

ピアレビュー有効時間比率計測による ピアレビュー会議の改善と品質改善の効果

久野 倫義^{†1}中島 毅^{†1}松下 誠^{†2}井上 克郎^{†2}

本論文は、ピアレビュー会議の改善に関するものである。ピアレビュー会議はソフトウェア開発における品質向上の重要な活動であり、ピアレビューの中心的な活動である。ピアレビュー会議は品質向上の活動として定着しているが、ピアレビュー会議を完了したソフトウェアにおいても、欠陥が残存している場合も多い。そこでピアレビュー会議測定ツールを用いて、ピアレビュー会議の課題を定量的に明確化した。課題を解決するため、ピアレビュー有効時間比率という指標を用い、ピアレビュー会議が欠陥抽出を中心とした活動になるように改善を行った。その結果ピアレビュー有効時間比率の平均が9.64%から26.4%に向上し、単位時間当たりの指摘件数が1.7件から2.1件へ改善した。さらにテスト段階へ流出する欠陥数が減少し、製品品質を向上できた。

An Improvement of Peer-Review-meetings Using the Peer-Review-Effectiveness Ratio and Effect of Quality Improvement

Noriyoshi Kuno^{†1}, Tsuyoshi Nakajima^{†1}, Makoto Matsushita^{†2}, Katsuro Inoue^{†2}

Abstract

In this paper, we show an improvement for peer-review-meetings which are center of the peer-reviews which have an important role for software quality improvement. In software developments, the peer-review-meetings are common activities. After peer-review-meetings, there are some bugs which should be removed in peers-review-meetings. So we made peer-review-meetings' problems clear quantitatively with a measurement tool. And we improved peer-review-meetings which focus on remove bugs by using a new metrics named peer-review-effectiveness ratio. As a result, peer-review-effectiveness ratio rise up from 9.64% to 26.4% and the number of bugs which removed in a peer-review meeting per unit time are increased from 1.7 to 2.1. Moreover, we show the number of bugs which removed in test is decreased.

1. はじめに

ソフトウェアの欠陥が引き起こすシステム障害の社会的な影響が増大する中、高品質なソフトウェアを開発するための検証手法及びそれらを使った品質管理方法を確立することが求められている。高品質なソフトウェアを最終検査工程だけで達成することは困難であり、開発各工程で確実に欠陥を除去していき後工程に流出させないことが必要で

ある [Kan2002][中島 2008].

要求分析からコーディングに至る上流工程では、設計文書やプログラムを対象としたピアレビューが主たる検証手段である [Gilb1993] [織田 2006].

【脚注】

- †1 三菱電機株式会社 設計システム技術センター
- †2 大阪大学 大学院情報科学研究科

ピアレビューは人手により実施するため参加者個人の技量に大きく依存し、その効果にバラツキが現れやすい[森崎2009]。検証手法として、このバラツキを軽減することを目的に、観点やチェックリストを用いる方法[野中2004]やプロセスを重視し組織力を活用する方法[細川2009]などが提案・評価されている。

Gilbはソフトウェアインスペクションを体系化し、その中心的な活動として欠陥抽出を行うピアレビュー会議(ソフトウェアインスペクションではロギングミーティングと呼ぶ)を定義した[Gilb1993]。

ピアレビュー会議に対する問題点として、ピアレビュー会議時間を短縮し、ピアレビュー会議におけるアイドル時間を削減する必要があることや、ピアレビュー会議がプロジェクト遅れを引き起こしており、ピアレビュー会議は不要であると主張している[Johnson1998][Glass1999]。

一方、レビュー会議の質を上げるため、レビュー会議の参加人数の適正化とファシリテーションの有効性を評価する取り組み、ピアレビュー速度/指摘密度/レビュー効率という指標でピアレビューを分析する取り組みなどがある[小室2005][中野2006]。

上記のいずれの研究においても、ピアレビュー会議において具体的にどのような活動を行っているかを定量的に評価していない。開発現場では、一言でピアレビュー会議と言っても千差万別であり、その中身を把握しなければ、その改善は困難である。そこでピアレビュー会議における読上げ、指摘、議論などの活動の内訳を定量化し、それらの活動の偏りを正し、企業活動で重要である、限られた時間内で効率と品質を改善することが可能となる。

本論文では、まずピアレビューの中心的活動であるピアレビュー会議を定量的に評価し、ピアレビュー会議の実施方法を改善することで、ピアレビュー会議をその目的である欠陥抽出活動に変更し製品品質を改善できることを示す。2節において、ピアレビュー会議の問題点を明確化する。次に、3節においては、レビュー会議を定量的に評価する手法を提案し、4節では提案手法の実プロジェクトへの適用結果を示しその有効性を示す。5節ではピアレビュー会議改善による製品品質向上の効果を示す。6節では関連研究について述べ、従来研究と本提案技法との関係を明確にする。

2. 従来研究と解決すべき課題

2.1. ピアレビュー会議

ソフトウェアインスペクションは、図1に示すように、計画策定、キックオフ、個人チェック、ピアレビュー会議(ソフトウェアインスペクションでは、ロギングミーティングと呼ぶ)、編集及びフォローアップで構成される[Gilb1993]。

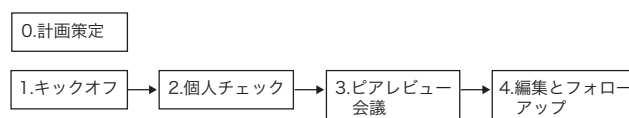


図1 ソフトウェアインスペクションの流れ

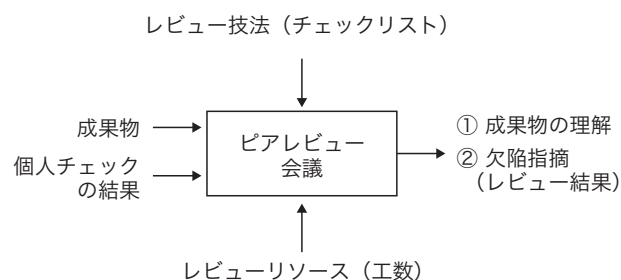


図2 ピアレビュー会議プロセス

インスペクションは品質向上の重要な活動と位置づけられ、多くの研究対象となっている。その中でも、中心的役割を担うピアレビュー会議に対する従来研究を述べ、本研究で解決すべき課題を明確にする。

Gilbは、ピアレビュー会議の目的を個人チェックにおいて検出した欠陥を報告し、会議の場で新たに発見した欠陥と共に記録することであると定義した[Gilb1993]。その目的を達成する為、ピアレビュー会議においては、資料内容の読上げや修正の提案、個人で検出した欠陥に対する議論を行わず、欠陥の抽出に特化することで、ピアレビューを効果的な活動にできると報告している。図2は、ピアレビュー会議の課題を明確化する為、Gilbの報告をプロセスとして整理したものである。ピアレビュー会議では、成果物と個人チェックの結果をインプットとし、レビュー技法とリソースを使い、成果物を理解し指摘を行う。

2.2. ピアレビュー会議の課題

2.1で示したようにピアレビュー会議は、欠陥の記録と会議で新たな欠陥を発見することが目的であり、その他の活動を行わないように制御する必要がある。ピアレビュー会議において、単に個人チェックにおいて検出された欠陥のみを報告し、会議時に追加欠陥が発見されなければ、ピアレビュー会議を行う必要はない。

Johnsonは、インスペクションをソフトウェア品質改善のための、ただ1つの重要な手法と述べている。しかし、企業におけるソフトウェアインスペクションの採用率は低く、その原因の1つがピアレビュー会議に工数が多くかかること、アイドル時間があることを挙げている[Johnson1998]。

Glassは、複数人による個人チェックが品質向上のためには十分で有り、ピアレビュー会議は不要であると主張している。さらにピアレビュー会議開催がプロジェクトの進捗

を遅らせていると報告している [Glass1999].

上記報告ではピアレビュー会議は不要であると主張しているが、個人の視点で抽出した欠陥から他者が新たな欠陥を抽出できるピアレビュー会議は欠陥を流出させないという点で最も重要な活動であり、開発の現場では必ず実施すべき活動である。

2.3. 本研究が扱う課題

ソフトウェア開発の現場において、ピアレビュー会議が品質向上の中心的活動として定着してきている。ソフトウェア開発の各フェーズにおいてピアレビュー会議を行い、欠陥抽出を行っている [森崎 2009]。しかし、ピアレビュー会議を完了しても、ピアレビュー不足やピアレビューで検出すべき欠陥がテスト段階に流出する欠陥が残存する場合がある。

2.2 節で示した課題や上記の開発現場における課題など様々な問題点は報告されているが、種々な要因がからみあっており、ピアレビューの欠陥抽出件数/ピアレビュー全体時間などの評価指標では直接ピアレビュー会議の問題を把握することはできない。そこで実際にピアレビュー会議において何が行われているかを把握し、それをどのように改善し、結果をどのように評価するかを明確化することが必要である。

3. ピアレビュー会議の改善手法

前節で述べたようにピアレビューの問題点を把握する為、ピアレビュー会議で実際に何が行われているかを測定する手法を明確化する。

3.1. ピアレビュー会議の定義

一般には、ピアレビュー会議がどのように行われたかを測定することは難しい。レビュー会議時間であれば、会議開始時間と終了時間から算出できるが、レビュー会議では、読上げや単なる質問に要する時間もあり、単純な会議時間とレビューに要した時間は異なる。そこで、まず表 1 のようにピアレビュー会議で行われる活動を整理した。

表 1 ピアレビュー会議で行われる活動

発言内容分類	内容
開始宣言	目的や欠陥指摘件数目標の説明
内容読上げ	ピアレビュー対象の作業成果物の説明
指摘	作業成果物に対する欠陥指摘
議論	指摘に対する議論や成果物以外の議論
修正案	指摘に対する修正案の検討
意図の質問	指摘に対する、その意図の確認
無発言	発言のない時間

表 1 の定義は、Gilb がピアレビュー会議で想定した活動(「開始宣言」「指摘」「意図の質問」と実際にソフトウェア開発の現場で行われると想定される発言から導いた。

測定者はピアレビュー会議に出席し、図 3 に示すピアレビュー会議測定ツールを用いて測定を行う。本ツールは、発言内容を測定するためのボタンと発言者ごとの発言時間を測定するボタンで構成される。測定者がピアレビュー会議に出席し、出席者の発言内容を確認し表 1 の活動に該当するボタンを押下する。その際に発言者に該当するボタンも押下することでピアレビュー出席者ごとの発言時間を記録する。

3.2. 調査結果と課題

ピアレビュー会議測定ツールを用いて、12 種の製品を開発する 12 部門における 31 回のピアレビュー会議を定量的に測定した結果一覧を表 2 に示す。今回の調査対象は、ドキュメントに対するレビューを対象とし、コードレビューは対象としていない。

表 2 ピアレビュー会議時間測定結果

会議 No	開始宣言	内容読上	指摘 (TF*)	議論	修正案	意図の質問	無発言	部門名
1	1.4%	22.4%	6.7%	56.2%	7.0%	4.3%	2.1%	A
2	0.3%	15.4%	9.5%	46.1%	16.9%	11.6%	0.3%	A
3	0.5%	11.5%	5.6%	69.8%	7.3%	0.6%	4.7%	A
4	2.9%	25.6%	10.2%	36.4%	14.2%	1.3%	9.5%	A
5	0.0%	86.1%	8.3%	1.5%	0.0%	0.0%	4.1%	B
6	0.1%	75.2%	12.3%	10.3%	0.8%	0.0%	1.4%	B
7	1.0%	34.0%	5.0%	47.0%	0.0%	6.0%	7.0%	C
8	0.2%	22.5%	13.9%	43.5%	0.4%	0.0%	19.7%	C
9	0.4%	20.9%	25.4%	49.9%	0.2%	1.8%	1.4%	C
10	0.1%	20.6%	15.3%	21.7%	35.4%	4.0%	2.8%	C
11	0.1%	10.3%	15.4%	63.0%	6.2%	4.4%	0.5%	C
12	1.0%	74.0%	5.0%	13.0%	2.0%	0.0%	5.0%	D
13	1.0%	62.0%	7.0%	24.0%	4.0%	0.0%	2.0%	D
14	0.4%	61.9%	9.0%	11.7%	12.7%	0.4%	4.0%	E
15	0.5%	68.4%	2.4%	0.6%	0.0%	4.7%	23.4%	E
16	0.0%	14.0%	14.0%	48.0%	4.0%	0.0%	20.0%	F
17	2.0%	37.7%	17.7%	23.3%	3.9%	9.7%	5.8%	G
18	0.0%	38.2%	14.3%	31.6%	1.6%	2.6%	11.7%	G
19	0.0%	9.9%	27.8%	52.3%	2.3%	4.0%	3.6%	G
20	0.8%	40.7%	15.9%	37.4%	3.6%	1.0%	0.7%	H
21	0.1%	41.2%	7.8%	24.9%	1.8%	8.1%	16.0%	H
22	0.5%	23.2%	26.3%	37.9%	6.5%	2.2%	3.3%	H
23	2.3%	13.4%	22.9%	39.3%	8.5%	7.6%	6.2%	I
24	1.8%	35.3%	0.7%	39.2%	6.6%	11.5%	4.8%	J
25	4.4%	59.5%	1.1%	18.8%	2.8%	13.1%	0.4%	J
26	1.0%	8.3%	20.0%	47.4%	0.0%	22.3%	1.0%	J
27	0.8%	21.1%	8.2%	44.2%	0.1%	2.0%	23.7%	K
28	0.0%	25.4%	4.7%	39.4%	2.1%	2.6%	25.7%	K
29	0.2%	7.9%	8.3%	28.5%	6.6%	9.1%	39.5%	L
30	0.2%	55.0%	6.7%	18.3%	3.4%	7.6%	8.8%	L
31	0.1%	29.3%	5.8%	24.9%	11.8%	5.6%	22.5%	L

※ TFについては、3.3 項に述べる。

発言内容 **C** の発言時間です。

B：開始宣言, C：読上げ, D：指摘, E：指摘に対する議論, F：修正案, G：意図の質問, H：無発言時間

開始宣言	内容読上げ	指摘	議論	修正案	意図の質問	無発言
------	-------	----	----	-----	-------	-----

発言内容のボタンを押すと、その前に押されたボタンの発言内容時間が終了します。

会議完了

発言者 **A** さんの発言中です。

A	B	C	D	E	F	G	H	I	J	K	L	M	N	発言なし
---	---	---	---	---	---	---	---	---	---	---	---	---	---	------

発言者のボタンを押すと、その前に押されたボタンの人の発言時間が終了します。

図3 ピアレビュー会議測定ツール

会議 No13

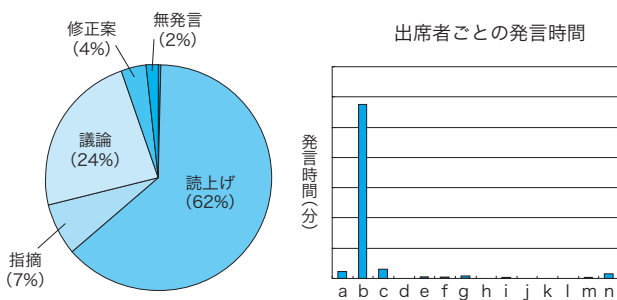


図4 仕様説明を中心としたピアレビュー会議結果

会議 No11

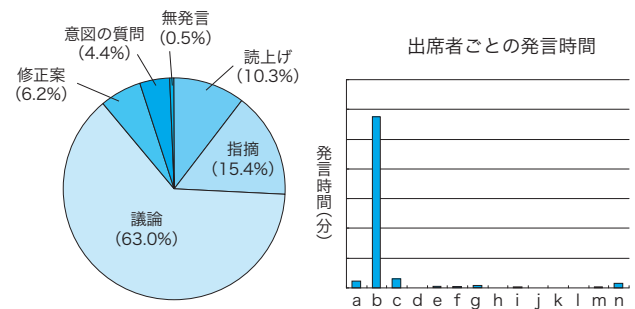


図5 設計活動を中心としたピアレビュー会議結果

表2より、部門ごとに差はあるが、以下の課題がある。

- i) ピアレビュー会議という同一の名称であっても内容は様々である。
- ii) 成果物を読上げる時間が30%を超えるものが半数近くある。
- iii) 無発言時間が10%を超えるものが3割ある。
- iv) 指摘時間が10%未満のものが半数を超える。

図4から図6は、測定結果をグラフ化したものであり、円グラフは表1の発言内容、棒グラフは出席者ごとの発言時間である。図4は、表2の会議No.13の測定結果である。図4の円グラフからはピアレビュー対象の作業成果物を説明する読上げ時間が、会議の60%以上を占めていること、棒グラフからは発言者が1名に集中していることが分かる。それぞれのピアレビュー会議の出席者が14名であり、そのうち1名のみが資料を読上げていることから、本ピアレビュー会議が実質的には仕様説明会であることを示している。

図5は、表2の会議No.11の測定結果である。図5の円グラフからは、欠陥指摘に対する議論や修正案の検討時間が長く、欠陥指摘時間は10%程度であり欠陥指摘活動というより設計自体を行っている。出席者の発言時間については、特定の出席者に偏っていないことから出席者の選定には問題がないことが判る。

会議 No19

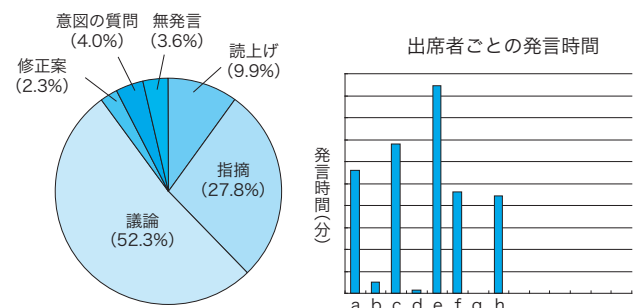


図6 欠陥抽出を中心としたピアレビュー会議結果

図6は、表2の会議No.19の測定結果である。図6の円グラフからは、議論の時間も長い欠陥抽出時間比率が25%を超えており欠陥抽出を中心としたピアレビュー会議であることが判る。出席者の発言時間については、特定の出席者に偏っていないことから出席者の選定には問題がないが、議論時間が長く欠陥抽出を十分できていない可能性がある。

3.3. ピアレビュー会議プロセス定義と有効指摘率を用いたピアレビュー会議改善

3.2項で示したように、本来は欠陥抽出を意図しているピアレビュー会議が説明会や設計活動になっており、ピアレビュー会議を本来の欠陥抽出に特化した活動とするために

会議 No2

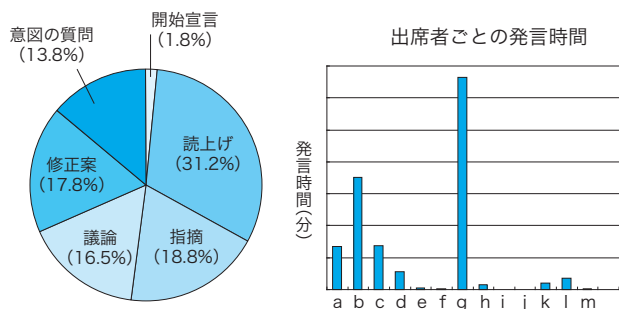


図7 部門Aのピアレビュー会議測定結果 (改善後)

は、内容の読上げ、修正案の検討、無発言を少なくする必要があります。まず、図4などの定量的に把握したピアレビュー会議の現状を開発者に説明を実施し、目指すべき欠陥抽出を中心とした活動とすべきことを合意した。その後、Gilbが定めたピアレビュー会議プロセスをベースにソフトウェア開発の現場に適合するピアレビュー会議プロセスを以下の通り定義した。

- (1) モデレータ、書記、読上げ者（できれば作成者以外）を決める。
- (2) レビュー会議の目的を文書化し、全員が見える場所に示す。
- (3) 目的に合致した参加者を決定する。
- (4) 事前査読時間と指摘件数を全員が報告し、事前査読が不十分なら延期を検討する。
- (5) 会議内容を記録（指摘内容だけではなく、発言で気になる点を記録する）。
- (6) 全員が目的を意識し、不要な議論（前提を基にした議論等）を排除する。
- (7) 全員が指摘をするように順番に指摘を促す。
- (8) モデレータは、発言内容、動作を観察し、納得していないようなら発言を促す。以下に注意する。
 - ① 資料自体の読上げをしていないか。
 - ② 指摘に対しその場で回答をしようとして議論になっていないか。
 - ③ 「以前に説明したように」という発言がある場合、今までも口頭で説明し記録されていないと判断し、発言を文書化するように促す。
 - ④ 必要であれば議論の内容を仕様書に記載することを促す。
 - ⑤ 用語の解釈が出席者間で異なっていないか。
- (9) 目的が達成できたかを、ピアレビュー会議終了時に確認する。

会議 No3

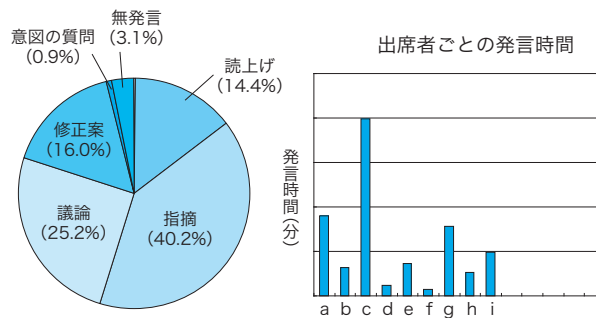


図8 部門Hのピアレビュー会議測定結果 (改善後)

- (10) 指摘をピアレビュー会議後に確認し、記録誤り、抜けがないことを確認する。

上記で定義したプロセスに基づくピアレビュー会議を実施し、ピアレビュー会議における発言内容ごとの時間を測定した。ピアレビュー会議が欠陥抽出を中心とした活動となることを推進する為、ピアレビュー有効時間比率（以降 TF と記述）と呼ぶ指標を導入した。 TF は、ピアレビュー会議測定ツールの「指摘ボタン」を押していた時間の割合であり、以下の式で表わされる。

$$TF = \sum (Ti) / T \quad (式 1)$$

ここで、

Ti: レビュー会議参加者 i が指摘を行った時間

i: レビュー会議参加者, T: 総レビュー会議時間

式1で示すように、 TF は、各レビュー参加者がレビュー会議中に指摘を行った時間の総和を総レビュー時間で割ったものである。 TF を高めることで、ピアレビュー会議を欠陥抽出中心とした活動とすることができる。なお、今回測定したピアレビュー会議時間は、過去のピアレビュー会議のデータから1頁にかけた平均時間により決定しているため、 TF を向上することは、目標ピアレビュー会議時間内で欠陥抽出の効果を最大化することである。ただし、目標ピアレビュー会議時間の設定がピアレビューの効果に影響を与える可能性があり、本来は流出欠陥が基準を越えたケースをはずれ値として除き、統計的に目標を決定することが望ましい。

4. 適用と評価

4.1. 適用

3.3節で示した改善プロセスを適用し、その適用結果を測定できた4部門における8回の会議結果を表3に示す。表3から改善前のように TF が1桁のピアレビュー会議はなくなっておりピアレビューが欠陥抽出活動に変化したことが分かる。

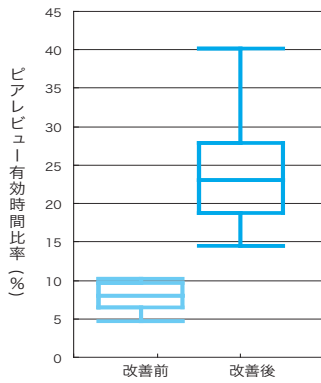


図9 改善前後のTFの箱ひげ

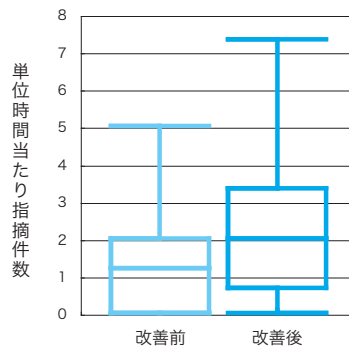


図10 改善前後の単位時間単位指摘件数の箱ひげ

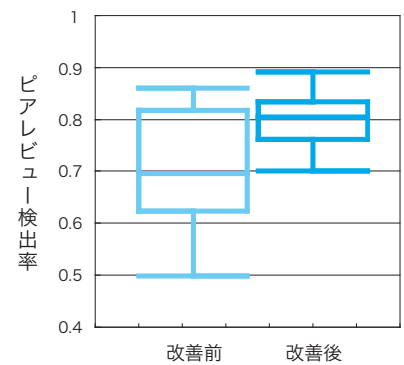


図11 全検出欠陥に占めるレビュー指摘欠陥割合(式2)の変化

表3 ピアレビュー会議測定結果(プロセス定義後)

会議No	開始宣言	内容読上	指摘(TF)	議論	修正案	意図の質問	無発言	部門名
1	1.5%	53.6%	23.5%	9.6%	6.9%	1.5%	3.4%	A
2	1.8%	31.2%	18.8%	16.5%	17.8%	13.8%	0.0%	A
3	0.2%	14.4%	40.2%	25.2%	16.0%	0.9%	3.1%	H
4	0.1%	9.9%	22.7%	47.1%	4.1%	3.4%	12.8%	H
5	0.8%	27.8%	18.5%	45.0%	4.1%	0.0%	3.8%	H
6	1.0%	2.0%	49.0%	35.0%	4.0%	7.0%	2.0%	K
7	1.0%	12.3%	23.7%	29.0%	17.9%	3.2%	12.8%	L
8	0.6%	40.4%	14.5%	15.5%	12.1%	6.0%	10.9%	L

改善後の測定結果において、TFが比較的低い会議とTFが比較的高い会議の測定結果を示す。図7は、表3の会議No.2の測定結果である。TFは18.8%であり、改善前に比べ向上したが、読上げ、議論、修正案、意図の質問に対する時間比率も多く、13名の参加者を集めた説明会、欠陥抽出活動、設計の混在した会議であった。

図8は、表3の会議No.3の測定結果である。9名の参加者でピアレビュー会議を実施し、同一製品における改善前の比率と比較して、指摘時間比率の向上、読上げ時間比率の減少など改善効果があった。

これらの結果から、プロセス定義に基づきピアレビュー会議を実施し、その結果を見る化し開発者と課題を共有することで、ピアレビューをGilbの示したピアレビュー会議のあるべき姿に近づけることができると言える。

4.2. 評価結果

4.1項の適用結果からピアレビュー会議におけるTFは平均9.64%から26.36%に改善した。本項では、改善前と改善後のTFの変化について評価し、その有効性を判定する。まず改善前後の2群の分散が同一であることを帰無仮説として分散比を検定した。表4からP値は0.022であり、有意水準5%で分散が同一であるという仮説は却下された。そこで、分散が異なるとして、2つの平均に差がないことを帰無仮説として検定を行った。表5からP値が0.003と

0.005であり、またt境界値よりも算出されたtの絶対値が大きく、有意水準5%で平均に差がないという仮説は却下され、平均値は統計的に有意な差を持つと結論できた。

表4 改善前後の分散比の検定

	平均	分散	観測数	自由度	観測された分散比	P(F<=f)片側	F境界値片側
改善前	9.64	36.1	12	11	3.927	0.022	3.012
改善後	26.36	141.77	8	7			

表5 改善前後の平均値の差の検定

	平均	分散	観測数	自由度	t	P(T<=t)片側	t境界値片側	P(T<=t)両側	t境界値両側
改善前	9.64	36.1	12	9	-3.673	0.003	1.833	0.005	2.262
改善後	26.36	141.77	8						

図9は、改善前と改善後のTFの測定結果を箱ひげ図を用いて示したものであり、改善前後のTFに差異があることが判る。なお、ひげの両端は最大最小値を示している。

5. レビュー会議改善による品質改善に対する評価

5.1. 単位時間当たりの指摘数の評価

これまで示した改善により、ピアレビュー会議を欠陥抽出を中心とした会議に変更できたことを示した。さらに、ピアレビュー会議における単位時間当たりの指摘件数は、平均1.7件から2.1件に向上した。この違いを評価する為、まず2つの分散が同一であることを帰無仮説として分散比を検定した。表6は、改善前後の単位時間当たりの指摘件数の分散比の検定結果である。表6から、P値は0.0001となり有意水準5%で分散が同一であるという仮説は却下された。そこで、分散が異なるとして、2つの平均に差がないことを帰無仮説として検定を行った。表7からP値が0.00026と0.0005であり、有意水準5%で平均に差がないという仮説は却下され、改善前後の平均値が統計的に有意な差を持つと結論できた。なお、指摘件数に関しては欠陥

種別により分類し、上位仕様書に要求事項として記載されていない改善事項（例えば保守性に関する指摘）や体裁に関する欠陥などはカウントしていない。

表 6 改善前後の単位時間当たりの指摘件数分散比の検定

	平均	分散	観測数	自由度	観測された分散比	P(F<=f)片側	F境界値片側
改善前	1.666	2.419	213	212	1.548	0.0001	1.213
改善後	2.136	3.744	557	556			

表 7 改善前後の単位時間当たりの指摘件数平均値の差の検定

	平均	分散	観測数	自由度	t	P(T<=t)片側	t境界値片側	P(T<=t)両側
改善前	1.666	2.419	213	474	3.496	0.00026	1.648075	0.000517
改善後	2.136	3.744	557					

図 10 は、改善前と改善後の単位時間当たりの指摘件数の測定結果を箱ひげ図を用いて示したものであり、改善前後の単位時間当たりの指摘件数に差異があることが判る。

これまで説明してきたように、ピアレビュー会議プロセスを測定、プロセスの問題を定量的に把握し、目標となる指標を決定した上で改善活動とその効果を測定することでプロセスの問題を解決できることを示した。

開発の現場においては、ピアレビュー工数とピアレビュー指摘件数に目標値を設けピアレビュープロセスの統一を図っている。今回の改善により、目標ピアレビュー工数におけるピアレビュー指摘件数を向上することができ、テストフェーズに流出する欠陥数を抑制することができた。

5.2. ピアレビュー改善による製品品質向上の評価

ソフトウェア開発において、上流工程で品質を作り込むことが重要である。そのためには、設計書様式を決定し欠陥混入を防止すると共に、ピアレビューにより欠陥を検出する。さらにピアレビューで検出できない欠陥をテストで検出する。しかしテスト段階で全ての欠陥を検出することは困難であり、設計段階で欠陥混入を防止することが製品品質を向上する方法である。図 11 は、ピアレビュー改善を実施した複数のプロジェクトにおけるソフトウェアライフサイクル全体で検出する欠陥数のうち、ピアレビューの欠陥数の比率を示したものである。

ピアレビュー検出率 =

$$\text{ピアレビュー検出数} / \text{ライフサイクルの欠陥数} \quad (\text{式 2})$$

改善前後において、ピアレビューで検出した欠陥数が増加しており、テストで検出する欠陥比率が減少していることが判る。

上記改善前後のピアレビュー検出率についても、検定を行った。まず 2 群の分散が同一であることを帰無仮説とし

て分散比を検定した。表 8 から P 値は 0.032 であり、有意水準 5% で分散が同一であるという仮説は却下された。そこで、分散が異なるとして、2 つの平均に差がないことを帰無仮説として検定を行った。表 9 から P 値が 0.012 と 0.025 であり、有意水準 5% で平均に差がないという仮説は却下され、改善前後のピアレビュー検出率の平均値は統計的に有意な差を持つと結論できた。

表 8 改善前後のピアレビュー検出率の分散比の検定

	平均	分散	観測数	自由度	観測された分散比	P(F<=f)片側	F境界値片側
改善前	0.703	0.0141	14	13	2.911	0.032	2.577
改善後	0.791	0.0048	14	13			

表 9 改善前後のピアレビュー検出率平均値の差の検定

	平均	分散	観測数	自由度	t	P(T<=t)片側	t境界値片側	P(T<=t)両側
改善前	0.703	0.0141	14	21	-2.41	0.01272	1.720743	0.02543
改善後	0.791	0.0048	14					

5.3. ピアレビュー参加者による主観的評価及び改善活動の展開

ピアレビュー改善を実施した開発現場から、以下の意見がありピアレビュー改善が有効に機能したと評価できた。

- (1) ピアレビューの問題点が定量的にはっきりし改善が進んだ。
 - (2) ピアレビューを実施することが目的となりがちであったが、ピアレビューが本来の目的である欠陥抽出に有効な活動となった。
- 一方、以下のような課題も上げられた。
- (3) 定着化が課題である。改善に関わったメンバ以外を巻き込むには教育が必要。
 - (4) なぜ、議論をしてはいけないのか。有識者が集まれる時間は少ない。

上記のような課題に対しては、開発現場の実態を考慮し 3 段階で改善を進める方法や欠陥抽出と修正案検討の場を分離するなどの提案を行った [久野 2009]。その結果をガイドラインとしてまとめ、ソフトウェア開発のベストプラクティスとして全社ソフトウェア開発を行う事業所から利用できるようにした。またソフトウェア開発のプロジェクトリーダーに対する教育において今回の改善の講義を行い、ピアレビュー改善を全社的に展開している。

6. 関連研究

ピアレビューの結果から品質を評価するために定量データを用いて評価する手法として、以下のような手法が提案されている。

中野らは、500プロジェクトのコードレビューデータから、レビュー効率(ライン数/レビュー工数)に対してレビュー指摘密度(指摘数/ライン数)が大きい場合は、テスト段階で欠陥が多く検出される傾向を示した[13]。しかし、レビュープロセスについては、全プロジェクトで統一されており、プロジェクトごとにプロセスの変化がないことを前提としており、具体的にレビュー会議においてどのような活動を実施したかは把握していない。

「定量的品質予測のススメ」では、レビュー工数を式3で定義した[SEC2008]。

$$\text{レビュー工数} = \sum \text{各レビューアのレビュー実施時間} \quad (\text{式} 3)$$

その際に、有識者以外(育成等を目的とした要員)のレビューアの工数は、レビュー工数から除外するなど、有識者以外のレビュー参加者の工数を適切な係数で補正することが望ましいとあるが、具体的な係数については言及されていない。またレビュープロセスの評価フローとして、レビュー工数密度の評価と適切さを評価することになっているが、無発言時間などを含むレビュー工数全体を用いて評価するため、その適切さ自体に誤りが入る可能性がある。

野中はインスペクションの定量的管理に用いる指標として欠陥指摘工数密度を定義した[野中2009]。

$$\text{欠陥指摘工数密度} = \frac{\sum NFi}{\sum Ti} \quad (\text{式} 4)$$

NFiは欠陥数、Tiは総レビュー時間であり、iはレビュー参加者を示す。

式4における各値の測定方法が一貫していること、すなわち開発プロセスが標準化され安定していることが必要であるとされている。

効率的なレビューやレビュープロセス改善手段としては、レビュー会議では欠陥の抽出に集中することや、ピアレビュー会議の工数をレビュープロセスの評価メトリクスとして用いることを提案している[飯山2008][安達2006]。

またRobbinsは、レビューアがドキュメントやレビュー用シナリオを読む時間などの相対時間を分析し、ドキュメントを読む時間が全体の29.64%であることを報告しているが、29.64%であることの有効性や課題については言及していない[Robbins2009]。

上記の各研究においては、各種指標の測定方法を一貫性のあるものにすることや、開発プロセスを標準化することの重要性は述べているが、具体的な方法については示されていない。本文で示した方法により、ピアレビュー会議を欠陥抽出中心の活動とすることで、ピアレビュー工数の精度を向上でき、ピアレビュー評価手法を改善できると考える。

7. おわりに

本論文では、ピアレビューの中心的活動であるピアレビュー会議を定量的に評価し、ピアレビュー会議の実施方法を改善することで、ピアレビュー会議をその目的である欠陥抽出活動に改善できることを示した。

実際にピアレビュー会議の改善を行うのは、そのプロセスを定義するだけではなく、実際にどのような活動を行っているかを定量的に測定し、開発者へフィードバックし、制御することが重要であることが分かった。さらにピアレビュー会議が欠陥抽出活動ではなく単なる説明会であったり、設計活動であったりした場合、ピアレビュー会議時間をピアレビュー工数として品質評価や品質予測を適切に行うことはできないことを示した。

今後は、ピアレビュー有効時間比率とテスト段階における欠陥数のデータを蓄積し、両値の相関を分析することで、ピアレビュー有効時間比率を品質判断の基準値として用いる上での適値範囲を決定していく予定である。また今回の改善はピアレビュー会議に焦点を絞った活動であるが、本手法は様々な会議の改善に用いることが可能である。

【参考文献】

- [Kan2002] S. H. Kan: Metrics and Models in Software Quality Engineering, Addison-Wesley, 2002.
- [中島2008] 中島毅, 東基衛: ソフトウェア開発における品質プロセスのコスト最適化のためのモデルとシミュレーションツール, 電子情報通信学会論文誌, Vol.J91-D, No.5, pp.1216-1230, 2008.
- [織田2006] 織田巖: ソフトウェア・レビュー技術, ソフトウェア・リサーチ・センター, 2006.
- [Gilb1993] T. Gilb and D. Graham: Software Inspection, Addison-Wesley, 1993.
- [森崎2009] 森崎修司: ソフトウェアインスペクションの動向, 情報処理, Vol.50, No.5, pp.377-380, 2009.
- [野中2004] 野中誠: 設計・ソースコードを対象とした個人レビュー手法の比較実験, 情報処理学会研究報告, SE-140-4, Vol.2004, No.118, pp.25-31, 2004.
- [細川2009] 細川宣啓: 第三者インスペクションによる品質検査と欠陥予測, 情報処理, Vol.50, No.5, pp.405-411, 2009.
- [Johnson1998] P. M. Johnson: Reengineering Inspection, Communications of the ACM, Vol.41, No.2, pp.49-52, 1998.
- [Glass1999] R. L. Glass: Inspection - Some Surprising Findings, Communications of the ACM, Vol.42, No.4, pp.17-19, 1999.
- [小室2005] 小室睦他: 開発現場の実態に基づいたピアレビュー手法の改善と改善効果の定量的分析, SEC journal, Vol.1, No.4, pp.6-15, 2005.
- [中野2006] 中野裕也, 水野修, 菊野亨, 阿南佳之, 田中又治: コードレビューの密度と効率がコード品質に与える影響, SEC journal, Vol.2, No.4, pp.10-17, 2006.
- [久野2009] 久野倫義, 丹羽友光, 前川隆昭: デザインレビューの効果的実施及び評価方法, 三菱電機技法, Vol.83, No.5, pp.18, 2009.
- [SEC2008] 情報処理推進機構ソフトウェア・エンジニアリングセンター: 定量的品質予測のススメ, オーム社, 2008.
- [野中2009] 野中誠: ソフトウェアインスペクションの効果と効率, 情報処理, Vol.50, No.5, pp.385-390, 2009.
- [飯山2008] 飯山俊介: 設計レビュー指標値の算出, 第28回ソフトウェア品質シンポジウム, 2008.
- [安達2006] 安達賢二: レビュープロセスの現実的な改善手段の提案, ソフトウェアテストシンポジウム, 2006.
- [Robbins2009] B. Robbins: Cognitive Factors in Perspective-Based Reading (PBR): A Protocol Analysis Study, Third International Symposium on Empirical Software Engineering and Measurement, pp.145-155, 2009.