Facing Communication Challenges in Distributed Software Development

Keynote for VirtuES 2013 workshop, Bari, Italy

Fabio Calefato
University of Bari, Italy

Collaborative Development Research Group
http://cdg.di.uniba.it
Software development as an intense collaborative process

- Software development as “an exercise in complex interrelationships” *
- Types of collaboration within a team
  - Communication
  - Coordination
  - Control

Communication’s key role in managing distributed SW projects

- Distance has both direct and indirect negative effect on coordination and control
- Communication disruption aggravates coordination and control breakdowns

Direct vs. Indirect communication

Formal vs. Informal communication

<table>
<thead>
<tr>
<th></th>
<th>Formal</th>
<th>Informal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Message</td>
<td>Planned</td>
<td>Spontaneous</td>
</tr>
<tr>
<td>Content</td>
<td>Work-related</td>
<td>Relational</td>
</tr>
<tr>
<td>Purpose</td>
<td>Organizational</td>
<td>Personal</td>
</tr>
<tr>
<td>Location / channels</td>
<td>Official</td>
<td>Random</td>
</tr>
</tbody>
</table>
Communication challenges “faced” in this talk

What

• Direct formal communication
  – Finding best fitting media for communication-intensive tasks
  – Overcoming language barriers with machine translation

• Direct informal communication
  – Establishing personal, trust-based connections in distributed teams

How

• Studies to inform software practices or tool designs
• Mixed research methods:
  – Analysis of software artifacts
  – Questionnaires
  – Interviews
  – Content analysis
FINDING BEST FITTING MEDIA FOR COMMUNICATION-INTENSIVE TASKS

Research partners:
Daniela Damian (UVic), Filippo Lanubile (Univ. of Bari)
Research goal

- To assess the support offered by synchronous text-based communication media (CMC) to distributed groups involved in requirements elicitation and negotiations as compared to traditional face-to-face communication (F2F)
Computer-Mediated Communication (CMC)

- Media can be characterized along three dimensions of information exchange:
  - Time (when)
  - Space (where)
  - Richness (how much)

<table>
<thead>
<tr>
<th></th>
<th>Synchronous (same time)</th>
<th>Asynchronous (different time)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2F</td>
<td></td>
<td>Collocated (same space)</td>
</tr>
<tr>
<td>Videoconference Telephone</td>
<td></td>
<td>Distributed (different space)</td>
</tr>
<tr>
<td>Chat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Billboard</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Media-Richness continuum in the Time/Space Matrix

Main theories on CMC

• Social Presence Theory (Short et al., 1976)
  – Lean single-channel media have low sense of presence (inability to convey non-verbal cues)
  – Lean media better for task-focused communication, rich media for relational communication

• Media Richness Theory (Daft & Lengel, 1984)
  – The more complex the task, the richer the media to use
  – Lean media better for uncertain tasks, rich media better for equivocal tasks

• Common Ground Theory (Clark & Brennan, 1991)
  – Argues that communications is not simply the sending of messages
  – There is no best medium in absolute
Main theories on CMC

• Channel Expansion Theory (Goodhue et al., 1995)
  – Factors other than channel characteristics affect CMC effectiveness
  – Group’s shared experience and media use experience

• Media Synchronicity Theory (Dennis & Valacich, 1998)
  – Highly synch media when convergence is the key process to task accomplishment
  – Lowly synch media when conveyance is the key process to task accomplishment

• Cognitive-Based View (Robert & Dennis, 2005)
  – Sense of presence not as vital as the ability to process information
  – Media Richness Paradox: The richer the media, the harder to properly process information
Task/Technology Fit (TTF)

- Effectiveness of CMC varies on the type of task
  - Differences in tasks and media generate differences in group performance
- Rich media do not always provide the best solution for any given task
  - Too much or too few media richness for a given task represents a poor TTF
- Good TTF only when information richness required by task is proportional to that conveyed by media
  - TTF theories suggest how to appropriately match task characteristics to medium properties
Putting all the pieces together

Rich media
+ High motivation/attention - Low reprocessability
+ High synchronicity - Low parallelism
+ High social presence
+ High comfort
+ Fosters common ground

Lean media
- Low motivation/attention + High reprocessability
- Low synchronicity + High parallelism
- Low social presence
- Low comfort
- Impairs common ground

F2F Negotiation (+)
CMC Negotiation (-)

CMC theories and properties relevant to task/technology fit
Committed
- Commitment
- Reprocessability
- Convergent Thinking

Relational
- Social Presence
- Task-focused
- Convergence
- Conveyance
- Conflictual Tasks
- Divergent Thinking

Equivocal
- Media Richness
- Uncertain

Media Synchronicity
- Media Synchronicity
- F2F Elicitation (+)
- CMC Elicitation (-)

CMC theories and properties relevant to task/technology fit
Cognitive-based View
- Convergent Thinking
- Divergent Thinking
- Cooperative Tasks

Task-focused
- Relational
- Convergent Thinking
- Divergent Thinking
- Cooperative Tasks

Cognitive-based View
- Convergent Thinking
- Divergent Thinking
- Cooperative Tasks

Task-focused
- Relational
- Convergent Thinking
- Divergent Thinking
- Cooperative Tasks

Cognitive-based View
- Convergent Thinking
- Divergent Thinking
- Cooperative Tasks
Empirical study

• Goal
  – Compare F2F and synchronous text-based communication (CMC) modes in distributed requirements elicitations and negotiations

• Setting
  – RE undergraduate course at University of Victoria, Canada
  – 40 students in 6 groups of stakeholders involved in the definition of sw requirements for 6 projects

• Research hypotheses
  – **H1** – F2F requirements workshops are better appreciated (i.e., perceived as more comfortable and satisfying with outcome) than CMC requirements workshops
  – **H2** – CMC Elicitation represents a better task/technology fit (i.e., produce better results) than F2F Elicitation. F2F Negotiation represents a better task/technology fit (i.e., produce better results) than CMC Negotiation
Data

- Gathered through two satisfaction questionnaires and chat logs
- Questionnaires administered at the end of the elicitation and negotiation sessions
eConference

http://code.google.com/p/econference4

Perspectives

Agenda

Message board

Decisions place

Presence

Hand raise
Dependant variables and measures – H1

• Subjects’ responses to satisfaction questionnaires coded to perform quantitative analysis
  – Used 4-point Likert scales and closed questions

• Differences between stakeholders’ perception of requirements workshops conceptualized as:
  – Satisfaction with performance
  – Comfort with communication mode
Dependant variables and measures – H2

- **Requirements workshop and Communication mode** factors created two variants in the rqmt definition process
  - CE - FN
  - FE - CN

- Differences in rqmt workshops effectiveness conceptualized as:
  - Group performance
  - Shared understanding => lack of common ground
Coding schema

- Performed content analysis on the logs of the CMC workshops to collect data on negative evidence and grounding chains constructs
- 9 major categories (thematic unit)

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES-NO</td>
<td>A question that takes a yes/no answer or just a few words (e.g., Q: &quot;How many beds are available in the hospital overall?&quot;, A: &quot;There are 400&quot;). It may initiate a grounding chain.</td>
</tr>
<tr>
<td>QUESTION</td>
<td>Any other question not covered by the YES-NO QUESTION category, such as the wh-questions or those that aggregate multiple questions in a single utterance (e.g., &quot;What information will each employee use to login to the system? Is a key card a consideration?&quot;). It may initiate a grounding chain.</td>
</tr>
<tr>
<td>YES-NO</td>
<td>An answer to a question that takes a few words or a yes-no utterance (e.g., yes, no, yep, nope, y, n, yeah, &quot;yeah, 400&quot;). Normally appears after a YES-NO QUESTION or CHECK. It may close a grounding chain.</td>
</tr>
<tr>
<td>ANSWER</td>
<td>Any other answer to a question not covered by the YES-NO ANSWER category (e.g., Q: &quot;Are you going to do the display public events portion for the final product?&quot;, A: &quot;Yes&quot;). It may close a grounding chain.</td>
</tr>
<tr>
<td>COMPLEX</td>
<td>Any utterance that explicitly looks for confirmation of acceptance through provisional, try-marked statements (e.g., &quot;So we decided for 400 beds, right?&quot;). It is normally followed by an AGREEMENT or an ANSWER.</td>
</tr>
<tr>
<td>PROVISIONAL</td>
<td>Any utterance that explicitly gives confirmation of acceptance through by verbatim copying a previous utterances (e.g., &quot;Backup monthly on a tape&quot;, &quot;Ok, once a month on tape&quot;). It is normally followed by an AGREEMENT.</td>
</tr>
<tr>
<td>CHECK</td>
<td>Any utterance that provides evidence that a previously entered utterance was not accepted (e.g., &quot;I'm not sure I get the question&quot;, &quot;What?&quot;). It may initiate a grounding chain and is normally followed by a TASK or an ANSWER.</td>
</tr>
</tbody>
</table>
Summary of study findings (1/3)

- The role (i.e., being a customer or a developer) has no effect on media selection
  - No significant differences in satisfaction and comfort between CMC and F2F requirements meetings
No Conclusive Evidence that F2F is The Most Preferred Medium for Communication

- F2F provided more opportunity to familiarize with other participants, better ability to express complex ideas and to understand others’ opinions in both elicitations and negotiations
- CMC was more comfortable during both elicitations and negotiations to better participate and openly discuss conflicting issues more openly with group members
- Stakeholders more satisfied with the performance in the F2F negotiations than in the CMC negotiations
- No differences in the perceived satisfaction with performance between F2F and CMC elicitations
Summary of study findings (3/3)

- **Group performance not affected by communication medium**
  - No significant differences in the number of defects found in the final RS docs produced at the end of the process

- **Computer-Mediated Elicitations Offer Support to Achieving Shared Understanding**
  - Data on grounding chains did not allow to draw conclusions about support given by CMC elicitations as compared to the CMC negotiations
  - CMC elicitations had less negative evidence than the CMC negotiations
OVERCOMING LANGUAGE BARRIERS WITH MACHINE TRANSLATION

Research Partners:

Tayana Conte (UFAM), Filippo Lanubile (Univ. of Bari), Rafael Prikladnicki (PUCRS)
Motivation

• Global software projects challenged by language differences
  – Especially requirements meetings

• Machine translation technology for remote meetings in countries with
  – Opportunities for global projects
  – Lack of English speaking professionals

• Example: Brazil
  – Large pool of IT professionals
  – Only 10M speaking English (< 6% of the population)
Research questions

• RQ1: Can MT services be used in distributed multilingual requirements meetings? (instead of English)

• RQ2: How does the adoption of MT affect group interaction? (in distributed multilingual requirements meetings)

• RQ3: Do individuals with a low English proficiency level benefit more than individuals with a high level from MT?
1st study: Simulation


• Simulation
  – Requirements meetings logs as data source

• Assessment of time performance:
  – Delay is negligible

• Assessment of translation quality:
  – Google Translate (corpus-based) more accurate than Apertium (rule-based)
2nd study: Controlled experiment


- 32 participants: 16 students from Bari (IT)
  8 from PUCRS, Porto Alegre (BR)
  8 from Fed. Univ. of Amazonas, Manaus (BR)
- Multilingual groups arranged by proficiency level of English (high vs. low)
Experimental tasks

**T1 – requirements prioritization (30 min.)**

- Customer’s perspective
  1. Assign 16 mobile phone features to 3 piles: *very important, important, less important*
  2. Rank the features within piles

**T2 – release planning (60 min.)**

- Developer’s perspective
  1. Distribute 1000 story points to each feature as an estimate of implementation costs
  2. Plan 3 releases based on priorities (T1) and cost estimates
## Experimental design*

### Original experiment (high proficiency)

<table>
<thead>
<tr>
<th>MT</th>
<th>EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gr1, Gr3 execute T1</td>
<td>Gr2, Gr4 execute T1</td>
</tr>
</tbody>
</table>

### Replicated experiment (low proficiency)

<table>
<thead>
<tr>
<th>MT</th>
<th>EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gr6, Gr8 execute T1</td>
<td>Gr5, Gr7 execute T1</td>
</tr>
</tbody>
</table>

### Run 2

<table>
<thead>
<tr>
<th>MT</th>
<th>EN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gr5, Gr7 execute T2</td>
<td>Gr6, Gr8 execute T2</td>
</tr>
</tbody>
</table>

Data sources:
- post-task questionnaires
- meeting logs

*We have doubled the # of groups to 16, new data still to be analyzed*
eConferenceMT

http://code.google.com/p/econference-mt-plugin
Findings

• **RQ1 - Can MT services be used in distributed multilingual requirements meetings?**
  – Machine translation is not disruptive of the conversation flow and is accepted with favor

• **RQ2 - How does the adoption of MT affect group interaction?**
  – More balanced discussions when using native language with MT

• **RQ3 - Do individuals with a low English proficiency level benefit more than individuals with a high level from MT?**
  – So far NO, although people with low English skills are more prone to use MT again
However…

- Messaging is easier than talking for a non-native English speaker
- Therefore, as future work we should:
  - Replicate with voice conferences
  - Compare with groups including native English speakers
  - Replicate with more distant languages couples (e.g., Chinese, Russian)
ESTABLISHING PERSONAL TRUST-BASED CONNECTIONS IN DISTRIBUTED TEAMS

Research Partner:
Filippo Lanubile (Univ. of Bari)
Challenges of distributed software development

<table>
<thead>
<tr>
<th></th>
<th>Geographical Distance</th>
<th>Temporal Distance</th>
<th>Sociocultural Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Communication</strong></td>
<td>Decreased frequency of communication</td>
<td>Delay in responses</td>
<td>Language differences and misunderstandings</td>
</tr>
<tr>
<td></td>
<td>Lack of informal exchange</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Providing right technical infrastructure</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Coordination</strong></td>
<td>Lack of awareness</td>
<td>Reduced hours for same time collaboration</td>
<td>Doubtful of others’ capabilities</td>
</tr>
<tr>
<td></td>
<td>Reduced trust</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>Push for heavy-weight processes</td>
<td>Management of project artifacts subject to delays</td>
<td>Jobs perceived as under threat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Different perceptions of authority</td>
</tr>
</tbody>
</table>

Adapted from: Ågerfalk, P.J., and Fitzgerald, B. Flexible and Distributed Software Processes: Old Petunias in New Bowls?, CACM, 49(10), 2006
Group awareness

“An understanding of the activities of others which provides a context for your own activity” *

- Informal awareness
- Group-structural awareness
- Workspace awareness
- Social awareness

* Dourish, P. and Bellotti, V. Awareness and coordination in shared workspaces. Proc. CSCW '92.
How to increase awareness & build trust in distributed teams?

- Meeting Room
- Remote Conferencing
- Informal Communication
- Social Networks
How can social software mitigate the negative effects of distance in globally distributed development?

- Current approach: Have all team members use a single, project-oriented, social networking site
- Problems
  - Project-oriented networking sites do not capture the full social identity of a software engineer
  - On the other hand, nobody would like to get project news feeds into the personal timelines of Facebook or Twitter
  - In large global teams, you might do not know people you should be aware of (awareness network)
  - The awareness network can be very dynamic*

Is social awareness needed?

• Previous research
  – Bradner, E., Mark, G. Why distance matters: effects on cooperation, persuasion and deception. CSCW '02, New Orleans, USA, 2002
  – Shami, N.S. et al. Making sense of strangers' expertise from signals in digital artifacts. CHI '09, Boston, USA
Theoretical Model

**INTEGRITY**
The adherence to intrinsic moral norms which makes a trustee reliable

**BENEVOLENCE**
The perceived level of courtesy and positive attitude

**ABILITY**
Capability of a trustee (based on knowledge, competence, skills) to perform tasks within a specific domain

**PREDICTABILITY**
The degree to which a person is liable and accountable and meets the expectation of another person

**Cognitive Trust**
Trustee’s antecedents to trust

**Affective Trust**
Trustor’s antecedent to trust

**PROPENSITY TO TRUST**
A general, not experience-based inclination to display faith and adopt a trusting attitude toward others
So, is social awareness needed???

YES!

We are in the **social era** of software engineering where **social media ecosystems** are an integral component of software repositories, and vice versa

– M.A. Storey, MSR’12 Keynote
Our approach

• Information shared on social media can surrogate the social awareness on which affective trust grows
  – Developers’ personal content from social media aggregated and made it contextually available into the shared workspace

• Great impact of social software on collaborative development environments
  – How to provide social awareness?

• Challenge for empirical studies
  – How can we measure the actual benefit of being personally connected to other developers, wherever they are located?
The SocialCDE Project

http://socialtfs.codeplex.com
https://github.com/collab-uniba/socialcde4eclipse
## Group awareness in Collaborative Development Environments

<table>
<thead>
<tr>
<th></th>
<th>Informal awareness</th>
<th>Group-structural awareness</th>
<th>Workspace awareness</th>
<th>Social awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trac</td>
<td>—</td>
<td>YES</td>
<td>YES</td>
<td>—</td>
</tr>
<tr>
<td>Google Code</td>
<td>—</td>
<td>YES</td>
<td>YES</td>
<td>—</td>
</tr>
<tr>
<td>Assembla</td>
<td>—</td>
<td>YES</td>
<td>YES</td>
<td>—</td>
</tr>
<tr>
<td>Jazz</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>partial (IBM Connections)</td>
</tr>
<tr>
<td>TFS</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>—</td>
</tr>
<tr>
<td>CodePlex</td>
<td>—</td>
<td>YES</td>
<td>YES</td>
<td>—</td>
</tr>
<tr>
<td>GitHub</td>
<td>—</td>
<td>YES</td>
<td>YES</td>
<td>partial (view developers’ connections and @mentions)</td>
</tr>
</tbody>
</table>

Research Model & Hypotheses

• $H_1$ - There is a positive relationship between the amount of social awareness gained through social media and the level of affective trust mutually established among distant teams

• $H_2$ - There is a positive relationship between the level of affective trust mutually established among distant teams and project performance
CONCLUDING REMARKS
Conclusions

Finding best fitting media for communication-intensive tasks

• Lean media can be effectively used instead of F2F for communication-intensive tasks
• Tasks mostly involving idea generation (divergent thinking)

Overcoming language barriers with machine translation

• State-of-the-art MT solutions still far from perfect
• Can be used to complete communication-intensive tasks with some delay due to mistranslations

Establishing personal, trust-based connections in distributed teams

• The actual benefit of being personally connected to other developers still under investigation
Facing communication issues takeaway

- Communication, coordination and control are all equally important to collaboration within (virtual) teams

YET

- Communication should be first among equals!
  - Avoiding communication disruption and breakdowns is paramount
  - Any effort put to resolve coordination and control issues is ineffective otherwise
Thanks for your attention!

Questions & comments

Contacts:
• fabio.calefato@uniba.it
• http://cdg.di.uniba.it/calefato
• https://www.researchgate.net/profile/Fabio_Calefato
References


