2)
library(lattice)
library(influence.ME)
library(party)
```

Load data

```
d = read.csv("../data/FinalSignalData.csv")
```

Work out number of turns in each trial.

```
# Number of turns in each trial
numTurns = tapply(d$turnString, d$trialString,
                  function(X){length(unique(X))})
d$numberOfTurns = numTurns[d$trialString]

matcherResponds = tapply(d$turnType, d$trialString,
                         function(X){"T2" %in% X})
d$matcherResponds = matcherResponds[d$trialString]
```

Variable for length of first T1

```
T1L = tapply(d[d$turnType=="T1",]$turnLength,
            d[d$turnType=="T1",]$trialString, head, n=1)
d$T1Length = T1L[d$trialString]
d$T1Length[is.na(d$T1Length)] = mean(d$T1Length, na.rm=T)
d$T1Length.log = log(d$T1Length)
d$T1Length.log = d$T1Length.log - mean(d$T1Length.log)
```

We don't need info on every signal in each turn, just the trial time. Keep only 1st signal in each trial.

```
d = d[!duplicated(d$trialString),]
```

Descriptive stats

Here is a graph showing the distribution of accuracy by conditions:

Make a variable to represent proportion of games played:

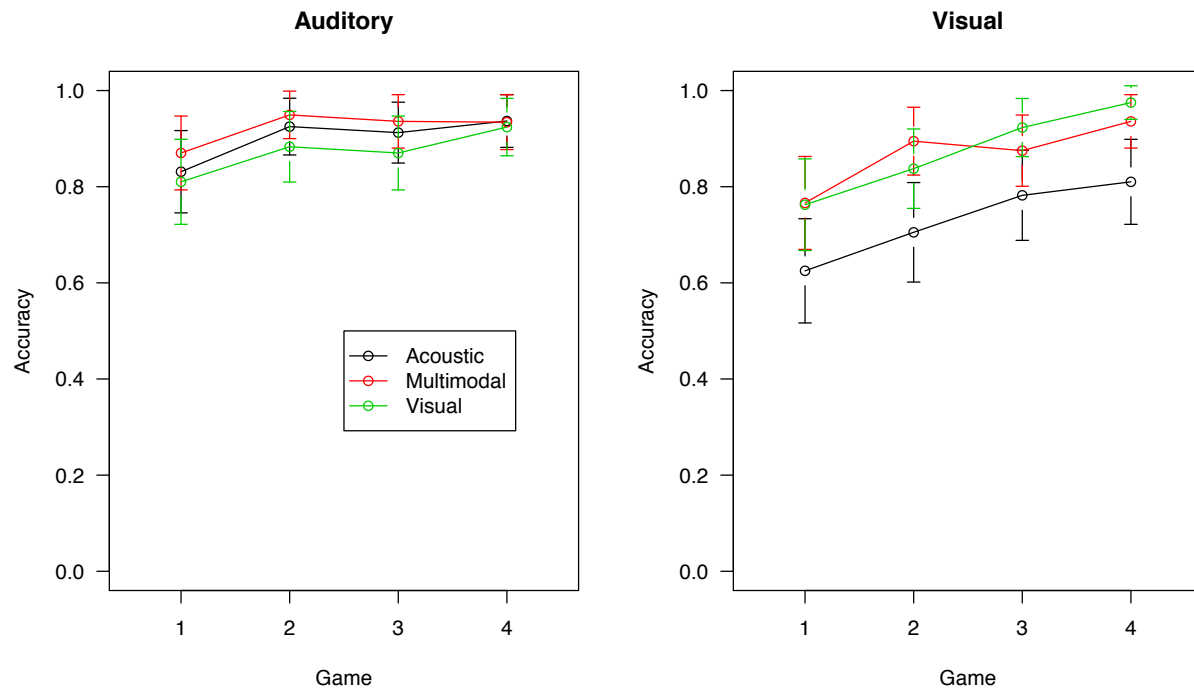


Figure 1: The efficiency of trials in different conditions

```
# Make a variable that represents the number of trials played
d$trialTotal = d$trial + (d$game * (max(d$trial)+1))
# Convert to proportion of games played, so that estimates reflect change per game.
d$trialTotal = d$trialTotal / 16
# Center the trialTotal variable so intercept reflects after the first game
d$trialTotal = d$trialTotal
```

Make a variable for which stimuli the players experienced first.

```
firstBlock = tapply(as.character(d$condition), d$dyadNumber, head, n=1)
d$firstBlock = as.factor(firstBlock[match(d$dyadNumber, names(firstBlock))])
```

Variable to indicate whether T1 is multimodal.

```
turnD = read.csv("../data/Final_Turn_data.csv")
turnD = turnD[turnD$turnType=="T1",]
turnD = turnD[turnD$role == "Director",]
d$multimodal = turnD[match(d$trialString, turnD$trialString),]$turnModalityType == "multi"
d$multimodal[is.na(d$multimodal)] = F
```

Make a variable to represent proportion of games played:

```
# Make a variable that represents the number of trials played
d$trialTotal = d$trial + (d$game * (max(d$trial)+1))
# Convert to proportion of games played, so that estimates reflect change per game.
d$trialTotal = d$trialTotal / 16
# Center the trialTotal variable so intercept reflects after the first game
d$trialTotal = d$trialTotal - 2
```

Transformed trial time.

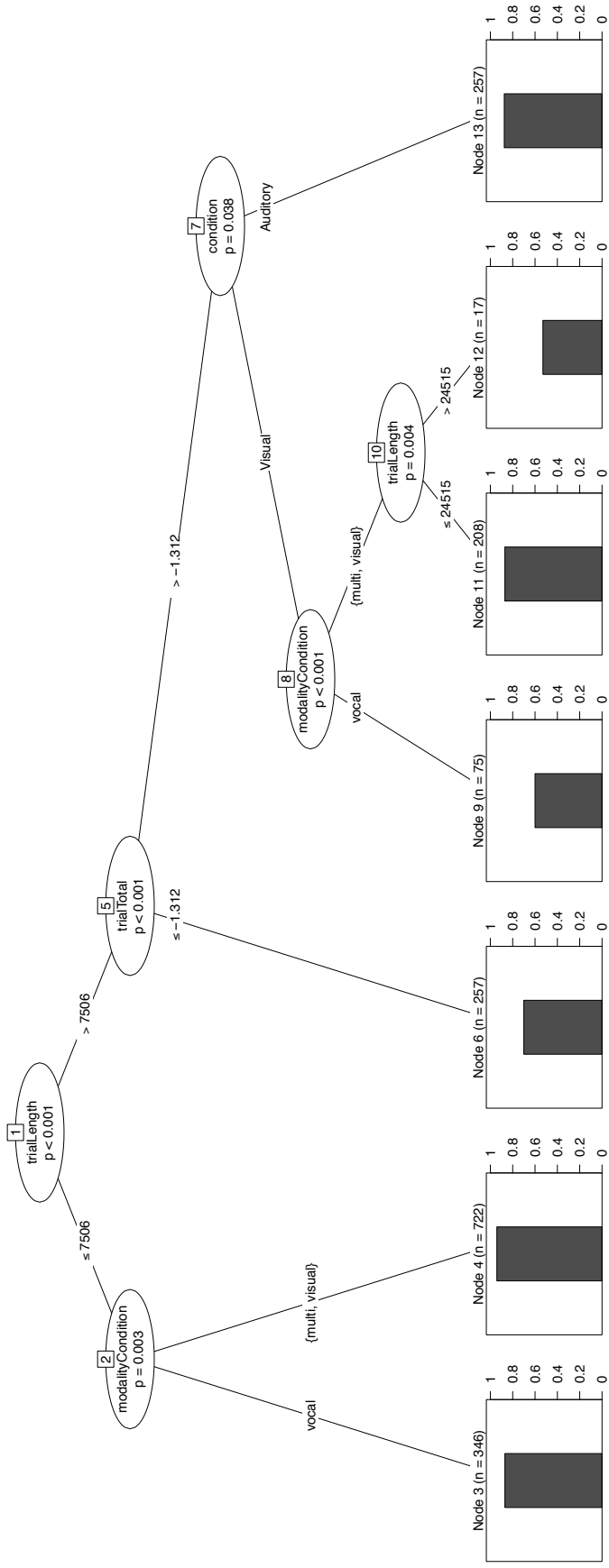
```
d$trialLength.log = log(d$trialLength)
meanLogTrialLength = mean(d$trialLength.log)
d$trialLength.log = d$trialLength.log - meanLogTrialLength
```

Get an idea of the structure of the data from a binary tree:

```
cx = ctree(correct ~ modalityCondition + condition +
            trialTotal +
            trialLength +
            numberOfTurns +
            matcherResponds +
            T1Length +
            multimodal+
            firstBlock,
            data=d)
```



```
plot(cx, terminal_panel=node_barplot(cx))
```



Mixed models

There are ceiling effects in the data, which reduces variance and makes model convergence difficult. Experimentation revealed that random effects other than random intercepts for dyad and item lead to non-convergence.

The final models do not converge within standard tolerances, but the convergence is acceptable.

No fixed effects

```
gc = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun=50000))

m0 = glmer(correct ~ 1 +
            (1 | dyadNumber) +
            (1 | itemId) ,
            data=d, family=binomial,
            control = gc)

game = glmer(correct ~ 1 +
              trialTotal +
              (1 | dyadNumber) +
              (1 | itemId) ,
              data=d, family=binomial,
              control = gc)

trialL = glmer(correct ~ 1 +
               trialTotal +
               trialLength.log +
               (1 | dyadNumber) +
               (1 | itemId) ,
               data=d, family=binomial,
               control = gc)

t1L = glmer(correct ~ 1 +
             trialTotal +
             trialLength.log +
             T1Length.log +
             (1 | dyadNumber) +
             (1 | itemId) ,
             data=d, family=binomial,
             control = gc)

multi = glmer(correct ~ 1 +
               trialTotal +
               trialLength.log +
               T1Length.log +
               multimodal +
               (1 | dyadNumber) +
               (1 | itemId) ,
               data=d, family=binomial,
               control = gc)

mod = glmer(correct ~ 1 + modalityCondition +
             trialTotal +
             trialLength.log +
             T1Length.log +
```

```

        multimodal+
        (1 |dyadNumber) +
        (1 |itemId) ,
data=d, family=binomial,
control = gc)

con = glmer(correct ~ 1 + modalityCondition + condition +
        trialTotal +
        trialLength.log +
        T1Length.log +
        multimodal+
        (1 |dyadNumber) +
        (1 |itemId) ,
data=d, family=binomial,
control = gc)

modXcon = glmer(correct ~ 1 + modalityCondition * condition +
        trialTotal +
        trialLength.log +
        T1Length.log +
        multimodal+
        (1 |dyadNumber) +
        (1 |itemId) ,
data=d, family=binomial,
control = gc)

trialLXmo = glmer(correct ~ 1 + modalityCondition * condition +
        trialTotal +
        trialLength.log * modalityCondition+
        T1Length.log +
        multimodal+
        trialLength.log +
        (1 |dyadNumber) +
        (1 |itemId) ,
data=d, family=binomial,
control = gc)

t1LXmo = glmer(correct ~ 1 + modalityCondition * condition +
        trialTotal +
        trialLength.log * modalityCondition+
        T1Length.log *modalityCondition +
        multimodal+
        trialLength.log +
        (1 |dyadNumber) +
        (1 |itemId) ,
data=d, family=binomial,
control = gc)

block = glmer(correct ~ 1 + modalityCondition * condition +
        trialTotal +
        trialLength.log * modalityCondition+
        T1Length.log *modalityCondition +

```

```
multimodal+  
trialLength.log +  
firstBlock +  
(1 |dyadNumber) +  
(1 |itemId) ,  
data=d, family=binomial,  
control = gc)
```

Results

Compare the fit of the models:

```
modelComparison = anova(m0,mod,con,modXcon,
                        game, trialL,trialLXmo,
                        t1L, t1LXmo,
                        multi, block)

modelComparison

## Data: d
## Models:
## m0: correct ~ 1 + (1 | dyadNumber) + (1 | itemId)
## game: correct ~ 1 + trialTotal + (1 | dyadNumber) + (1 | itemId)
## trialL: correct ~ 1 + trialTotal + trialLength.log + (1 | dyadNumber) +
## trialL:      (1 | itemId)
## t1L: correct ~ 1 + trialTotal + trialLength.log + T1Length.log + (1 |
## t1L:      dyadNumber) + (1 | itemId)
## multi: correct ~ 1 + trialTotal + trialLength.log + T1Length.log + multimodal +
## multi:      (1 | dyadNumber) + (1 | itemId)
## mod: correct ~ 1 + modalityCondition + trialTotal + trialLength.log +
## mod:      T1Length.log + multimodal + (1 | dyadNumber) + (1 | itemId)
## con: correct ~ 1 + modalityCondition + condition + trialTotal + trialLength.log +
## con:      T1Length.log + multimodal + (1 | dyadNumber) + (1 | itemId)
## modXcon: correct ~ 1 + modalityCondition * condition + trialTotal + trialLength.log +
## modXcon:      T1Length.log + multimodal + (1 | dyadNumber) + (1 | itemId)
## trialLXmo: correct ~ 1 + modalityCondition * condition + trialTotal + trialLength.log *
## trialLXmo:      modalityCondition + T1Length.log + multimodal + trialLength.log +
## trialLXmo:      (1 | dyadNumber) + (1 | itemId)
## t1LXmo: correct ~ 1 + modalityCondition * condition + trialTotal + trialLength.log *
## t1LXmo:      modalityCondition + T1Length.log * modalityCondition + multimodal +
## t1LXmo:      trialLength.log + (1 | dyadNumber) + (1 | itemId)
## block: correct ~ 1 + modalityCondition * condition + trialTotal + trialLength.log *
## block:      modalityCondition + T1Length.log * modalityCondition + multimodal +
## block:      trialLength.log + firstBlock + (1 | dyadNumber) + (1 | itemId)
##      Df    AIC    BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
## m0      3 1405.1 1421.8 -699.56   1399.1
## game     4 1355.9 1378.0 -673.95   1347.9 51.2377      1 8.183e-13 ***
## trialL   5 1327.1 1354.8 -658.54   1317.1 30.8054      1 2.852e-08 ***
## t1L      6 1329.0 1362.2 -658.48   1317.0  0.1207      1  0.72831
## multi    7 1331.0 1369.8 -658.48   1317.0  0.0004      1  0.98419
## mod      9 1329.2 1379.0 -655.59   1311.2  5.7777      2  0.05564 .
## con     10 1328.8 1384.2 -654.41   1308.8  2.3686      1  0.12380
## modXcon  12 1306.1 1372.6 -641.06   1282.1 26.7068      2 1.587e-06 ***
## trialLXmo 14 1306.8 1384.4 -639.41   1278.8  3.2903      2  0.19298
## t1LXmo   16 1309.6 1398.2 -638.78   1277.6  1.2646      2  0.53138
## block    17 1311.5 1405.7 -638.77   1277.5  0.0096      1  0.92178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pick final model for estimates:

```
finalModel = block
```

Model estimates:

```
summary(finalModel)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula:
## correct ~ 1 + modalityCondition * condition + trialTotal + trialLength.log *
## modalityCondition + T1Length.log * modalityCondition + multimodal +
## trialLength.log + firstBlock + (1 | dyadNumber) + (1 | itemId)
## Data: d
## Control: gc
##
##      AIC      BIC   logLik deviance df.resid
##  1311.5   1405.7   -638.8   1277.5     1865
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -8.2077  0.1304  0.2469  0.4077  1.5545
##
## Random effects:
## Groups      Name                Variance Std.Dev.
## itemId      (Intercept)  0.8636     0.9293
## dyadNumber   (Intercept)  0.1678     0.4096
## Number of obs: 1882, groups: itemId, 16; dyadNumber, 15
##
## Fixed effects:
##                                     Estimate Std. Error z value
## (Intercept)                        2.70441    0.57476   4.705
## modalityConditionvisual             -0.08443    0.51424  -0.164
## modalityConditionvocal              0.16822    0.52890   0.318
## conditionVisual                    -0.20518    0.64920  -0.316
## trialTotal                         0.28021    0.07059   3.969
## trialLength.log                    -0.76613    0.25435  -3.012
## T1Length.log                      -0.25848    0.25398  -1.018
## multimodalTRUE                     0.09001    0.44758   0.201
## firstBlockVisual                   -0.02671    0.27128  -0.098
## modalityConditionvisual:conditionVisual 0.42941    0.53255   0.806
## modalityConditionvocal:conditionVisual -1.41920    0.51275  -2.768
## modalityConditionvisual:trialLength.log -0.38138    0.37631  -1.013
## modalityConditionvocal:trialLength.log  0.13783    0.35785   0.385
## modalityConditionvisual:T1Length.log   0.08486    0.34497   0.246
## modalityConditionvocal:T1Length.log    0.30451    0.29834   1.021
##                                     Pr(>|z|)
## (Intercept)                        2.53e-06 ***
## modalityConditionvisual             0.86959
## modalityConditionvocal              0.75043
## conditionVisual                     0.75196
## trialTotal                         7.21e-05 ***
## trialLength.log                     0.00259 **
## T1Length.log                       0.30881
## multimodalTRUE                      0.84062
## firstBlockVisual                    0.92158
## modalityConditionvisual:conditionVisual 0.42005
## modalityConditionvocal:conditionVisual 0.00564 **
```

```
## modalityConditionvisual:trialLength.log 0.31084
## modalityConditionvocal:trialLength.log 0.70011
## modalityConditionvisual:T1Length.log 0.80570
## modalityConditionvocal:T1Length.log 0.30740
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 15 > 12.
## Use print(x, correlation=TRUE) or
##   vcov(x)      if you need it
```

Plot the fixed effects

Relabel the effects:

```
feLabels = matrix(c(
  "(Intercept)"           , "Intercept"           , NA,
  "modalityConditionvisual" , "Visual modality" , "mod",
  "modalityConditionvocal"  , "Acoustic modality" , "mod",
  "conditionVisual"        , "Visual stimuli" , "con",
  "trialTotal"             , "Game" , "game",
  "modalityConditionvisual:conditionVisual" , "Visual modality:Visual stimuli" , "modXcon",
  "modalityConditionvocal:conditionVisual" , "Acoustic modality:Visual stimuli" , "modXcon",
  "firstBlockVisual" , "Visual stims first" , "block",
  "trialLength.log" , "Trial length" , "trialL",
  "modalityConditionvisual:trialLength.log" , "Visual modality:Trial length" , 'trialLXmo',
  "modalityConditionvocal:trialLength.log" , "Acoustic modality:Trial length" , 'trialLXmo',
  "multimodalTRUE" , "Multimodal T1" , "multi",
  "trialLength.log" , 'Trial Length' , 'trialL',
  "T1Length.log" , "T1 length" , "t1L",
  "modalityConditionvisual:T1Length.log" , "T1 length:Visual modality" , "t1LXmo",
  "modalityConditionvocal:T1Length.log" , "T1 length:Acoustic modality" , "t1LXmo"

), ncol=3, byrow = T)
feLabels1 = as.vector(feLabels[match(names(fixef(finalModel)), feLabels[,1]),1])
feLabels2 = as.vector(feLabels[match(names(fixef(finalModel)), feLabels[,1]),2])
feModel = as.vector(feLabels[match(names(fixef(finalModel)), feLabels[,1]),3])
```

Plot the strength of the fixed effects:

```
sjp.glmer(finalModel, 'fe',
  show.intercept = T,
  sort.est=NULL,
  axis.labels = feLabels2[2:length(feLabels2)],
  axis.title="Odds of correct selection",
  geom.colors = c(1,1),
  show.values = F,
  show.p = F,
  fade.ns = T,
  string.interc="Intercept")
```

```
## Warning: Deprecated, use tibble::rownames_to_column() instead.
```

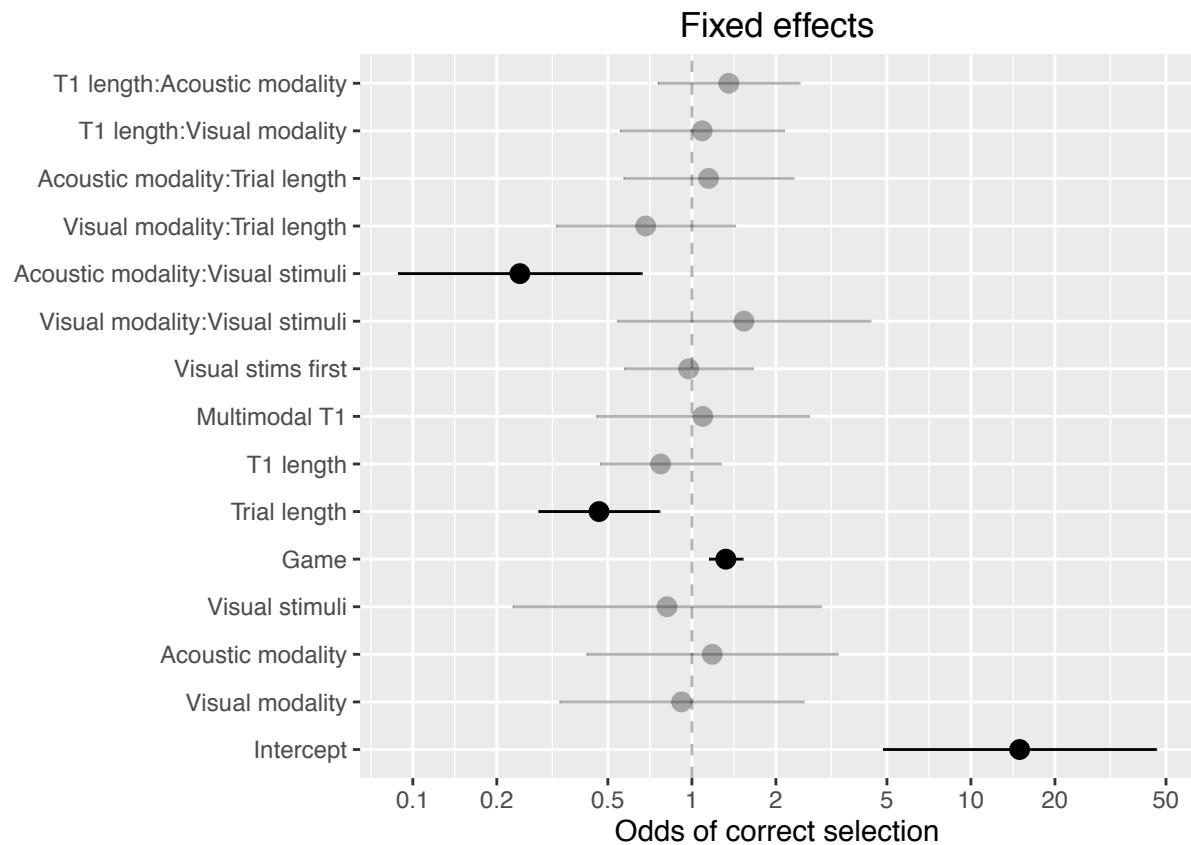
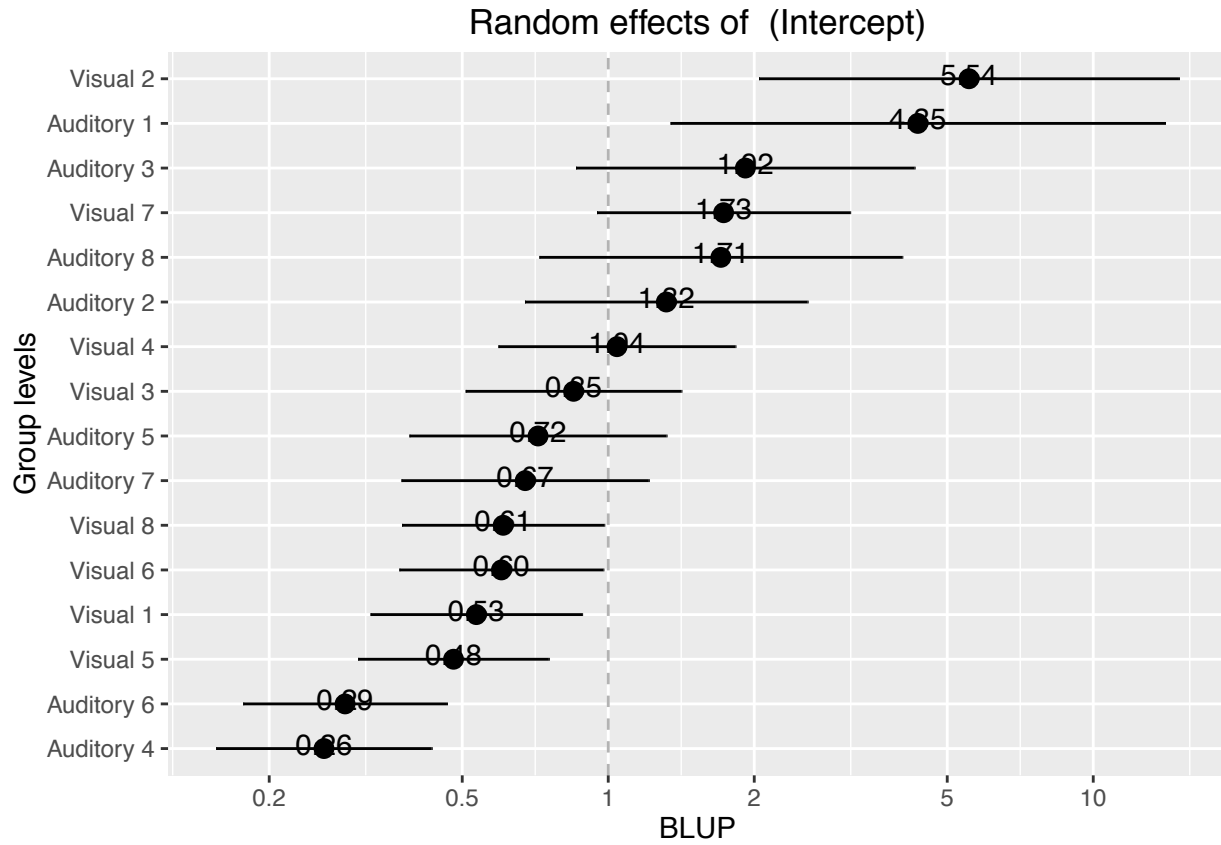


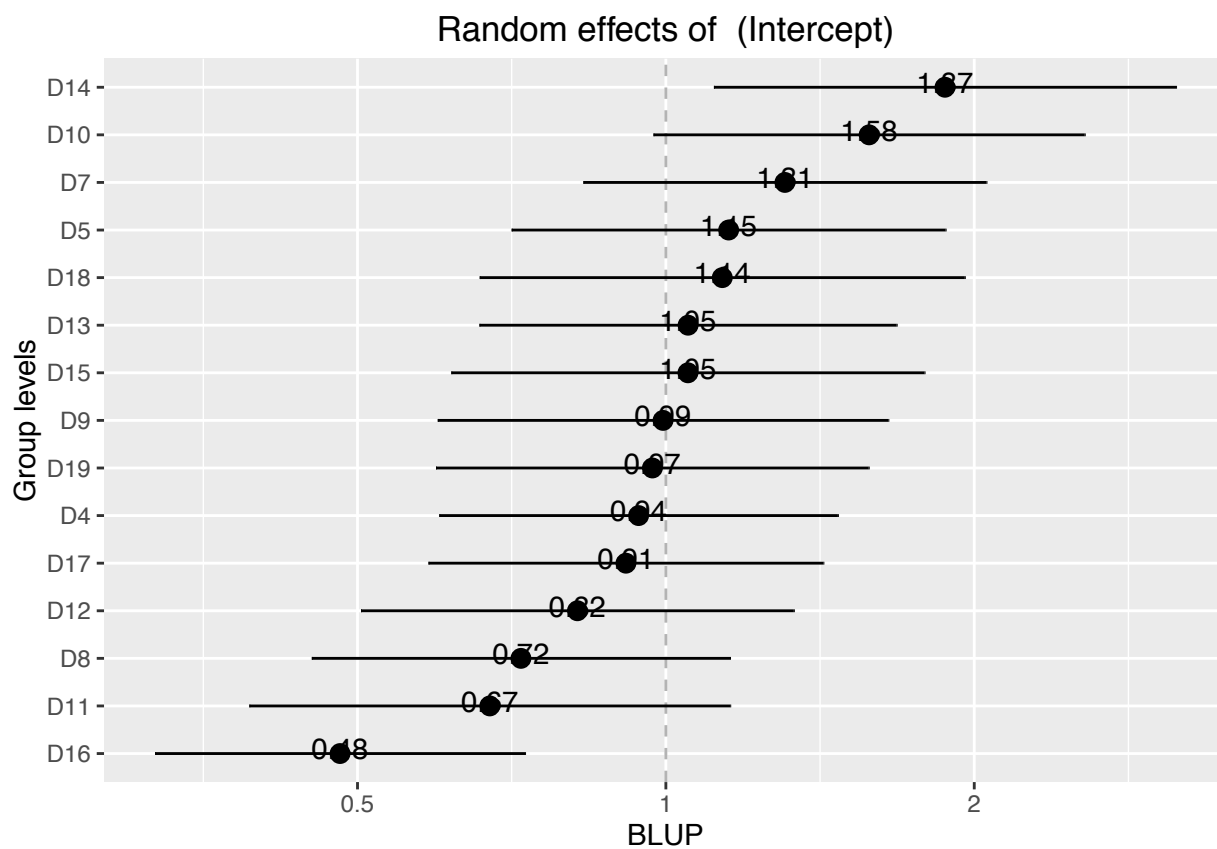
Table of results

```
x = as.data.frame(summary(finalModel)$coef)
mc = as.data.frame(modelComparison)
write.csv(cbind(x,mc[feModel,]), "../results/tables/Accuracy_FixedEffects.csv")
```


Random effects

```
sjp.glmer(finalModel, 're', sort.est = "sort.all",  
  facet.grid = F,  
  geom.colors = c(1,1))
```

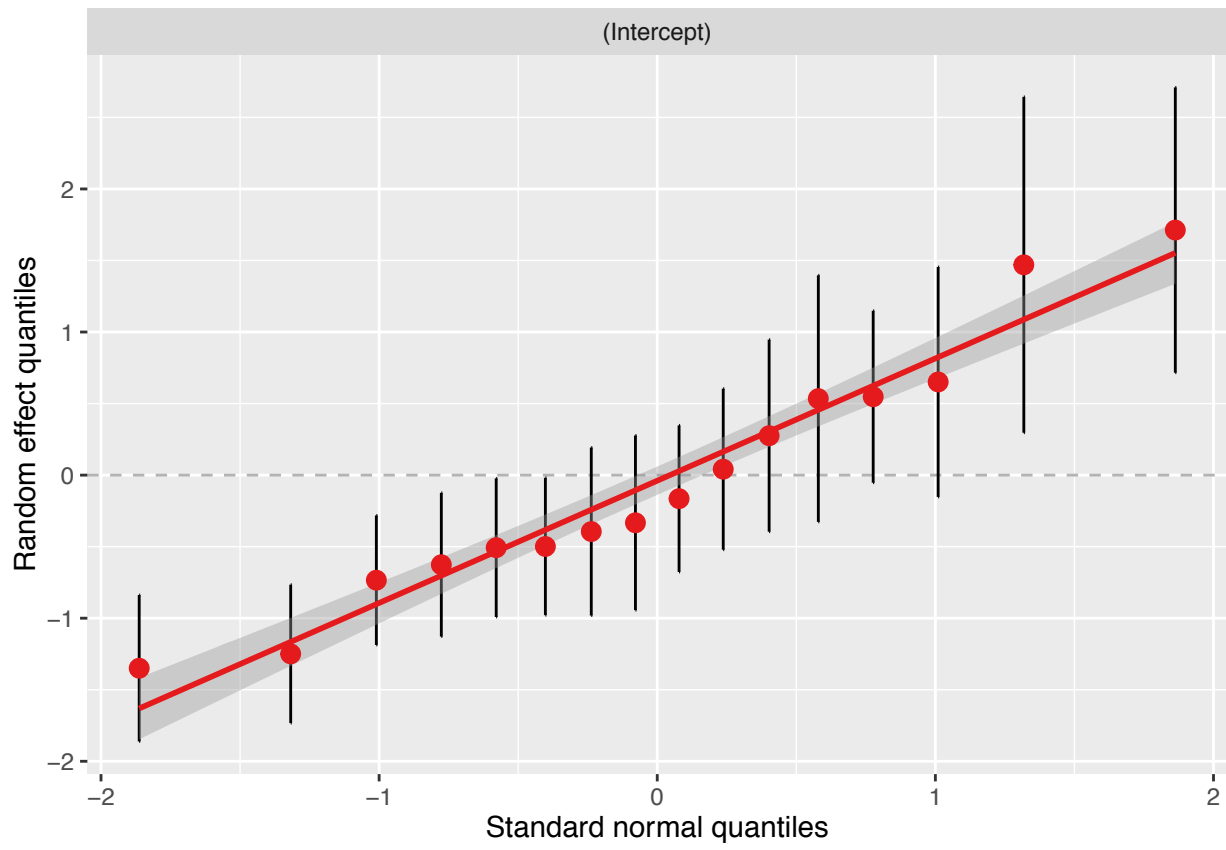




qq-plots of random effects

```
sjp.glmer(finalModel, type = "re.qq")
```

Testing for normal distribution. Dots should be plotted along the line.



Modality effects in a signalling game

Intro

This script uses data compiled by *analyseData.R*.

Load libraries

```
library(lme4)
library(sjPlot)
library(ggplot2)
library(lattice)
library(influence.ME)
```

Load data

```
d = read.csv("../data/FinalSignalData.csv")
```

Work out number of turns in each trial.

```
# Number of turns in each trial
numTurns = tapply(d$turnString, d$trialString,
                  function(X){length(unique(X))})
d$numOfTurns = numTurns[d$trialString]
```

Variable for length of first T1

```
T1L = tapply(d[d$turnType=="T1",]$turnLength,
             d[d$turnType=="T1",]$trialString, head, n=1)
d$T1Length = T1L[d$trialString]
d$T1Length[is.na(d$T1Length)] = mean(d$T1Length, na.rm=T)
d$T1Length.log = log(d$T1Length)
d$T1Length.log = d$T1Length.log - mean(d$T1Length.log)
```

We don't need info on every signal in each turn, just the trial time. Keep only 1st signal in each trial.

```
d = d[!duplicated(d$trialString),]
```

Descriptive stats

Here is a graph showing the distribution of trial lengths by conditions:

The distribution of trial times is very skewed:

```
hist(d$trialLength)
```

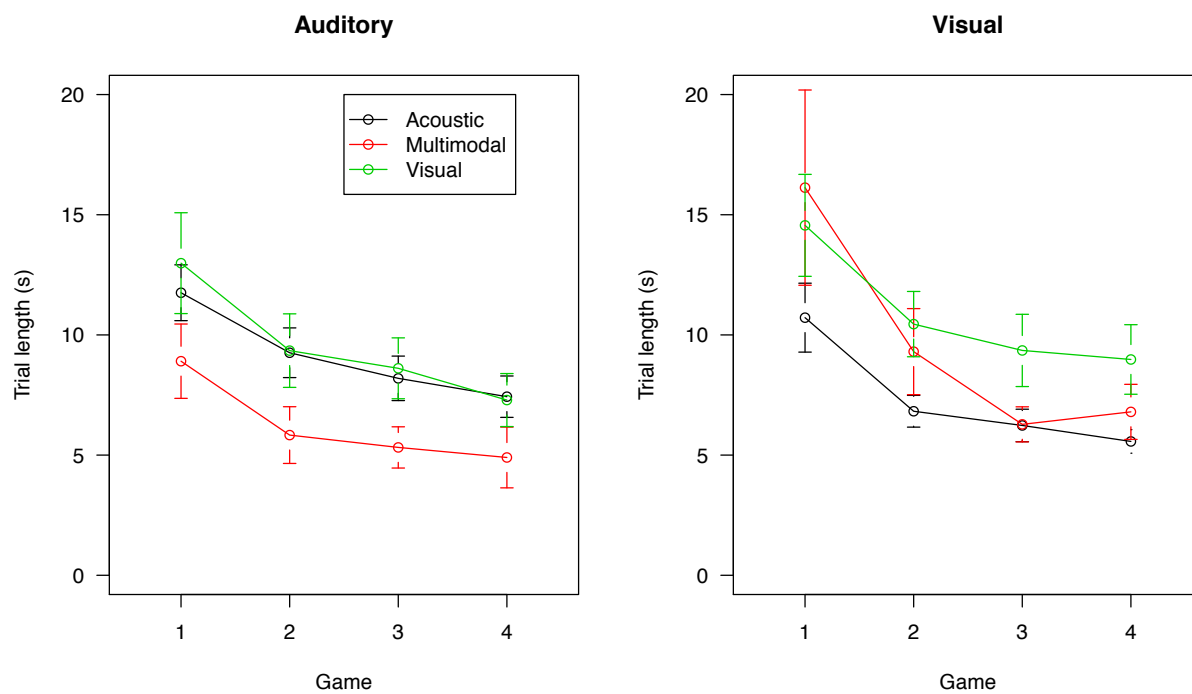
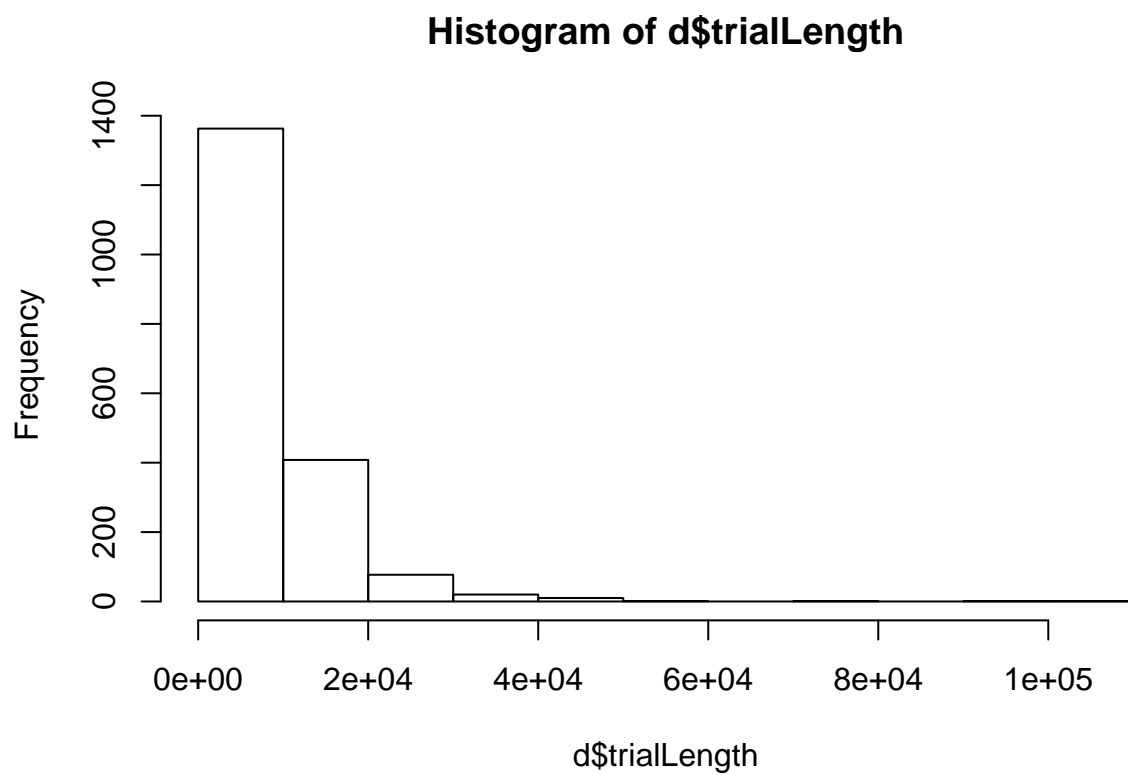
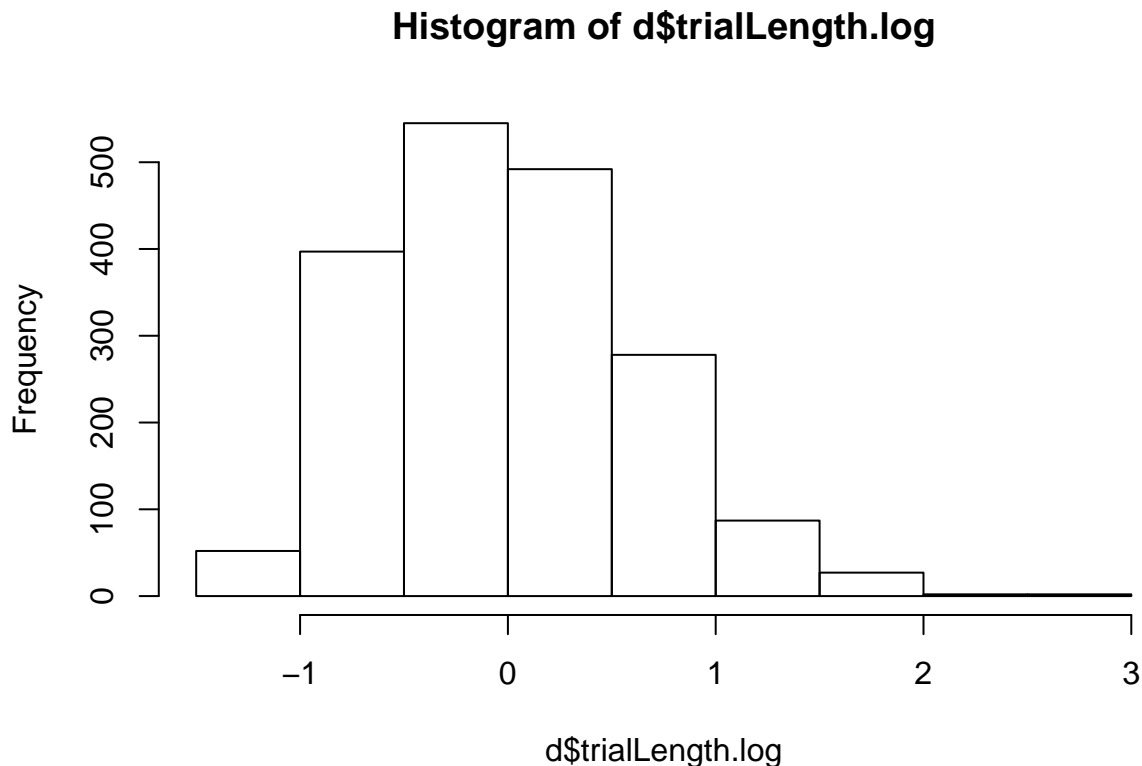


Figure 1: The efficiency of trials in different conditions



So we transform it using a log transform, then center the data.

```
d$trialLength.log = log(d$trialLength)
meanLogTrialLength = mean(d$trialLength.log)
d$trialLength.log = d$trialLength.log - meanLogTrialLength
hist(d$trialLength.log)
```



Make a variable to represent proportion of games played:

```
# Make a variable that represents the number of trials played
d$trialTotal = d$trial + (d$game * (max(d$trial)+1))
# Convert to proportion of games played, so that estimates reflect change per game.
d$trialTotal = d$trialTotal / 16
# Center the trialTotal variable so intercept reflects after the first game
d$trialTotal = d$trialTotal - 2
```

Make a variable for which stimuli the players experienced first.

```
firstBlock = tapply(as.character(d$condition), d$dyadNumber, head, n=1)
d$firstBlock = as.factor(firstBlock[match(d$dyadNumber, names(firstBlock))])
```

Reorder some levels so that the intercept reflects the most frequent condition.

```
d$incorrect = !d$correct
```

Variable for whether T1 was a multimodal signal.

```
turnD = read.csv("../data/Final_Turn_data.csv")
turnD = turnD[turnD$turnType=="T1",]
turnD = turnD[turnD$role == "Director",]
d$multimodal = turnD[match(d$trialString, turnD$trialString),]$turnModalityType == "multi"
d$multimodal[is.na(d$multimodal)] = F
```

Mixed models

Make a series of models with random effects for dyad, director (nested within dyad) and item.

Not all random slopes are appropriate. For example, items are used in only one stimulus condition, so a random slope for condition by item is not appropriate. Similarly, each dyad only plays in one modality condition.

It is reasonable to have a random slope for trial by dyad, but this caused unreliable model convergence, so is not included.

The final random slopes were for condition and incorrectness by dyad/player, and modality condition by item.

```
# No fixed effects
m0 = lmer(trialLength.log ~ 1 +
          (1 + condition + incorrect | dyadNumber/playerId) +
          (1 + modalityCondition | itemId),
          data=d, REML = FALSE)
```

Now we add a series of possible confounding factors such as the number of turns etc. We add the main experimental factors at the end to ensure that they're really contributing to the model over and above the confounds.

```
# Add number of turns
nTurns = lmer(trialLength.log ~ 1 +
              numberOfTurns +
              (1 + condition + incorrect | dyadNumber/playerId) +
              (1 + modalityCondition | itemId),
              data=d, REML = FALSE)
```

```
# Add whether the response was incorrect
incor = lmer(trialLength.log ~ 1 +
             numberOfTurns +
             incorrect +
             (1 + condition + incorrect | dyadNumber/playerId) +
             (1 + modalityCondition | itemId),
             data=d, REML = FALSE)
```

```
# Add multimodal signal

multim = lmer(trialLength.log ~ 1 +
              numberOfTurns +
              incorrect +
              multimodal +
              (1 + condition + incorrect | dyadNumber/playerId) +
              (1 + modalityCondition | itemId),
              data=d, REML = FALSE)
```

```
# Add effect of trial

game = lmer(trialLength.log ~ 1 +
            trialTotal +
            numberOfTurns +
            incorrect +
            multimodal +
            (1 + condition + incorrect | dyadNumber/playerId) +
            (1 + modalityCondition | itemId),
```

```

    data=d, REML = FALSE)

# Add the quadratic effect of trial
gamQuad = lmer(trialLength.log ~ 1 +
  trialTotal + I(trialTotal^2) +
  numberOfTurns +
  incorrect +
  multimodal +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

# Add modality condition
modality = lmer(trialLength.log ~ 1 + modalityCondition +
  trialTotal + I(trialTotal^2) +
  numberOfTurns +
  incorrect +
  multimodal +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

# Add stimulus condition
cond = lmer(trialLength.log ~ 1 + modalityCondition + condition +
  trialTotal + I(trialTotal^2) +
  numberOfTurns +
  incorrect +
  multimodal +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

# Add interaction between modality and stimulus condition
modXcond = lmer(trialLength.log ~ 1 + modalityCondition*condition +
  trialTotal + I(trialTotal^2) +
  numberOfTurns +
  incorrect +
  multimodal +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

# Add interaction between condition and trial
conXgame = lmer(trialLength.log ~ 1 + modalityCondition*condition +
  trialTotal + I(trialTotal^2) +
  condition:trialTotal +
  numberOfTurns +
  incorrect +
  multimodal +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

# Add interaction between modality and trial
modXgame = lmer(trialLength.log ~ 1 + modalityCondition*condition +
  trialTotal + I(trialTotal^2) +

```



```

        condition:trialTotal + modalityCondition:trialTotal +
        numberOfTurns +
        incorrect +
        multimodal +
        (1 + condition + incorrect |dyadNumber/playerId) +
        (1 + modalityCondition|itemId),
    data=d, REML = FALSE)

```

Add 3-way interaction

```

moXcoXga = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +
        I(trialTotal^2) +
        numberOfTurns +
        incorrect +
        multimodal +
        (1 + condition + incorrect |dyadNumber/playerId) +
        (1 + modalityCondition|itemId),
    data=d, REML = FALSE)

```

Interactions

interaction between turns and modality

```

nTurnXmo = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +
        I(trialTotal^2) +
        numberOfTurns + numberOfTurns:modalityCondition +
        incorrect +
        multimodal +
        (1 + condition + incorrect |dyadNumber/playerId) +
        (1 + modalityCondition|itemId),
    data=d, REML = FALSE)

```

```

nTurnXco = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +
        I(trialTotal^2) +
        numberOfTurns + numberOfTurns:modalityCondition +
        numberOfTurns:condition +
        incorrect +
        multimodal +
        (1 + condition + incorrect |dyadNumber/playerId) +
        (1 + modalityCondition|itemId),
    data=d, REML = FALSE)

```

Turn x modality x condition

*# Note that the acousitc modality had hardly any matcher turns,
#so the factor is dropped*

```

tuXmoXco = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +
        I(trialTotal^2) +
        numberOfTurns*modalityCondition*condition +
        incorrect +
        multimodal +
        (1 + condition + incorrect |dyadNumber/playerId) +
        (1 + modalityCondition|itemId),
    data=d, REML = FALSE)

```

fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

```

# Add the interaction between modality and incorrectness
moXincor = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +
  I(trialTotal^2) +
  numberOfTurns*modalityCondition*condition +
  incorrect + incorrect:modalityCondition +
  multimodal +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

```

fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

```

# Add the interaction between condition and incorrectness
coXincor = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +
  I(trialTotal^2) +
  numberOfTurns*modalityCondition*condition +
  incorrect + incorrect:modalityCondition + incorrect:condition +
  multimodal +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

```

fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

```

# Add the three-way interaction between condition, modality and incorrectness
coXmoXin = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +
  I(trialTotal^2) +
  numberOfTurns*modalityCondition*condition +
  incorrect *modalityCondition*condition +
  multimodal +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

```

fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

```

# Interaction between multimodality and condition
multiXco = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +
  I(trialTotal^2) +
  numberOfTurns*modalityCondition*condition +
  incorrect *modalityCondition*condition +
  multimodal + multimodal:condition +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

```

fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

```

# Add interaction between quadratic effect of trial and modality
modXgamQ = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +
  I(trialTotal^2) +(modalityCondition:I(trialTotal^2)) +
  numberOfTurns*modalityCondition*condition +
  incorrect *modalityCondition*condition +
  multimodal + multimodal:condition +
  (1 + condition + incorrect |dyadNumber/playerId) +
  (1 + modalityCondition|itemId),
  data=d, REML = FALSE)

```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

Check block has no effect

```
# Add block order
```

```
block = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +  
             I(trialTotal^2) +(modalityCondition:I(trialTotal^2)) +  
             numberOfTurns*modalityCondition*condition +  
             incorrect *modalityCondition*condition +  
             multimodal + multimodal:condition +  
             firstBlock +  
             (1 + condition + incorrect |dyadNumber/playerId) +  
             (1 + modalityCondition|itemId),  
             data=d, REML = FALSE)
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
# Add interaction between block order and modality
```

```
blocXmod = lmer(trialLength.log ~ 1 + modalityCondition*condition*trialTotal +  
                I(trialTotal^2) +(modalityCondition:I(trialTotal^2)) +  
                numberOfTurns*modalityCondition*condition +  
                incorrect *modalityCondition*condition +  
                multimodal + multimodal:condition +  
                firstBlock*modalityCondition +  
                (1 + condition + incorrect |dyadNumber/playerId) +  
                (1 + modalityCondition|itemId),  
                data=d, REML = TRUE) # Last model is REML to get estimates
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

Results

Compare the fit of the models:

```
modelComparison = anova(m0,modality,cond,game,modXcond,conXgame, modXgame,
  moXcoXga,nTurns,nTurnXmo,nTurnXco,tuXmoXco,
  incor,moXincor,coXincor,coXmoXin,
  multim,multiXco,
  gamQuad,modXgamQ,block, blocXmod)
```

```
## refitting model(s) with ML (instead of REML)
```

```
modelComparison
```

```
## Data: d
## Models:
## m0: trialLength.log ~ 1 + (1 + condition + incorrect | dyadNumber/playerId) +
## m0:      (1 + modalityCondition | itemId)
## nTurns: trialLength.log ~ 1 + numberOfTurns + (1 + condition + incorrect |
## nTurns:      dyadNumber/playerId) + (1 + modalityCondition | itemId)
## incor: trialLength.log ~ 1 + numberOfTurns + incorrect + (1 + condition +
## incor:      incorrect | dyadNumber/playerId) + (1 + modalityCondition |
## incor:      itemId)
## multim: trialLength.log ~ 1 + numberOfTurns + incorrect + multimodal +
## multim:      (1 + condition + incorrect | dyadNumber/playerId) + (1 +
## multim:      modalityCondition | itemId)
## game: trialLength.log ~ 1 + trialTotal + numberOfTurns + incorrect +
## game:      multimodal + (1 + condition + incorrect | dyadNumber/playerId) +
## game:      (1 + modalityCondition | itemId)
## gamQuad: trialLength.log ~ 1 + trialTotal + I(trialTotal^2) + numberOfTurns +
## gamQuad:      incorrect + multimodal + (1 + condition + incorrect | dyadNumber/playerId) +
## gamQuad:      (1 + modalityCondition | itemId)
## modality: trialLength.log ~ 1 + modalityCondition + trialTotal + I(trialTotal^2) +
## modality:      numberOfTurns + incorrect + multimodal + (1 + condition +
## modality:      incorrect | dyadNumber/playerId) + (1 + modalityCondition |
## modality:      itemId)
## cond: trialLength.log ~ 1 + modalityCondition + condition + trialTotal +
## cond:      I(trialTotal^2) + numberOfTurns + incorrect + multimodal +
## cond:      (1 + condition + incorrect | dyadNumber/playerId) + (1 +
## cond:      modalityCondition | itemId)
## modXcond: trialLength.log ~ 1 + modalityCondition * condition + trialTotal +
## modXcond:      I(trialTotal^2) + numberOfTurns + incorrect + multimodal +
## modXcond:      (1 + condition + incorrect | dyadNumber/playerId) + (1 +
## modXcond:      modalityCondition | itemId)
## conXgame: trialLength.log ~ 1 + modalityCondition * condition + trialTotal +
## conXgame:      I(trialTotal^2) + condition:trialTotal + numberOfTurns +
## conXgame:      incorrect + multimodal + (1 + condition + incorrect | dyadNumber/playerId) +
## conXgame:      (1 + modalityCondition | itemId)
## modXgame: trialLength.log ~ 1 + modalityCondition * condition + trialTotal +
## modXgame:      I(trialTotal^2) + condition:trialTotal + modalityCondition:trialTotal +
## modXgame:      numberOfTurns + incorrect + multimodal + (1 + condition +
## modXgame:      incorrect | dyadNumber/playerId) + (1 + modalityCondition |
## modXgame:      itemId)
## moXcoXga: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## moXcoXga:      I(trialTotal^2) + numberOfTurns + incorrect + multimodal +
```

```

## moXcoXga:      (1 + condition + incorrect | dyadNumber/playerId) + (1 +
## moXcoXga:      modalityCondition | itemId)
## nTurnXmo: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## nTurnXmo:      I(trialTotal^2) + numberOfTurns + numberOfTurns:modalityCondition +
## nTurnXmo:      incorrect + multimodal + (1 + condition + incorrect | dyadNumber/playerId) +
## nTurnXmo:      (1 + modalityCondition | itemId)
## nTurnXco: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## nTurnXco:      I(trialTotal^2) + numberOfTurns + numberOfTurns:modalityCondition +
## nTurnXco:      numberOfTurns:condition + incorrect + multimodal + (1 + condition +
## nTurnXco:      incorrect | dyadNumber/playerId) + (1 + modalityCondition |
## nTurnXco:      itemId)
## tuXmoXco: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## tuXmoXco:      I(trialTotal^2) + numberOfTurns * modalityCondition * condition +
## tuXmoXco:      incorrect + multimodal + (1 + condition + incorrect | dyadNumber/playerId) +
## tuXmoXco:      (1 + modalityCondition | itemId)
## moXincor: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## moXincor:      I(trialTotal^2) + numberOfTurns * modalityCondition * condition +
## moXincor:      incorrect + incorrect:modalityCondition + multimodal + (1 +
## moXincor:      condition + incorrect | dyadNumber/playerId) + (1 + modalityCondition |
## moXincor:      itemId)
## coXincor: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## coXincor:      I(trialTotal^2) + numberOfTurns * modalityCondition * condition +
## coXincor:      incorrect + incorrect:modalityCondition + incorrect:condition +
## coXincor:      multimodal + (1 + condition + incorrect | dyadNumber/playerId) +
## coXincor:      (1 + modalityCondition | itemId)
## coXmoXin: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## coXmoXin:      I(trialTotal^2) + numberOfTurns * modalityCondition * condition +
## coXmoXin:      incorrect * modalityCondition * condition + multimodal +
## coXmoXin:      (1 + condition + incorrect | dyadNumber/playerId) + (1 +
## coXmoXin:      modalityCondition | itemId)
## multiXco: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## multiXco:      I(trialTotal^2) + numberOfTurns * modalityCondition * condition +
## multiXco:      incorrect * modalityCondition * condition + multimodal +
## multiXco:      multimodal:condition + (1 + condition + incorrect | dyadNumber/playerId) +
## multiXco:      (1 + modalityCondition | itemId)
## modXgamQ: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## modXgamQ:      I(trialTotal^2) + (modalityCondition:I(trialTotal^2)) + numberOfTurns *
## modXgamQ:      modalityCondition * condition + incorrect * modalityCondition *
## modXgamQ:      condition + multimodal + multimodal:condition + (1 + condition +
## modXgamQ:      incorrect | dyadNumber/playerId) + (1 + modalityCondition |
## modXgamQ:      itemId)
## block: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## block:      I(trialTotal^2) + (modalityCondition:I(trialTotal^2)) + numberOfTurns *
## block:      modalityCondition * condition + incorrect * modalityCondition *
## block:      condition + multimodal + multimodal:condition + firstBlock +
## block:      (1 + condition + incorrect | dyadNumber/playerId) + (1 +
## block:      modalityCondition | itemId)
## blocXmod: trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## blocXmod:      I(trialTotal^2) + (modalityCondition:I(trialTotal^2)) + numberOfTurns *
## blocXmod:      modalityCondition * condition + incorrect * modalityCondition *
## blocXmod:      condition + multimodal + multimodal:condition + firstBlock *
## blocXmod:      modalityCondition + (1 + condition + incorrect | dyadNumber/playerId) +
## blocXmod:      (1 + modalityCondition | itemId)
##
##      Df      AIC      BIC      logLik deviance      Chisq Chi Df Pr(>Chisq)

```

```
## m0      20 2686.0 2796.8 -1323.01 2646.0
## nTurns  21 2187.9 2304.2 -1072.95 2145.9 500.1165      1 < 2.2e-16 ***
## incor   22 2163.3 2285.2 -1059.67 2119.3 26.5501      1 2.568e-07 ***
## multim  23 2163.9 2291.3 -1058.95 2117.9 1.4555      1 0.2276463
## game    24 1736.8 1869.8 -844.41 1688.8 429.0695      1 < 2.2e-16 ***
## gamQuad 25 1687.2 1825.7 -818.59 1637.2 51.6477      1 6.641e-13 ***
## modality 27 1690.2 1839.8 -818.09 1636.2 1.0026      2 0.6057547
## cond    28 1691.8 1847.0 -817.92 1635.8 0.3314      1 0.5648583
## modXcond 30 1681.6 1847.8 -810.81 1621.6 14.2279      2 0.0008137 ***
## conXgame 31 1683.4 1855.2 -810.71 1621.4 0.2007      1 0.6541651
## modXgame 33 1674.0 1856.9 -804.02 1608.0 13.3687      2 0.0012503 **
## moXcoXga 35 1676.4 1870.3 -803.20 1606.4 1.6525      2 0.4376777
## nTurnXmo 37 1679.4 1884.4 -802.72 1605.4 0.9496      2 0.6219993
## nTurnXco 38 1681.2 1891.8 -802.61 1605.2 0.2143      1 0.6434528
## tuXmoXco 39 1679.1 1895.2 -800.55 1601.1 4.1286      1 0.0421637 *
## moXincor 41 1678.0 1905.2 -798.02 1596.0 5.0600      2 0.0796592 .
## coXincor 42 1679.8 1912.5 -797.89 1595.8 0.2665      1 0.6056599
## coXmoXin 44 1682.7 1926.5 -797.37 1594.7 1.0347      2 0.5960836
## multiXco 45 1684.4 1933.7 -797.22 1594.4 0.2999      1 0.5839638
## modXgamQ 47 1682.2 1942.6 -794.10 1588.2 6.2397      2 0.0441644 *
## block   48 1684.0 1949.9 -793.98 1588.0 0.2433      1 0.6218431
## blocXmod 50 1686.4 1963.4 -793.20 1586.4 1.5501      2 0.4606780
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pick final model for estimates:

```
finalModel = modXgamQ
```

Final model estimates:

```
summary(finalModel)

## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula:
## trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
## I(trialTotal^2) + (modalityCondition:I(trialTotal^2)) + numberOfTurns *
## modalityCondition * condition + incorrect * modalityCondition *
## condition + multimodal + multimodal:condition + (1 + condition +
## incorrect | dyadNumber/playerId) + (1 + modalityCondition |
## itemId)
## Data: d
##
##      AIC      BIC    logLik deviance df.resid
## 1682.2   1942.6   -794.1   1588.2     1835
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.460 -0.626 -0.047  0.561  5.052
##
## Random effects:
##      Groups                Name                Variance Std.Dev. Corr
## playerId:dyadNumber (Intercept)            0.0378093 0.19445
##                      conditionVisual        0.0255327 0.15979 -0.54
##                      incorrectTRUE           0.0102262 0.10112 -0.88
## itemId (Intercept)              0.0238895 0.15456
```

```

##          modalityConditionvisual 0.0024441 0.04944 0.97
##          modalityConditionvocal 0.0105709 0.10281 -0.08
## dyadNumber (Intercept)          0.0540030 0.23239
##          conditionVisual         0.0130060 0.11404 -0.20
##          incorrectTRUE           0.0009362 0.03060 -0.16
## Residual                        0.1206359 0.34733
##
##
## 0.19
##
##
## 0.15
##
##
## -0.94
##
## Number of obs: 1882, groups:
## playerId:dyadNumber, 30; itemId, 16; dyadNumber, 15
##
## Fixed effects:
##
##          Estimate Std. Error
## (Intercept)      -0.826423 0.144223
## modalityConditionvisual      0.355442 0.186475
## modalityConditionvocal      0.467219 0.257923
## conditionVisual      0.275850 0.124580
## trialTotal      -0.181853 0.017336
## I(trialTotal^2)      0.064879 0.012027
## numberOfTurns      0.298496 0.028196
## incorrectTRUE      0.280224 0.085600
## multimodalTRUE      0.054218 0.057072
## modalityConditionvisual:conditionVisual -0.063808 0.136929
## modalityConditionvocal:conditionVisual -0.594223 0.128863
## modalityConditionvisual:trialTotal      0.046677 0.024499
## modalityConditionvocal:trialTotal      0.032054 0.024559
## conditionVisual:trialTotal      -0.013144 0.024796
## modalityConditionvisual:I(trialTotal^2) -0.039148 0.016823
## modalityConditionvocal:I(trialTotal^2) -0.006688 0.016822
## modalityConditionvisual:numberOfTurns      0.102780 0.043898
## modalityConditionvocal:numberOfTurns      0.098083 0.181649
## conditionVisual:numberOfTurns      0.059961 0.037388
## modalityConditionvisual:incorrectTRUE -0.109282 0.112701
## modalityConditionvocal:incorrectTRUE -0.250303 0.116511
## conditionVisual:incorrectTRUE      -0.003167 0.099083
## conditionVisual:multimodalTRUE      -0.071988 0.103008
## modalityConditionvisual:conditionVisual:trialTotal      0.026921 0.034966
## modalityConditionvocal:conditionVisual:trialTotal -0.004871 0.034882
## modalityConditionvisual:conditionVisual:numberOfTurns -0.114333 0.057102
## modalityConditionvisual:conditionVisual:incorrectTRUE -0.023298 0.134163
## modalityConditionvocal:conditionVisual:incorrectTRUE      0.101245 0.130970
##
##          t value
## (Intercept)      -5.730
## modalityConditionvisual      1.906
## modalityConditionvocal      1.811

```

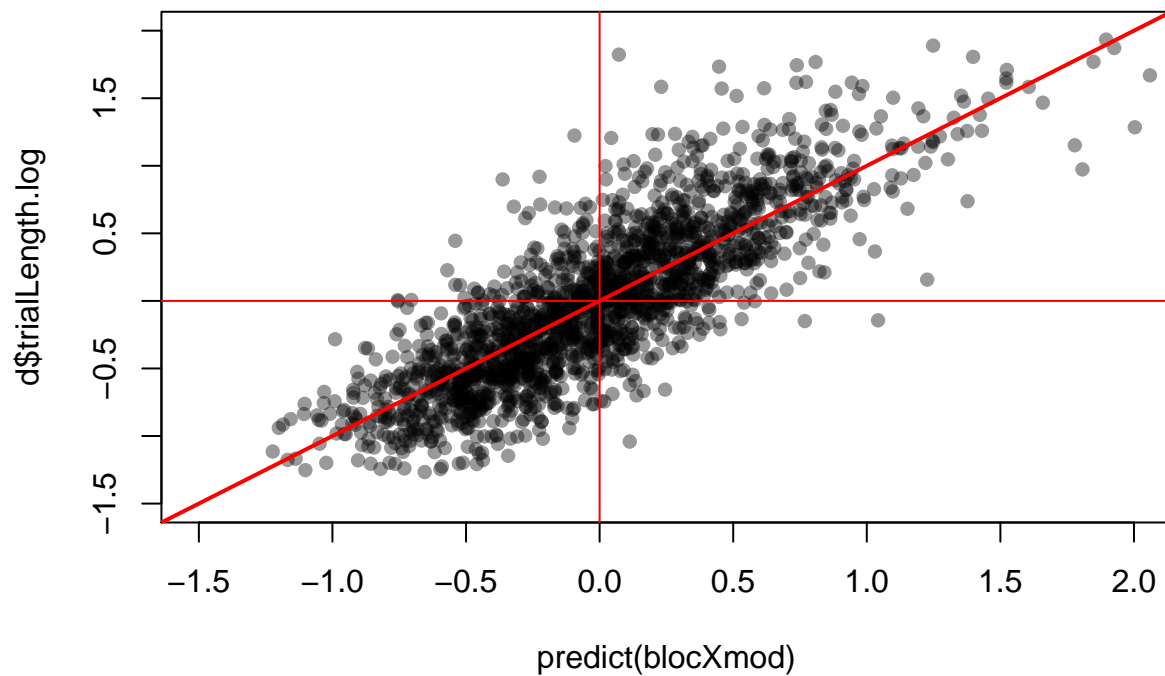
```
## conditionVisual                2.214
## trialTotal                    -10.490
## I(trialTotal^2)                5.395
## numberOfTurns                 10.587
## incorrectTRUE                 3.274
## multimodalTRUE                0.950
## modalityConditionvisual:conditionVisual -0.466
## modalityConditionvocal:conditionVisual -4.611
## modalityConditionvisual:trialTotal    1.905
## modalityConditionvocal:trialTotal    1.305
## conditionVisual:trialTotal          -0.530
## modalityConditionvisual:I(trialTotal^2) -2.327
## modalityConditionvocal:I(trialTotal^2) -0.398
## modalityConditionvisual:numberOfTurns 2.341
## modalityConditionvocal:numberOfTurns 0.540
## conditionVisual:numberOfTurns        1.604
## modalityConditionvisual:incorrectTRUE -0.970
## modalityConditionvocal:incorrectTRUE -2.148
## conditionVisual:incorrectTRUE        -0.032
## conditionVisual:multimodalTRUE       -0.699
## modalityConditionvisual:conditionVisual:trialTotal 0.770
## modalityConditionvocal:conditionVisual:trialTotal -0.140
## modalityConditionvisual:conditionVisual:numberOfTurns -2.002
## modalityConditionvisual:conditionVisual:incorrectTRUE -0.174
## modalityConditionvocal:conditionVisual:incorrectTRUE 0.773

##
## Correlation matrix not shown by default, as p = 28 > 12.
## Use print(x, correlation=TRUE) or
##   vcov(x)      if you need it

## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

Check model predictions. The model predictions are in the right range and direction, fitting linear quite well:

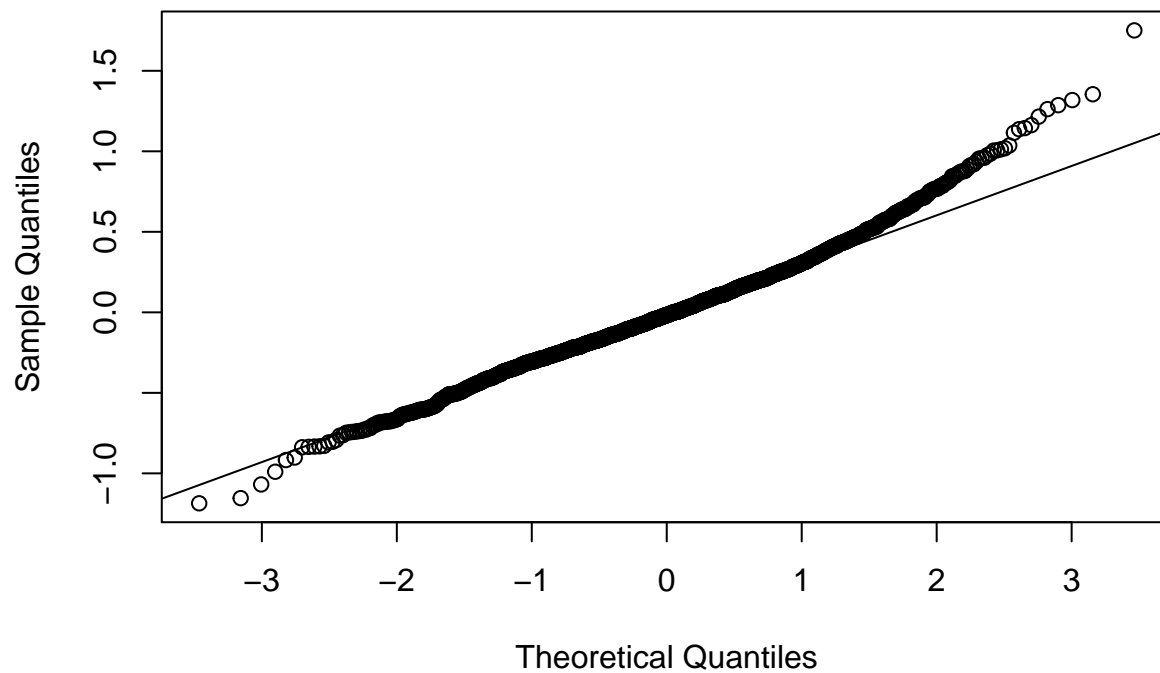
```
plot(predict(blocXmod),d$trialLength.log, pch=16, col=rgb(0,0,0,0.4),
      ylim=c(-1.5,2),xlim=c(-1.5,2))
abline(a=0,b=1, col=2, lwd=2)
abline(h=0, col=2)
abline(v=0, col=2)
```

The residuals are ok, though it tends to do worse at higher values. This is expected from using the log scale.

```
qqnorm(resid(blocXmod))
qqline(resid(blocXmod))
```

Normal Q-Q Plot



Plot the fixed effects

Relabel the effects:

```
feLabels = matrix(c(
  "(Intercept)" , "Intercept" , NA,
  "modalityConditionvisual" , "Visual modality", "modality",
  "modalityConditionvocal" , "Acoustic modality", "modality",
  "conditionVisual" , "Visual stimuli", "cond",
  "trialTotal" , "Game", "game",
  "modalityConditionvisual:conditionVisual" , "Visual modality:Visual stimuli", "modXcond",
  "modalityConditionvocal:conditionVisual" , "Acoustic modality:Visual stimuli", "modXcond",
  "modalityConditionvisual:trialTotal" , "Visual modality:Game", "modXgame",
  "modalityConditionvocal:trialTotal" , "Acoustic modality:Game", "modXgame",
  "conditionVisual:trialTotal" , "Visual stimuli:Game", "conXgame",
  "modalityConditionvisual:conditionVisual:trialTotal", "Visual modality:Visual stimuli:Game", "moXcoXga",
  "modalityConditionvocal:conditionVisual:trialTotal", "Acoustic modality:Visual stimuli:Game", "moXcoXga",
  "incorrectTRUE", "Incorrect", "incor",
  "modalityConditionvisual:incorrectTRUE", "Visual modality:Incorrect", "moXincor",
  "modalityConditionvocal:incorrectTRUE", "Acoustic modality:Incorrect", "moXincor",
  "modalityConditionvisual:I(trialTotal^2)", "Visual modality:Game^2", "modXgamQ",
  "modalityConditionvocal:I(trialTotal^2)", "Acoustic modality:Game^2", "modXgamQ",
  "I(trialTotal^2)", "Game^2", "gamQuad",
  "firstBlockVisual", "Visual stims first", "block",
  "modalityConditionvisual:firstBlockVisual", "Visual modality:Visual stim first", "blocXmod",
  "modalityConditionvocal:firstBlockVisual", "Acoustic modality:Visual stim first", "blocXmod",
  "conditionVisual:incorrectTRUE", "Visual stimuli:incorrect", "coXincor",
  "modalityConditionvisual:conditionVisual:incorrectTRUE", "Visual modality:Visual stimuli:incorrect", "coXincor",
  "modalityConditionvocal:conditionVisual:incorrectTRUE", "Acoustic modality:Visual stimuli:incorrect", "coXincor",
  "modalityConditionvisual:conditionVisual:numberOfTurns", "VisualModality:Visual stim:NumTurns", "tuXmoXco",
  "modalityConditionvocal:conditionVisual:numberOfTurns", "Vocal Modality:Visual stim:NumTurns", "tuXmoXco",
  "conditionVisual:numberOfTurns", "Visual stim:NumTurns", "nTurnXco",
  "modalityConditionvisual:numberOfTurns", "VisualModality:NumTurns", "nTurnXmo",
  "modalityConditionvocal:numberOfTurns", "Vocal Modality:NumTurns", "nTurnXmo",
  "numberOfTurns", "Number of turns", "nTurns",
  "multimodalTRUE", "Multimodal T1", "multim",
  "conditionVisual:multimodalTRUE", "VisualStim:MultimodalT1", "multiXco"
), ncol=3, byrow = T)

feLabels2 = as.vector(feLabels[match(names(fixef(finalModel)), feLabels[,1]), 2])
feModel = as.vector(feLabels[match(names(fixef(finalModel)), feLabels[,1]), 3])

sig = modelComparison$`Pr(>Chisq)`
names(sig) = rownames(modelComparison)

sig.data = data.frame(estimate = fixef(finalModel),
  y=1:length(fixef(finalModel)),
  sig=sig[feModel])

cols= c("black", 'red')
sig.data$pointCol = cols[1]
sig.data$pointCol[!is.na(sig.data$sig)] =
  cols[1 + (sig.data$sig[!is.na(sig.data$sig)] < 0.05)]
```

```
# Mark marginal effects
#sig.data$pointCol[!is.na(sig.data$sig) &
#                      sig.data$sig < 0.1 &
#                      sig.data$sig >=0.05] = "orange"
```

```
sig.data$fade = sig.data$sig > 0.05
```

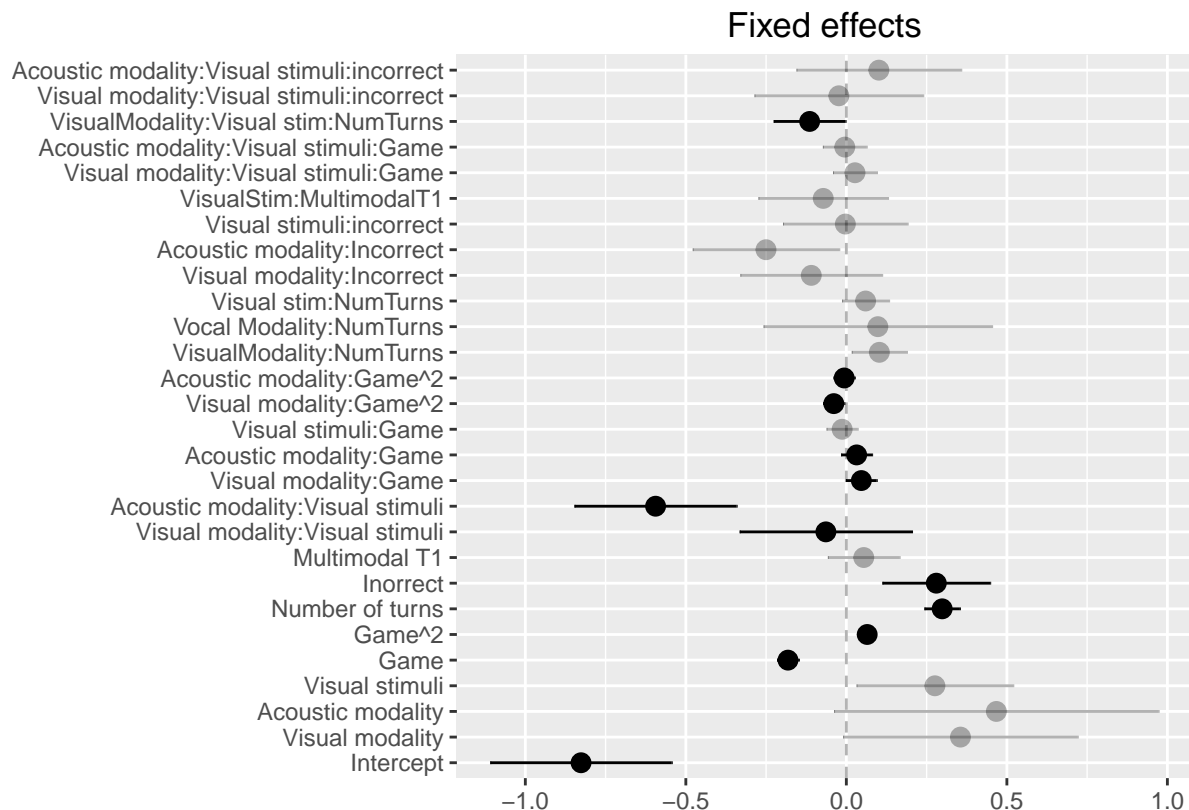
Plot the strength of the fixed effects:

```
x = sjp.lmer(finalModel, 'fe',
  show.intercept = T,
  sort.est=NULL,
  axis.labels = feLabels2[2:length(feLabels2)],
  xlab="Trial time (log ms)",
  geom.colors = c(1,1),
  show.p=F,
  show.values = F,
  p.kr = FALSE,
  string.interc="Intercept",
  prnt.plot = F)
```

```
## Computing p-values via Wald-statistics approximation (treating t as Wald z).
```

```
x$plot.list[[1]]$data$fade = sig.data$fade
```

```
x$plot.list[[1]]
```



Attempt plot with axes in milliseconds.

```
convertEst = function(X){
  exp(meanLogTrialLength+X) - exp(meanLogTrialLength)
}

x$plot.list[[1]]$data$estimate = convertEst(x$plot.list[[1]]$data$estimate)
x$plot.list[[1]]$data$conf.low = convertEst(x$plot.list[[1]]$data$conf.low)
x$plot.list[[1]]$data$conf.high = convertEst(x$plot.list[[1]]$data$conf.high)

sig.data2 = sig.data
sig.data2$estimate = x$plot.list[[1]]$data$estimate
sig.data2$estimate.lower = x$plot.list[[1]]$data$conf.low
sig.data2$estimate.upper = x$plot.list[[1]]$data$conf.high

x$plot.list[[1]]$data$fade = sig.data2$fade

x$plot.list[[1]] +
  scale_y_continuous(name="Difference (ms)") +
  scale_x_discrete(labels=feLabels2) +
  #geom_point(data=sig.data2,aes(y=estimate,x=y,fade=fade), color=sig.data$pointCol) +
  coord_flip(ylim=c(-5000,10000))
```

Scale for 'x' is already present. Adding another scale for 'x', which
will replace the existing scale.

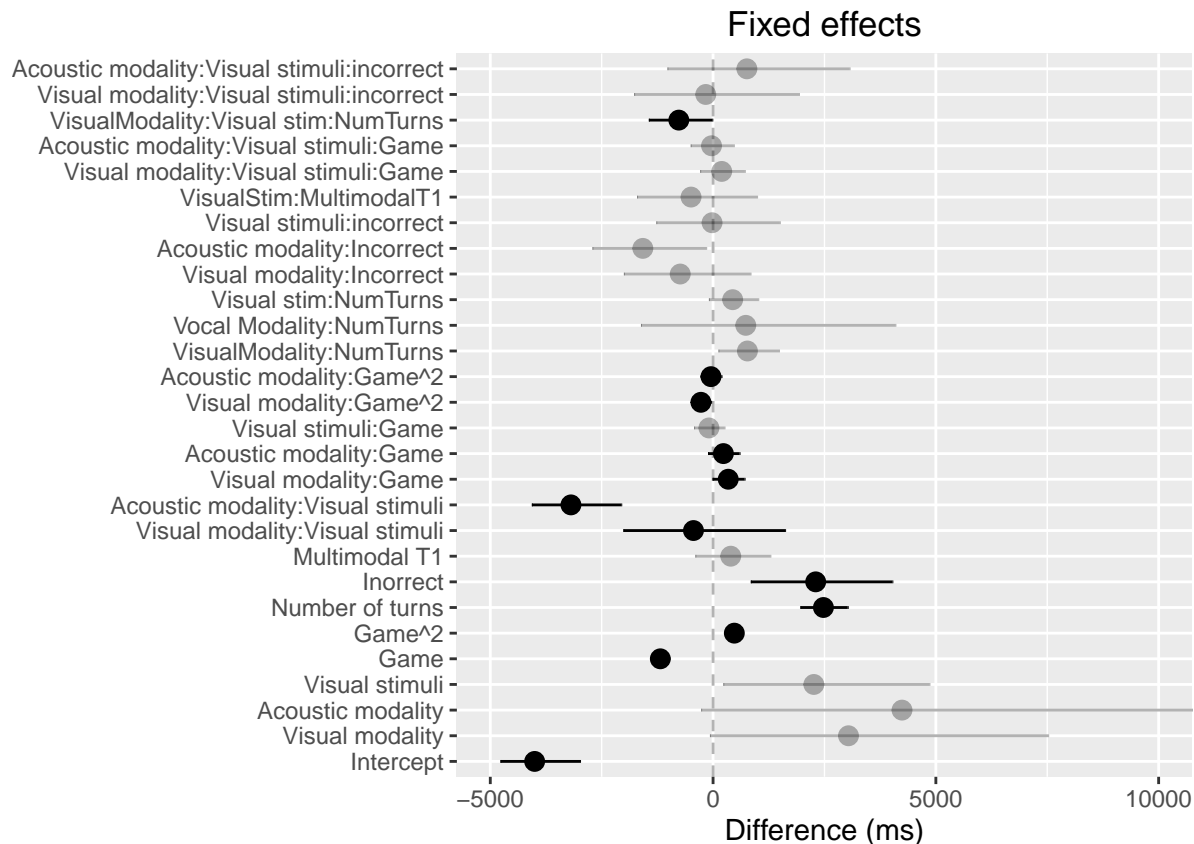


Table for paper

```
outdata = x$plot.list[[1]]$data[,c("estimate", "conf.low", 'conf.high')]

outdata$estimate = round(outdata$estimate)
outdata$conf.low = round(outdata$conf.low)
outdata$conf.high = round(outdata$conf.high)
#outdata = outdata[2:nrow(outdata),]

xd = as.data.frame(summary(finalModel)$coef)
#xd = xd[2:nrow(xd),]
outdata$wald.t = xd$t value`

sig = modelComparison$`Pr(>Chisq)`
names(sig) = rownames(modelComparison)
sigx = sig[feModel]
#sigx = sigx[2:length(sigx)]

outdata$model.comparison.p = sigx
outdata$estimate = paste(
  c("", "+")[1+(outdata$estimate>0)],
  as.character(outdata$estimate), sep='')

outdata$label = feLabels2

outdata = outdata[,c("label", "estimate", "conf.low",
  "conf.high", "wald.t",
  "model.comparison.p")]

write.csv(outdata[2:nrow(outdata),], file="../../results/tables/Efficiency_FixedEffects.csv")
```

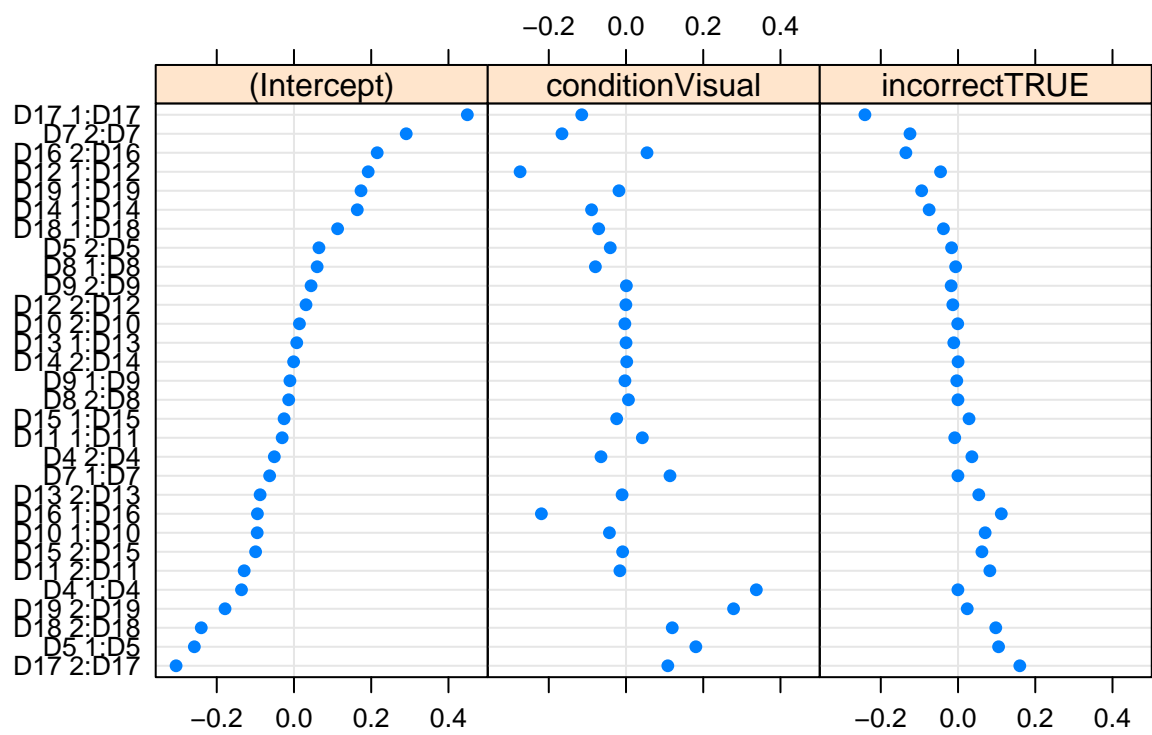
Random effects

There is a reasonable amount of variation in the random effects, suggesting that dyads and players differ. This justifies the use of mixed effects modelling.

```
dotplot(ranef(finalModel))
```

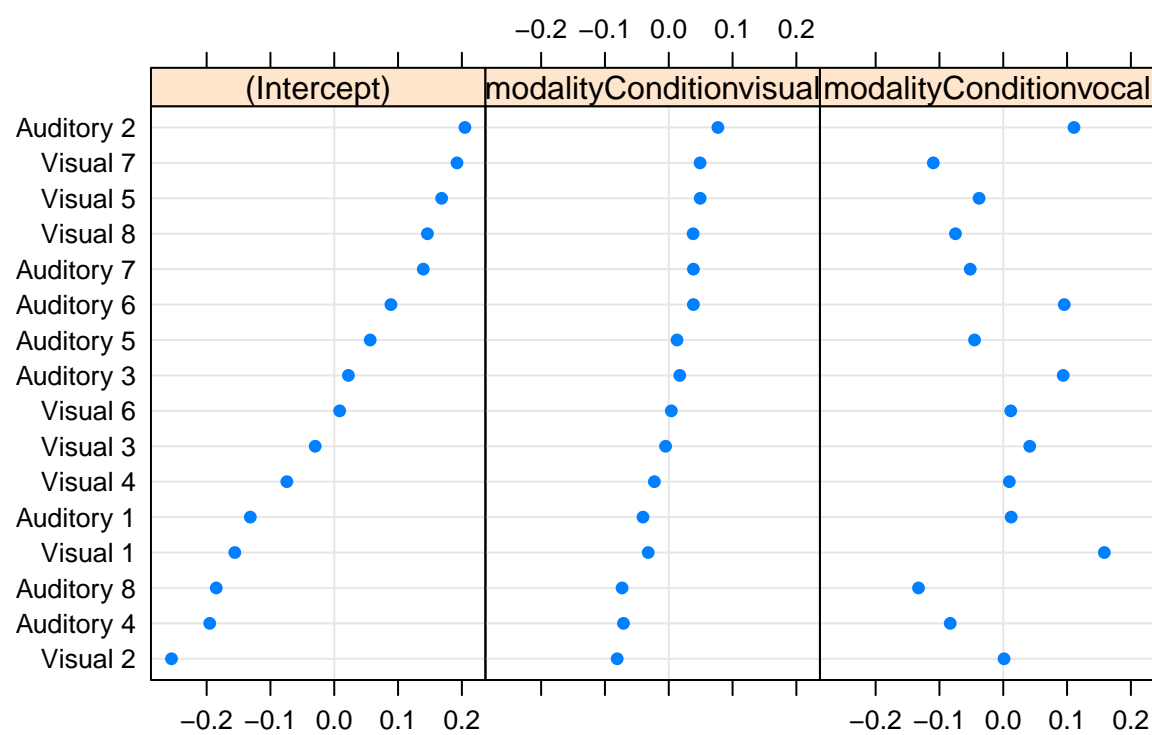
```
## $`playerId:dyadNumber`
```

playerId:dyadNumber



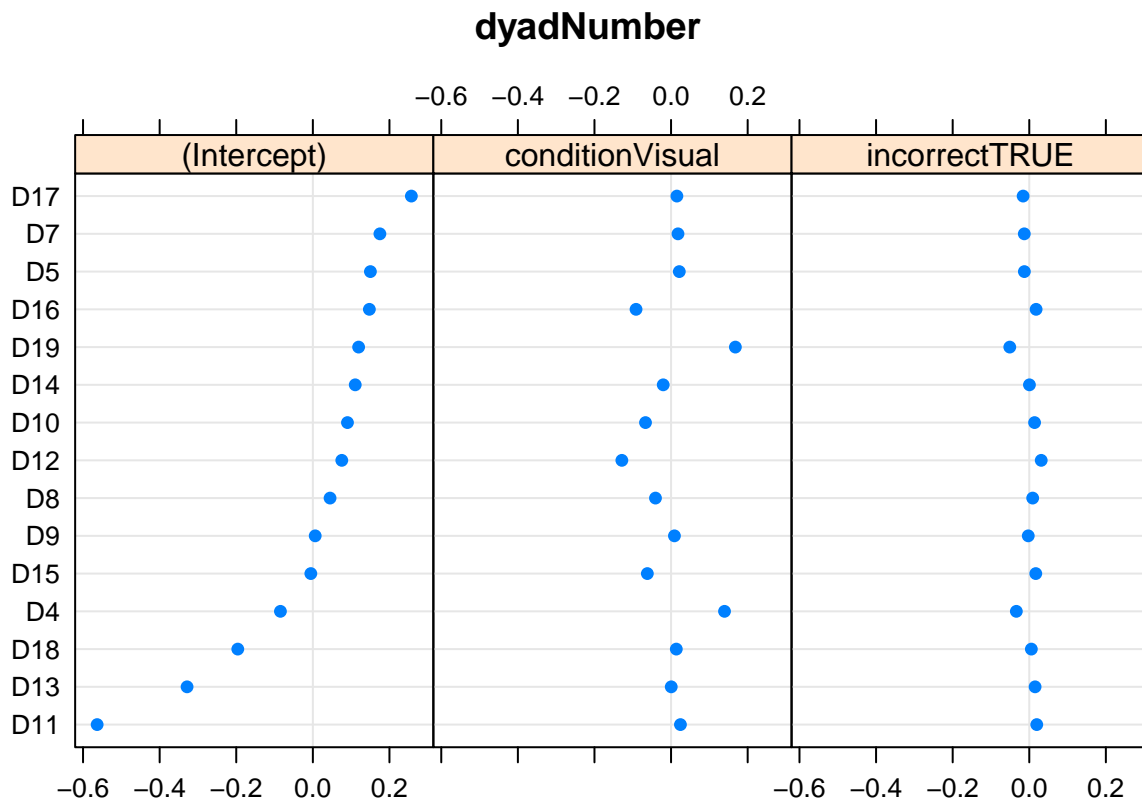
\$ItemId

itemId



##

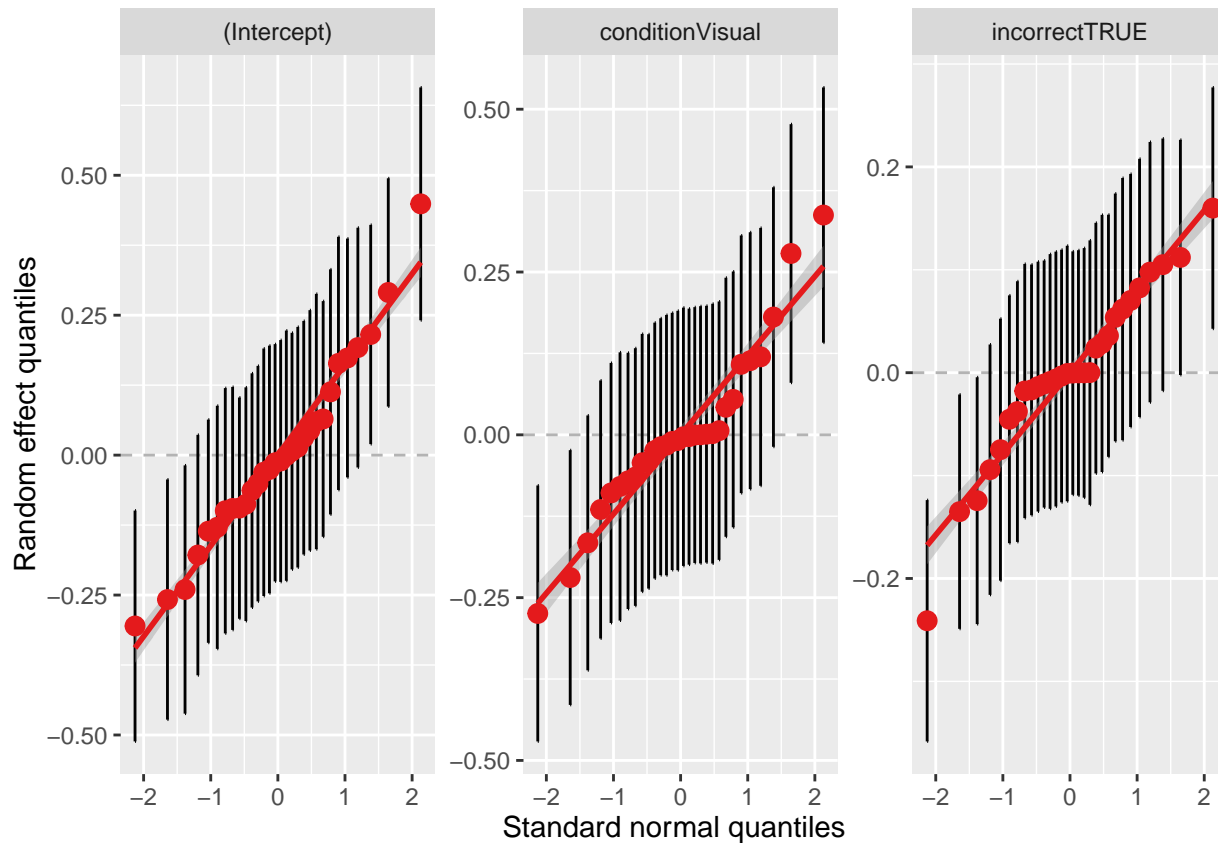
```
## $dyadNumber
```



qq-plots of random effects

```
sjp.lmer(finalModel, type = "re.qq")
```

```
## Testing for normal distribution. Dots should be plotted along the line.
```



Relevel factors to see other comparisons

```
d2 = d
d2$condition = relevel(d2$condition, "Visual")
fm2 = update(finalModel, data=d2)

## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
summary(fm2)

## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula:
## trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
##   I(trialTotal^2) + (modalityCondition:I(trialTotal^2)) + numberOfTurns *
##   modalityCondition * condition + incorrect * modalityCondition *
##   condition + multimodal + multimodal:condition + (1 + condition +
##   incorrect | dyadNumber/playerId) + (1 + modalityCondition |
##   itemId)
## Data: d2
##
##      AIC      BIC   logLik deviance df.resid
## 1682.2  1942.6  -794.1  1588.2    1835
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.460 -0.626 -0.047  0.561  5.052
```



```

##
## Random effects:
## Groups          Name                Variance Std.Dev. Corr
## playerId:dyadNumber (Intercept)      0.0296954 0.17232
##                  conditionAuditory    0.0255327 0.15979 -0.32
##                  incorrectTRUE        0.0102262 0.10112 -0.82
## itemId          (Intercept)      0.0238895 0.15456
##                  modalityConditionvisual 0.0024440 0.04944  0.97
##                  modalityConditionvocal 0.0105709 0.10281 -0.08
## dyadNumber      (Intercept)      0.0566260 0.23796
##                  conditionAuditory    0.0130061 0.11404 -0.29
##                  incorrectTRUE        0.0009362 0.03060 -0.60
## Residual                0.1206359 0.34733
##
##
## -0.19
##
##  0.15
##
##  0.94
##
## Number of obs: 1882, groups:
## playerId:dyadNumber, 30; itemId, 16; dyadNumber, 15
##
## Fixed effects:
##
##                                     Estimate
## (Intercept)                        -0.550573
## modalityConditionvisual              0.291633
## modalityConditionvocal             -0.127003
## conditionAuditory                  -0.275850
## trialTotal                         -0.194997
## I(trialTotal^2)                     0.064879
## numberOfTurns                      0.358456
## incorrectTRUE                      0.277057
## multimodalTRUE                     -0.017770
## modalityConditionvisual:conditionAuditory 0.063808
## modalityConditionvocal:conditionAuditory 0.594222
## modalityConditionvisual:trialTotal    0.073599
## modalityConditionvocal:trialTotal    0.027184
## conditionAuditory:trialTotal         0.013144
## modalityConditionvisual:I(trialTotal^2) -0.039148
## modalityConditionvocal:I(trialTotal^2) -0.006688
## modalityConditionvisual:numberOfTurns -0.011553
## modalityConditionvocal:numberOfTurns  0.098083
## conditionAuditory:numberOfTurns     -0.059961
## modalityConditionvisual:incorrectTRUE -0.132580
## modalityConditionvocal:incorrectTRUE  -0.149058
## conditionAuditory:incorrectTRUE      0.003167
## conditionAuditory:multimodalTRUE     0.071988
## modalityConditionvisual:conditionAuditory:trialTotal -0.026921
## modalityConditionvocal:conditionAuditory:trialTotal  0.004871

```

```

## modalityConditionvisual:conditionAuditory:numberOfTurns 0.114333
## modalityConditionvisual:conditionAuditory:incorrectTRUE 0.023298
## modalityConditionvocal:conditionAuditory:incorrectTRUE -0.101245
## Std. Error t value
## (Intercept) 0.137023 -4.018
## modalityConditionvisual 0.179036 1.629
## modalityConditionvocal 0.252686 -0.503
## conditionAuditory 0.124580 -2.214
## trialTotal 0.017745 -10.989
## I(trialTotal^2) 0.012027 5.395
## numberOfTurns 0.024884 14.405
## incorrectTRUE 0.070565 3.926
## multimodalTRUE 0.086214 -0.206
## modalityConditionvisual:conditionAuditory 0.136929 0.466
## modalityConditionvocal:conditionAuditory 0.128863 4.611
## modalityConditionvisual:trialTotal 0.024957 2.949
## modalityConditionvocal:trialTotal 0.024794 1.096
## conditionAuditory:trialTotal 0.024796 0.530
## modalityConditionvisual:I(trialTotal^2) 0.016823 -2.327
## modalityConditionvocal:I(trialTotal^2) 0.016822 -0.398
## modalityConditionvisual:numberOfTurns 0.037011 -0.312
## modalityConditionvocal:numberOfTurns 0.181649 0.540
## conditionAuditory:numberOfTurns 0.037388 -1.604
## modalityConditionvisual:incorrectTRUE 0.100971 -1.313
## modalityConditionvocal:incorrectTRUE 0.092505 -1.611
## conditionAuditory:incorrectTRUE 0.099083 0.032
## conditionAuditory:multimodalTRUE 0.103008 0.699
## modalityConditionvisual:conditionAuditory:trialTotal 0.034966 -0.770
## modalityConditionvocal:conditionAuditory:trialTotal 0.034882 0.140
## modalityConditionvisual:conditionAuditory:numberOfTurns 0.057102 2.002
## modalityConditionvisual:conditionAuditory:incorrectTRUE 0.134163 0.174
## modalityConditionvocal:conditionAuditory:incorrectTRUE 0.130970 -0.773

##
## Correlation matrix not shown by default, as p = 28 > 12.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it

## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient

feLabelsB = feLabels2
feLabelsB = gsub("Visual stimuli","Acoustic stimuli",feLabelsB)
feLabelsB = gsub("VisualStim","AcousticStim",feLabelsB)
feLabelsB = gsub("Visual stim","AcousticStim",feLabelsB)

x2 = sjp.lmer(fm2, 'fe',
  show.intercept = T,
  sort.est=NULL,
  axis.labels = feLabelsB[2:length(feLabelsB)],
  xlab="Trial time (log ms)",
  geom.colors = c(1,1),
  show.p=F,
  show.values = F,
  p.kr = FALSE,

```

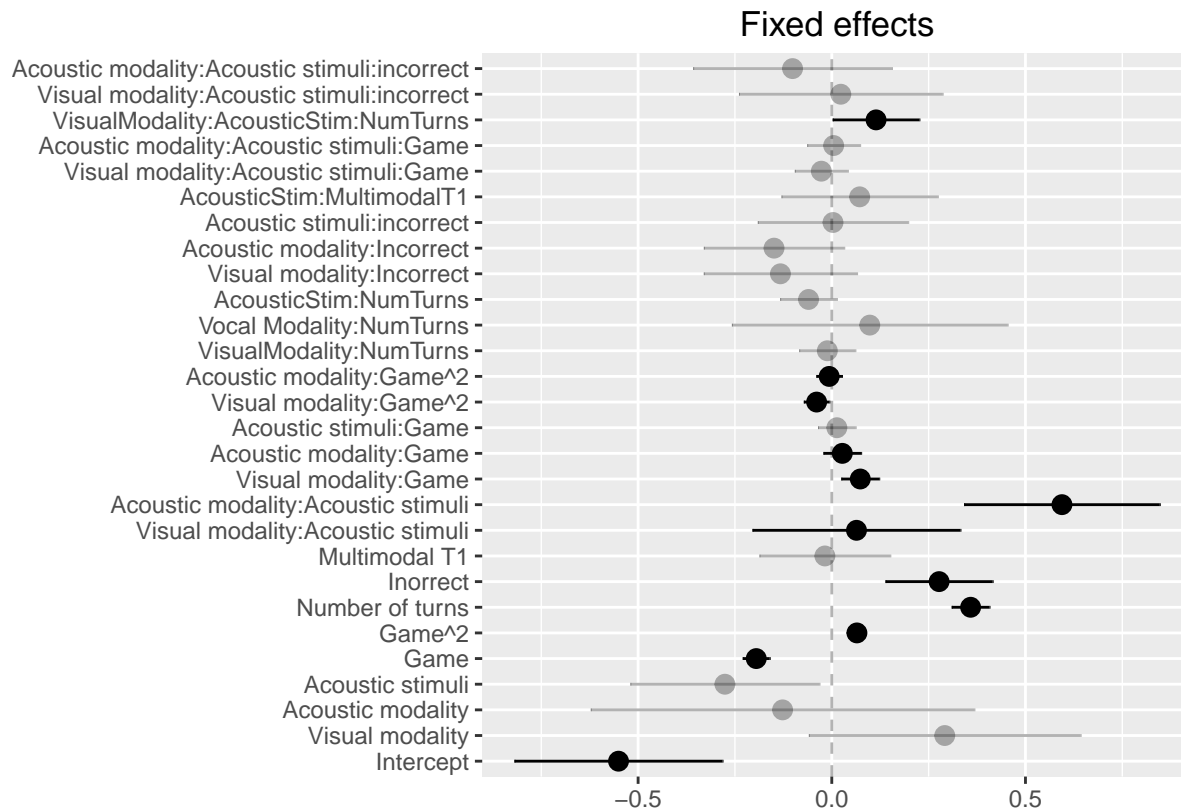
```
string.interc="Intercept",
prnt.plot = F)
```

```
## Computing p-values via Wald-statistics approximation (treating t as Wald z).
```

```
## Warning: Deprecated, use tibble::rownames_to_column() instead.
```

```
x2$plot.list[[1]]$data$fade = sig.data$fade
```

```
x2$plot.list[[1]]
```



```
d2 = d
```

```
d2$modalityCondition = relevel(d2$modalityCondition,"visual")
```

```
fm2 = update(finalModel, data=d2)
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
```

```
## $checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control
```

```
## $checkConv, : Model failed to converge: degenerate Hessian with 1 negative
```

```
## eigenvalues
```

```
summary(fm2)
```

```
## Linear mixed model fit by maximum likelihood ['lmerMod']
```

```
## Formula:
```

```
## trialLength.log ~ 1 + modalityCondition * condition * trialTotal +
```

```
## I(trialTotal^2) + (modalityCondition:I(trialTotal^2)) + numberOfTurns *
```

```
## modalityCondition * condition + incorrect * modalityCondition *
```

```

##      condition + multimodal + multimodal:condition + (1 + condition +
##      incorrect | dyadNumber/playerId) + (1 + modalityCondition |
##      itemId)
## Data: d2
##
##      AIC      BIC    logLik deviance df.resid
## 1682.4    1942.7   -794.2   1588.4     1835
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.4651 -0.6250 -0.0458  0.5585  5.0525
##
## Random effects:
## Groups              Name                Variance Std.Dev. Corr
## playerId:dyadNumber (Intercept)          0.037760 0.19432
##                      conditionVisual      0.025526 0.15977  -0.54
##                      incorrectTRUE         0.010192 0.10095  -0.88  0.18
## itemId              (Intercept)          0.041295 0.20321
##                      modalityConditionmulti 0.002396 0.04895  -1.00
##                      modalityConditionvocal 0.012296 0.11089  -0.48  0.48
## dyadNumber          (Intercept)          0.053966 0.23231
##                      conditionVisual      0.012967 0.11387  -0.20
##                      incorrectTRUE         0.000956 0.03092  -0.14 -0.94
## Residual                                0.120655 0.34735
## Number of obs: 1882, groups:
## playerId:dyadNumber, 30; itemId, 16; dyadNumber, 15
##
## Fixed effects:
##
##                                Estimate Std. Error
## (Intercept)                   -0.470549   0.147707
## modalityConditionmulti         -0.354645   0.186386
## modalityConditionvocal          0.112854   0.256126
## conditionVisual                 0.211717   0.137621
## trialTotal                     -0.135246   0.017358
## I(trialTotal^2)                 0.025736   0.011766
## numberOfTurns                   0.400891   0.033830
## incorrectTRUE                   0.171023   0.074831
## multimodalTRUE                  0.052137   0.057009
## modalityConditionmulti:conditionVisual 0.062662   0.136819
## modalityConditionvocal:conditionVisual -0.416301   0.123696
## modalityConditionmulti:trialTotal    -0.046487   0.024502
## modalityConditionvocal:trialTotal    -0.014616   0.024570
## conditionVisual:trialTotal           0.013900   0.024779
## modalityConditionmulti:I(trialTotal^2) 0.039123   0.016825
## modalityConditionvocal:I(trialTotal^2) 0.032469   0.016636
## modalityConditionmulti:numberOfTurns -0.102149   0.043909
## modalityConditionvocal:numberOfTurns -0.005647   0.182667
## conditionVisual:numberOfTurns        -0.054093   0.043390
## modalityConditionmulti:incorrectTRUE   0.110594   0.112705
## modalityConditionvocal:incorrectTRUE  -0.143147   0.108924
## conditionVisual:incorrectTRUE         -0.026671   0.092758
## conditionVisual:multimodalTRUE        -0.069179   0.102912
## modalityConditionmulti:conditionVisual:trialTotal -0.027219   0.034969
## modalityConditionvocal:conditionVisual:trialTotal -0.031860   0.034863

```

```

## modalityConditionmulti:conditionVisual:numberOfTurns  0.113970  0.057113
## modalityConditionmulti:conditionVisual:incorrectTRUE  0.022003  0.134163
## modalityConditionvocal:conditionVisual:incorrectTRUE  0.126984  0.126333
##
## t value
## (Intercept) -3.186
## modalityConditionmulti -1.903
## modalityConditionvocal 0.441
## conditionVisual 1.538
## trialTotal -7.792
## I(trialTotal^2) 2.187
## numberOfTurns 11.850
## incorrectTRUE 2.285
## multimodalTRUE 0.915
## modalityConditionmulti:conditionVisual 0.458
## modalityConditionvocal:conditionVisual -3.366
## modalityConditionmulti:trialTotal -1.897
## modalityConditionvocal:trialTotal -0.595
## conditionVisual:trialTotal 0.561
## modalityConditionmulti:I(trialTotal^2) 2.325
## modalityConditionvocal:I(trialTotal^2) 1.952
## modalityConditionmulti:numberOfTurns -2.326
## modalityConditionvocal:numberOfTurns -0.031
## conditionVisual:numberOfTurns -1.247
## modalityConditionmulti:incorrectTRUE 0.981
## modalityConditionvocal:incorrectTRUE -1.314
## conditionVisual:incorrectTRUE -0.288
## conditionVisual:multimodalTRUE -0.672
## modalityConditionmulti:conditionVisual:trialTotal -0.778
## modalityConditionvocal:conditionVisual:trialTotal -0.914
## modalityConditionmulti:conditionVisual:numberOfTurns 1.996
## modalityConditionmulti:conditionVisual:incorrectTRUE 0.164
## modalityConditionvocal:conditionVisual:incorrectTRUE 1.005

##
## Correlation matrix not shown by default, as p = 28 > 12.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it

## fit warnings:
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
## convergence code: 0
## unable to evaluate scaled gradient
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues

feLabelsB = feLabels2
feLabelsB = gsub("Visual modality", "Multimodal", feLabelsB)

x2 = sjp.lmer(fm2, 'fe',
  show.intercept = T,
  sort.est=NULL,
  axis.labels = feLabelsB[2:length(feLabelsB)],
  xlab="Trial time (log ms)",
  geom.colors = c(1,1),
  show.p=F,

```

```

show.values = F,
p.kr = FALSE,
string.interc="Intercept",
prnt.plot = F)

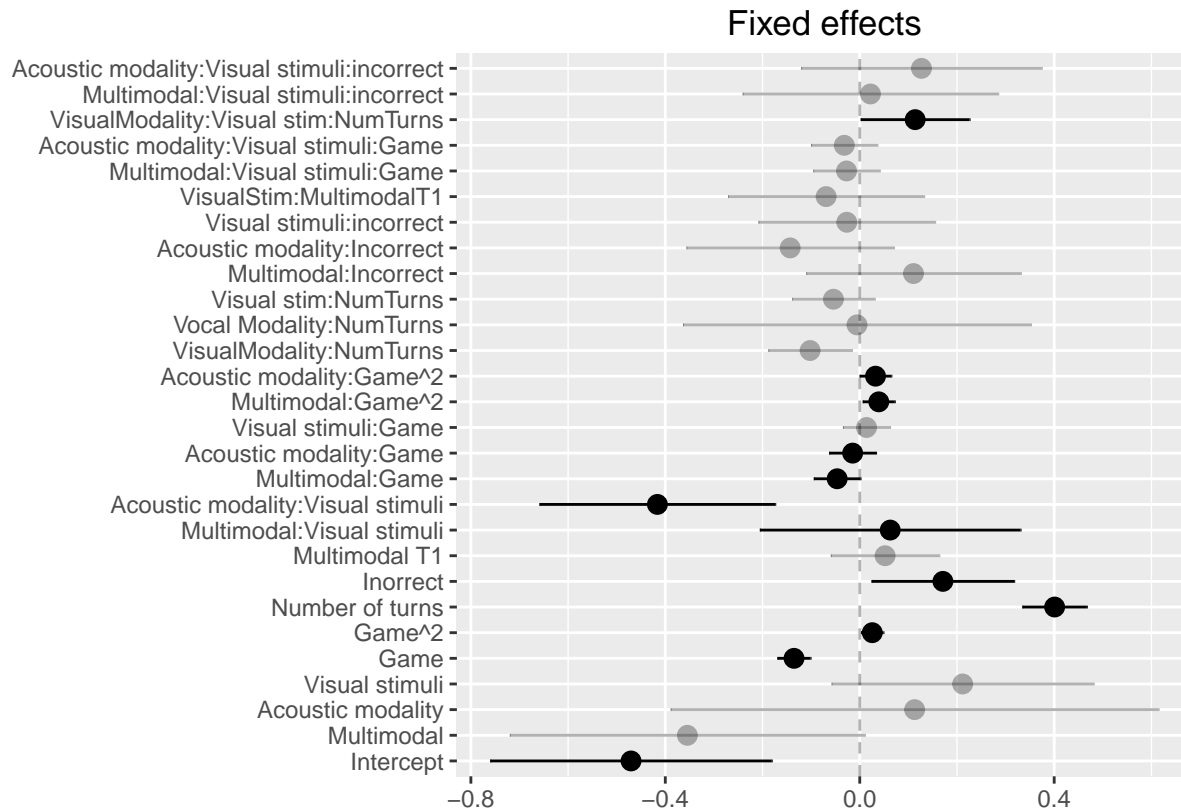
```

Computing p-values via Wald-statistics approximation (treating t as Wald z).

Warning: Deprecated, use `tibble::rownames_to_column()` instead.

```
x2$plot.list[[1]]$data$fade = sig.data$fade
```

```
x2$plot.list[[1]]
```



Analysis of multimodal condition

This script uses data compiled by *analyseData.R*.

Load libraries

```
library(party)
library(lme4)
library(sjPlot)
```

Load data

```
d = read.csv("../data/Final_Turn_data.csv", stringsAsFactors = F)
d = d[d$modalityCondition == "multi",]
```

Prepare variables

```
# Relabel modalities
d[d$turnModalityType=="multi",]$turnModalityType = "M"
d[d$turnModalityType=="unimodal acoustic",]$turnModalityType = "A"
d[d$turnModalityType=="unimodal visual",]$turnModalityType = "V"

# Only need one record per trial
d2 = d[!duplicated(d$trialString),]

# get turn modality type for T1
x = tapply(d[d$turnType=="T1",]$turnModalityType, d[d$turnType=="T1",]$trialString, head, n=1)
d2$turnModality.T1 = x[d2$trialString]

# remove NAs
d2 = d2[!is.na(d2$turnModality.T1),]
# relevel
d2$turnModality.T1 = relevel(factor(as.character(d2$turnModality.T1)), "V")

# get turn modality type for T2
x = tapply(d[d$turnType=="T2",]$turnModalityType, d[d$turnType=="T2",]$trialString, head, n=1)
d2$turnModality.T2 = x[d2$trialString]
d2$turnModality.T2[is.na(d2$turnModality.T2)] = "n"
d2$turnModality.T2 = relevel(factor(d2$turnModality.T2), 'n')

# get turn modality type for T3
x = tapply(d[d$turnType=="T3",]$turnModalityType, d[d$turnType=="T3",]$trialString, head, n=1)
d2$turnModality.T3 = x[d2$trialString]
d2$turnModality.T3[is.na(d2$turnModality.T3)] = "n"
d2$turnModality.T3 = relevel(factor(d2$turnModality.T3), 'n')
```

Make game variable.

```

d2$trialTotal = d2$trial + (d2$game * (max(d2$trial)+1))
# Convert to proportion of games played, so that estimates reflect change per game.
d2$trialTotal = d2$trialTotal / 16
# Center the trialTotal variable so intercept reflects after the first game
d2$trialTotal = d2$trialTotal - 1

d2$incorrect = !d2$correct

```

Scale trial length variable.

```

d2$trialLength.logcenter = log(d2$trialLength)
d2$trialLength.logcenter = d2$trialLength.logcenter - mean(d2$trialLength.logcenter)

```

Simple mixed effects model

```

m0 = lmer(trialLength.logcenter ~
  condition*trialTotal +
  I(trialTotal^2) +
  (1 | dyadNumber) +
  (1 | itemId),
  data = d2)

m1 = lmer(trialLength.logcenter ~
  condition*trialTotal +
  I(trialTotal^2) +
  turnModality.T1 +
  (1 | dyadNumber) +
  (1 | itemId),
  data = d2)

m2 = lmer(trialLength.logcenter ~
  condition*trialTotal +
  I(trialTotal^2) +
  turnModality.T1*condition +
  (1 | dyadNumber) +
  (1 | itemId),
  data = d2)

## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
anova(m0,m1,m2)

## refitting model(s) with ML (instead of REML)

## Data: d2
## Models:
## m0: trialLength.logcenter ~ condition * trialTotal + I(trialTotal^2) +
## m0:      (1 | dyadNumber) + (1 | itemId)
## m1: trialLength.logcenter ~ condition * trialTotal + I(trialTotal^2) +
## m1:      turnModality.T1 + (1 | dyadNumber) + (1 | itemId)
## m2: trialLength.logcenter ~ condition * trialTotal + I(trialTotal^2) +
## m2:      turnModality.T1 * condition + (1 | dyadNumber) + (1 | itemId)
##      Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m0  8 944.90 980.35 -464.45  928.90

```



```
## m1 10 948.08 992.39 -464.04 928.08 0.8212 2 0.6633
## m2 11 950.02 998.77 -464.01 928.02 0.0549 1 0.8147
```

There was no significant main effect of T1 signal modality (log likelihood difference = 0.41 , df = 2 , Chi Squared = 0.82 , p = 0.66).

There was no significant interaction between T1 signal modality and condition (log likelihood difference = 0.027 , df = 1 , Chi Squared = 0.05 , p = 0.81).

```
sjp.lmer(m2, 'fe',
  geom.colors = c(1,1),
  fade.ns = T)
```

```
## Warning: replacing previous import 'lme4::sigma' by 'stats::sigma' when
## loading 'pbkrtest'
```

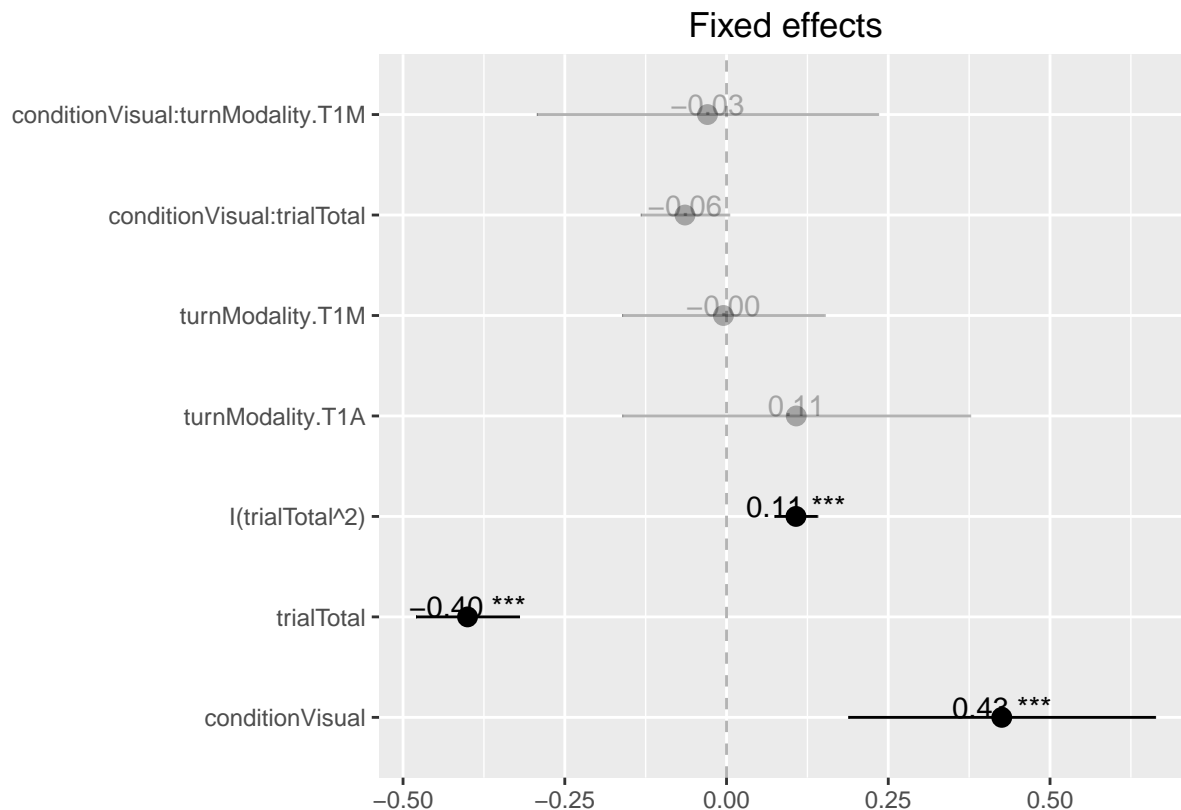
```
## Computing p-values via Kenward-Roger approximation. Use `p.kr = FALSE` if computation takes too long
```

```
## Warning in deviance.merMod(object, ...): deviance() is deprecated for REML
```

```
## fits; use REMLcrit for the REML criterion or deviance(.,REML=FALSE) for
```

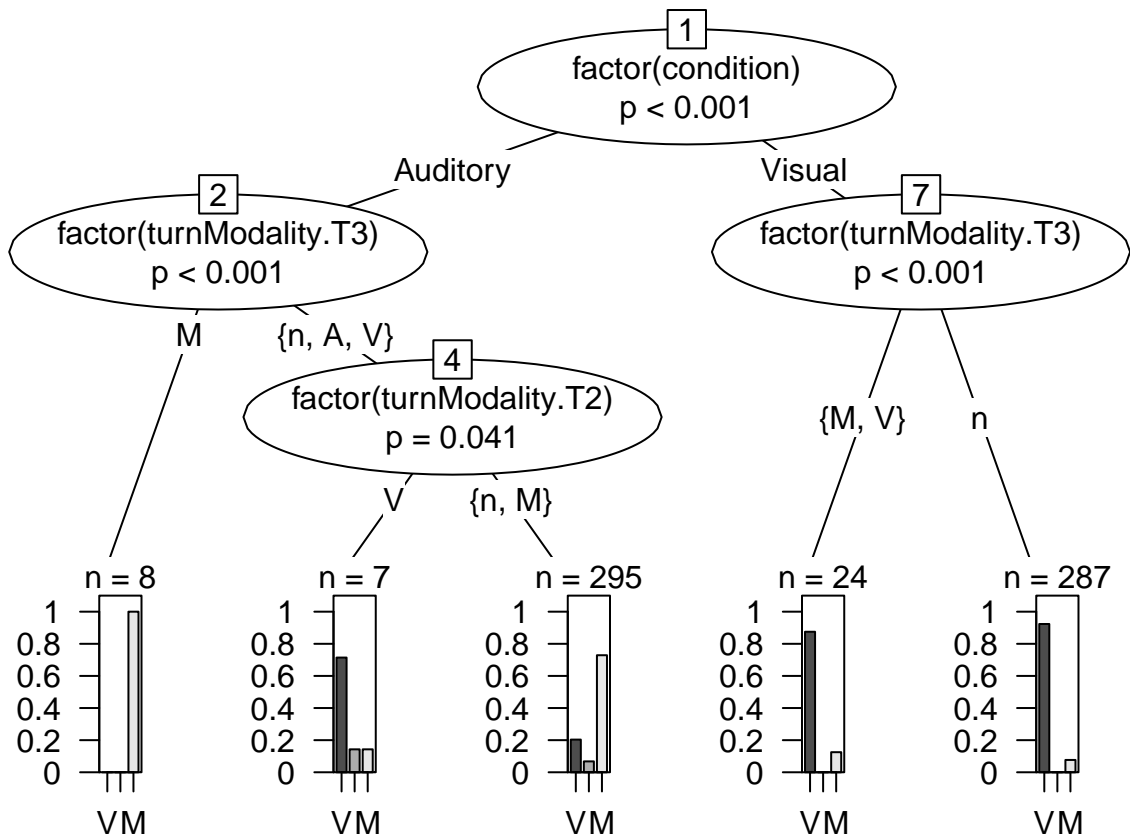
```
## deviance calculated at the REML fit
```

```
## Warning: Deprecated, use tibble::rownames_to_column() instead.
```



Binary trees

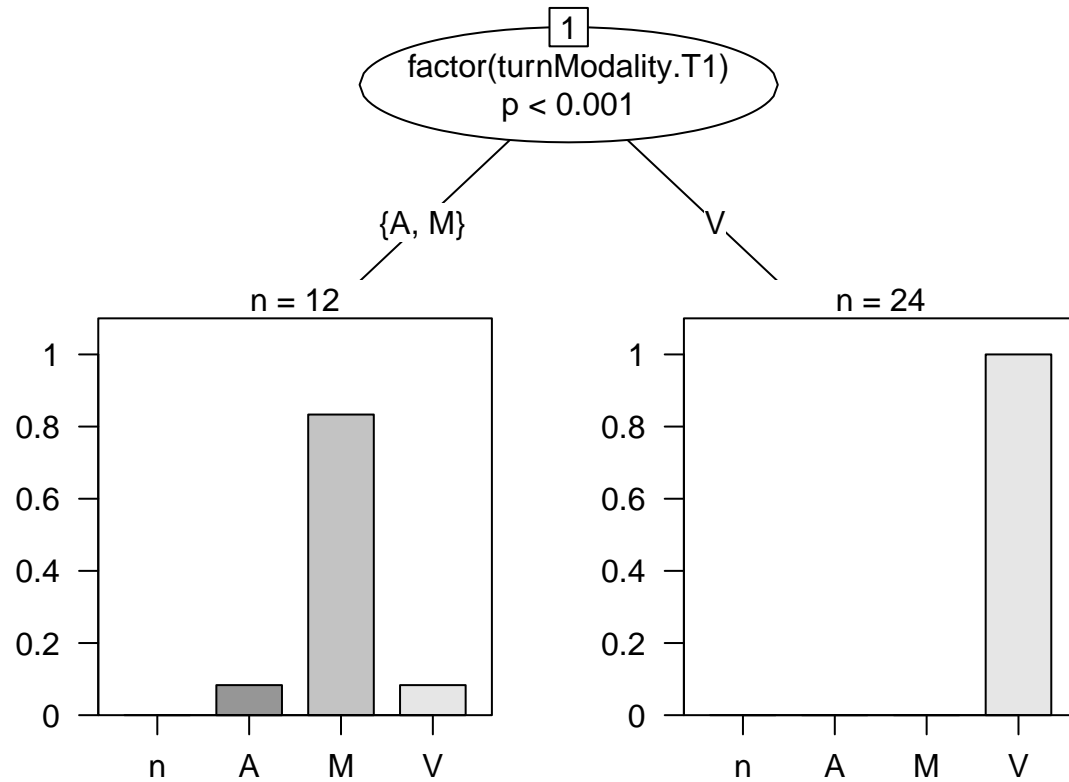
```
cx = ctree(turnModality.T1~
  factor(turnModality.T2)+
  factor(turnModality.T3) +
  factor(condition) +
  game
  , data=d2)
plot(cx, terminal_panel = node_barplot(cx, id=F))
```



```

cxT3 = ctree(turnModality.T3~
  factor(turnModality.T2)+
  factor(turnModality.T1) +
  factor(condition) +
  game
  , data=d2[d2$turnModality.T3!="n",])
plot(cxT3, terminal_panel = node_barplot(cxT3, id=F))

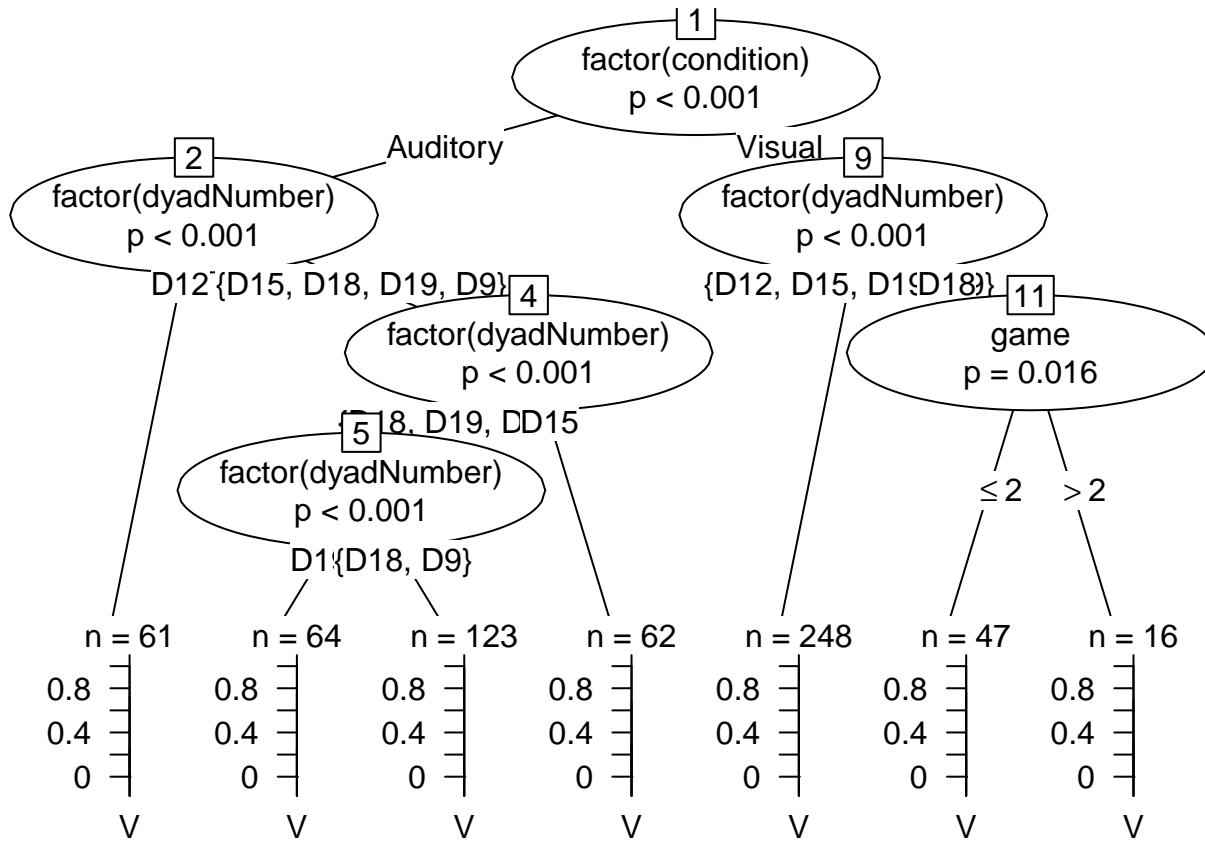
```



```

cx2 = ctree(turnModality.T1~
  factor(turnModality.T2)+
  factor(turnModality.T3) +
  factor(condition) +
  game +
  factor(dyadNumber)
  , data=d2)
plot(cx2, terminal_panel = node_barplot(cx2, id=F))

```



Graphs are also written to results/graphs/cTree/

```

## pdf
## 2

## pdf
## 2

## pdf
## 2

```