

ピアレビュー手法に基づく  
ソフトウェア品質の改善に関する研究

提出先 大阪大学大学院情報科学研究科

提出年月 2015年4月

久野 倫義



# 研究業績一覧

## 主要論文

- [1-1] 久野倫義, 中島毅, ピアレビュー網羅率を用いた品質評価技法の提案, 情報処理学会論文誌, Vol. 53, No. 2, pp. 622-630, 2012.
- [1-2] 久野倫義, 中島毅, 松下誠, 井上克郎, 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, SEC journal, Vol. 10, No. 1, pp. 16-23, 2014.
- [1-3] 久野倫義, 中島毅, 松下誠, 井上克郎, レビュー会議の有効性評価に関する一考察, 電子情報通信学会論文誌, 2015 (採録決定).
- [1-4] N. Kuno, T. Nakajima, M. Matsushita, K. Inoue, A Study on the Effectiveness of Peer Review Meeting, Proceedings of The 24<sup>th</sup> IEEE International Symposium on Software Reliability Engineering (ISSRE), Pasadena, CA, pp. 7-8, 2013.
- [1-5] N. Kuno, T. Niwa, T. Maekawa, The Effective Method for Design Reviews, Proceedings of 4<sup>th</sup> World Congress for Software Quality, Bethesda, Maryland, 2008.
- [1-6] 久野倫義, 丹羽友光, 前川隆昭, デザインレビューの効果的な実施方法, 第26回ソフトウェア品質シンポジウム, 2006.

## 関連論文

- [2-1] 藤岡卓, 田村直樹, 中島毅, 久野倫義, 真野哲也, ソフトウェア技術者エントリ層に対する職能教育コースの設計と実装, 公益社団法人日本工学教育協会 工学教育第 62 巻第 1 号, pp. 46-51, 2014.
- [2-2] 久野倫義, 丹羽友光, 前川隆昭, デザインレビューの効果的实施及び評価方法, 三電技法, Vol. 83, No. 5, p. 18, 2009.



## 内容梗概

ソフトウェア開発作業は、ハードウェア開発を含む製品の開発過程において、その後半部分に集中して行われることが多い。例えば、組込みシステム開発であれば、ソフトウェアに対する要求がハードウェア試作後に明確になることもある。さらに、ハードウェアからソフトウェアに対する制約の変化や開発後半における要求仕様変更などにより、実質的なソフトウェア開発期間の短縮が発生し、より効率的な機能実現と品質確保の取組が必須となっている。

このような開発期間の短縮、要求仕様の変更、品質の確保という 3 つの課題を解決するために、企業では様々な取組みを実施している。ソフトウェア品質を確保する取組としては、設計ドキュメント様式の充実、設計プロセスの定義、検証と妥当性確認の取り組みなど様々な改善活動を実施している。その中で、ピアレビューは、仕様・設計ドキュメントに混入した欠陥の除去だけでなく、仕様・設計ルールの不備、様式の不備を検出するなど多岐にわたり有用であり、開発における最重要活動の 1 つである。このような最重要活動の 1 つであることを認識されながら、ピアレビューに対する開発現場の取り組みは、他の要求定義・設計・テストに比べ十分ではない。もちろん、ピアレビューやソフトウェアインスペクションは、これまで様々な論文や文献で紹介されてきた。開発の現場では、一般的なピアレビュー手法を参考に、製品特性や組織風土などに合致するようにプロセスを定義しているが、適切にピアレビュー手法を定義するという事は難しく、適切な活動になっていないことが多い。

本論文では、ピアレビューという活動において、実開発の現場でピアレビューが真に有効な活動となっているか、その有効性を高めることができるかという視点から、以下の 3 つの課題について、これまで提案されていない測定手法や新しい指標を用いて定量的に課題を見える化し、その解決策を示した。

- ・課題 1：ピアレビュー後に不具合が流出することに対する効果的な活動がない
- ・課題 2：ピアレビュー会議の有効性について疑問があり決着がついていない
- ・課題 3：品質評価技法が不十分である

課題 1 に対しては、ピアレビュー後に不具合が流出する原因について調査し、その一つの要因がピアレビュー会議の実施方法に問題があると仮定し、定量的に測定した。その結果、ピアレビュー会議を不具合抽出に特化するプロセスを定義し、不具合流出を防止できることを示した。

課題 2 に対しては、ピアレビュー会議の有効性を評価する実験を行い、これまでの研究と異なるデータを示し、開発現場でピアレビュー会議が有効であることを示した。

課題 3 に対しては、これまでのピアレビュー評価技法であるゾーン分析における

課題を明確化し、その解決策としてピアレビュー網羅率という新たな指標を定義し、開発現場で適用し、その有効性を示した。

これらの 3 つの課題に対する解決策によって、これまで開発の現場で、その効果と適用の効率性が不十分であったピアレビューを真に有効な活動とし、ソフトウェア品質と開発効率を向上することに貢献できた。

# 謝辞

本研究の全般に関し、常日頃より適切なご指導を賜りました、大阪大学大学院情報科学研究科コンピュータサイエンス専攻 井上克郎教授に心から感謝申し上げます。

本研究を行うに当たり、具体的なご指導とご助言を賜りました、大阪大学大学院情報科学研究科コンピュータサイエンス専攻 松下誠准教授に厚く御礼申し上げます。

本論文を執筆するに当たり有益なご教示、ご助言を賜りました大阪大学大学院情報科学研究科 増澤利光教授、楠本真二教授に深く感謝致します。

本研究の全般において、多大なるご協力と適切なご助言を賜りました、三菱電機（株）設計システム技術センター 中島毅主任技師長に深く感謝申し上げます。

本研究を大阪大学大学院情報科学研究科で進める機会を与えて頂いた設計システム技術センターソフトウェアエンジニアリング部 前川隆昭元部長、本研究のきっかけを与えて頂いた三菱電機（株）名古屋製作所 丹羽友光氏に感謝申し上げます。また業務上の配慮を頂いた設計システム技術センター 中岡邦夫センター長、ソフトウェアエンジニアリング部 村松良晃部長、山下昭裕元部長、大野俊樹元部長に感謝申し上げます。

本研究は、筆者が設計システム技術センターにおいて行ったソフトウェア改善活動に基づいています。ソフトウェア改善活動を進める中で、ソフトウェア開発現場の皆様や設計システム技術センターソフトウェアエンジニアリング部の皆様からの貴重なご意見、ご指導により適用評価を進めることで、実りある成果をあげることができました。心から感謝申し上げます。





# 目次

|       |                                    |    |
|-------|------------------------------------|----|
| 1.1   | 高信頼化ソフトウェア開発におけるピアレビューに求められる役割     | 1  |
| 1.2   | ピアレビュー                             | 3  |
| 1.2.1 | 計画策定                               | 3  |
| 1.2.2 | キックオフ                              | 4  |
| 1.2.3 | 個人チェック                             | 4  |
| 1.2.4 | ピアレビュー会議                           | 5  |
| 1.2.5 | 編集とフォローアップ                         | 6  |
| 1.2.6 | ピアレビュー改善                           | 6  |
| 1.3   | ピアレビューの研究                          | 8  |
| 1.3.1 | ピアレビューの全般的な研究                      | 8  |
| 1.3.2 | これまでのピアレビュー管理技法の研究                 | 10 |
| 1.3.3 | これまでのピアレビュー実施技法の研究                 | 11 |
| 1.4   | ピアレビューの課題                          | 12 |
| 1.5   | 本論文の概要                             | 13 |
| 第2章   | ピアレビュー会議の改善に関する研究                  | 15 |
| 2.1   | ピアレビューにおけるピアレビュー会議の位置づけ            | 15 |
| 2.2   | 従来研究と解決すべき課題                       | 15 |
| 2.2.1 | ピアレビュー会議の課題                        | 16 |
| 2.2.2 | 本研究が扱う課題                           | 16 |
| 2.3   | ピアレビュー有効時間比率を用いたピアレビュー会議の改善        | 16 |
| 2.3.1 | ピアレビュー会議の定着                        | 17 |
| 2.3.2 | 調査結果と課題                            | 18 |
| 2.3.3 | 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 | 22 |
| 2.4   | 適用と評価                              | 24 |
| 2.4.1 | 適用                                 | 24 |
| 2.4.2 | 評価結果                               | 25 |
| 2.5   | レビュー会議改善による品質改善に対する評価              | 27 |
| 2.5.1 | 単位時間当たりの指摘数の評価                     | 27 |
| 2.5.2 | ピアレビュー改善による製品品質向上の評価               | 28 |
| 2.5.3 | ピアレビュー参加者による主観的評価および改善活動の展開        | 29 |
| 2.6   | 関連研究                               | 29 |

|         |                               |    |
|---------|-------------------------------|----|
| 2.7     | まとめと今後の課題                     | 30 |
| 第3章     | レビュー会議の有効性評価に関する一考察           | 33 |
| 3.1     | まえがき                          | 33 |
| 3.2     | レビュー会議に関する有効性評価               | 33 |
| 3.2.1   | 実験条件                          | 33 |
| 3.2.2   | Porterの実験結果                   | 34 |
| 3.2.3   | 今回の実験結果                       | 34 |
| 3.2.4   | 2つの実験結果の統計的評価                 | 35 |
| 3.2.5   | 2つの実験結果の相違に関する考察              | 35 |
| 3.3     | まとめと今後の課題                     | 38 |
| 第4章     | ピアレビュー網羅率を用いた品質評価技法に関する研究     | 39 |
| 4.1     | まえがき                          | 39 |
| 4.2     | 従来研究と解決すべき課題                  | 40 |
| 4.2.1   | ゾーン分析法                        | 40 |
| 4.2.2   | ゾーン分析法の課題                     | 41 |
| 4.2.3   | 本研究が扱う課題                      | 42 |
| 4.3     | ピアレビュー評価手法を用いた品質評価技法の提案       | 43 |
| 4.3.1   | 事前調査                          | 43 |
| 4.3.1.1 | レビューでの欠陥検出パターン                | 43 |
| 4.3.2   | 作業成果物への混入欠陥の分布                | 45 |
| 4.4     | 提案技法                          | 46 |
| 4.4.1   | ピアレビュー網羅率                     | 46 |
| 4.4.2   | ピアレビュー網羅率の分割                  | 47 |
| 4.4.3   | ゾーン分析とピアレビュー網羅率を組合せた品質評価技法    | 47 |
| 4.5     | 適用と評価                         | 48 |
| 4.5.1   | ピアレビュー網羅率の評価                  | 48 |
| 4.5.2   | 分割したピアレビュー網羅率の評価              | 49 |
| 4.5.3   | ゾーン分析とピアレビュー網羅率を組合せた品質評価技法の評価 | 50 |
| 4.6     | 関連研究                          | 52 |
| 4.6.1   | 欠陥数推定法とレビュー評価への適用上の特徴と制約      | 52 |
| 4.6.2   | 欠陥数推移法と提案技法との関係               | 55 |
| 4.7     | まとめと今後の課題                     | 55 |
| 第5章     | むすび                           | 57 |
| 5.1     | まとめ                           | 57 |
| 5.2     | 今後の研究方針                       | 59 |

|        |                       |    |
|--------|-----------------------|----|
| 付録 1   | ピアレビュー定義.....         | 67 |
| 付録 1.1 | ピアレビューの種類とその目的.....   | 67 |
| 付録 1.2 | ピアレビューの主要アクティビティ..... | 68 |
| 付録 1.3 | ピアレビューに必要な作業成果物.....  | 68 |
| 付録 1.4 | ピアレビュー枠組み.....        | 68 |
| 付録 1.5 | 自己チェックリスト作成方法例.....   | 69 |
| 付録 1.6 | チェックリストの一般的な使用方法..... | 71 |
| 付録 1.7 | ロギングミーティングの運営方法.....  | 71 |
| 付録 1.8 | 留意すべき発言とその解釈例.....    | 72 |
| 付録 1.9 | レビューの評価.....          | 72 |

## 図目次

|      |                              |    |
|------|------------------------------|----|
| 図 1  | ソフトウェア開発ライフサイクルモデル（V字開発モデル）  | 2  |
| 図 2  | ピアレビューの流れ                    | 3  |
| 図 3  | ピアレビュー会議プロセス                 | 6  |
| 図 4  | ピアレビュー改善の樹形図                 | 10 |
| 図 5  | ピアレビュー会議測定ツール                | 18 |
| 図 6  | 仕様説明を中心としたピアレビュー会議測定結果       | 20 |
| 図 7  | 設計活動を中心としたピアレビュー会議測定結果       | 21 |
| 図 8  | 欠陥抽出を中心としたピアレビュー会議測定結果       | 22 |
| 図 9  | 部門Aのピアレビュー会議測定結果（改善後）        | 24 |
| 図 10 | 部門Hのピアレビュー会議測定結果（改善後）        | 25 |
| 図 11 | 改善前後のピアレビュー有効時間比率の箱ひげ図       | 26 |
| 図 12 | 改善前後の単位時間単位指摘件数の箱ひげ図         | 28 |
| 図 13 | 全検出欠陥に占めるレビュー指摘欠陥割合の変化       | 28 |
| 図 14 | ゾーン分析法                       | 40 |
| 図 15 | 従来のゾーン分析手法の適用実例              | 42 |
| 図 16 | ページ番号と指摘件数の関係（流出欠陥数の多かった機能）  | 44 |
| 図 17 | ページ番号と指摘件数の関係（流出欠陥数の少なかった機能） | 45 |
| 図 18 | レビュー指摘の成果物                   | 46 |
| 図 19 | ピアレビュー網羅率を用いたゾーン分析           | 48 |
| 図 20 | 総ページ数とレビュー網羅率( $D$ )の関係      | 49 |
| 図 21 | 総ページ数と後半レビュー網羅率( $D_b$ )の関係  | 50 |
| 図 22 | ピアレビュー網羅率（後半）付ゾーン分析結果        | 51 |
| 図 23 | 欠陥プロファイル推定法                  | 54 |
| 図 24 | ピアレビューの枠組み                   | 69 |
| 図 25 | 個人用チェックリスト作成フロー概要            | 70 |
| 図 26 | 欠陥型例                         | 70 |

## 表目次

|      |                                       |    |
|------|---------------------------------------|----|
| 表 1  | ピアレビューチームメンバーの役割 .....                | 3  |
| 表 2  | チェッカ役割の例.....                         | 4  |
| 表 3  | 個人チェック技法の例 .....                      | 5  |
| 表 4  | 基本測定量 .....                           | 7  |
| 表 5  | 導出測定量 .....                           | 7  |
| 表 6  | インスペクションメトリクス.....                    | 9  |
| 表 7  | ピアレビュー会議で行われる活動 .....                 | 17 |
| 表 8  | ピアレビュー会議時間測定結果 .....                  | 18 |
| 表 9  | ピアレビュー会議時間測定結果（プロセス定義後） .....         | 24 |
| 表 10 | 改善前後の分散比の検定 .....                     | 26 |
| 表 11 | 改善前後の平均値の差の検定 .....                   | 26 |
| 表 12 | 改善前後の単位時間当たりの指摘件数分散比の検定 .....         | 27 |
| 表 13 | 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 .....       | 27 |
| 表 14 | 実験の条件 .....                           | 33 |
| 表 15 | レビュー形態 .....                          | 34 |
| 表 16 | 今回の実験結果 .....                         | 35 |
| 表 17 | t 検定の結果.....                          | 35 |
| 表 18 | ゾーンの意味 .....                          | 41 |
| 表 19 | 過去の成果物に対するレビュー網羅率 .....               | 48 |
| 表 20 | 成果物 1, 2, 3, 4, 5 に対するピアレビュー網羅率 ..... | 49 |
| 表 21 | レビューの種類 .....                         | 67 |
| 表 22 | ピアレビューの主要アクティビティ .....                | 68 |



# 第 1 章 はじめに

## 1.1 高信頼化ソフトウェア開発におけるピアレビューに求められる役割

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、高信頼なソフトウェアに関する問題が議論され、高信頼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[1-1]。情報システムを取り巻く環境の変化として、近年の情報技術の進展により企業活動の重要なインフラとしての機能のみならず、一般国民の生活に直結する重要なインフラとしての役割も担ってきている。さらに、ネットワーク技術の進展により情報が広範囲に、かつリアルタイムに授受され、情報システム間に関わる相互接続性の多様性および障害時の影響範囲が単独システムから複合システム間へ拡大し、高信頼化ソフトウェアに対する要求はさらに高まっている[1-1]。上記の高品質なソフトウェアを開発するには、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[1-2][1-3]。そのためにはソフトウェア開発ライフサイクルを明確化し、工程毎に品質を確保しながら開発を進める必要がある。例えば、ISO/IEC12207 においては、ソフトウェア開発ライフサイクルの標準を定め各工程で実施すべき活動を定義している。

一般的に V 字開発モデル（図 1）と呼ばれるソフトウェア開発ライフサイクルでは、要求定義工程と設計工程を V 字の左側に、検証工程を V 字の右側に配置している。V 字開発モデルでは、V 字の左側が最上位の要求を段階的に詳細化する流れを示し、V 字の右側が、詳細化したソフトウェアを段階的に統合しながら検証する流れを示している。さらに V 字各工程の左右の関係で設計と検証の対応関係を示しており、V 字の左側で定義した設計が意図した通りに動作するかを該当する V 字の右側工程で検証するモデルである。検証段階では、要求定義書や設計書に記載された事項が正しく実現できているかという視点で検証を行うため、要求定義や設計における“暗黙の了解”、“行間”と呼ばれる文書化されていない事項を検証段階で発見することは難しい。このような“暗黙の了解”や“行間”に伴う欠陥を検証する手段として、“記載されていないことを指摘できる”ピアレビューが有効である。ピアレビューなくして欠陥のない製品を作ることは困難であり、開発ライフサイクルにかかわらず、次の作業を実施する前に、各主要成果物が通過すべき品質の関所であると言われている[1-4]。

フロントローディング[1-1]は、要求定義段階や設計の初期段階で多くの欠陥を検出し、検証段階へ流出する欠陥数を減らすことで工程遅延や市場への不具合流出を防

止するという活動である。フロントローディングを推進するには、仕様書様式を整備するなどの欠陥混入防止とピアレビューによる欠陥流出防止活動が必要である。さらに、実開発においては要求と設計間の一貫性欠如という欠陥が多くある。要求と設計間の一貫性の検証手段としては、やはりピアレビューが主たる検証手段である。

上記のように、ピアレビューは高信頼化ソフトウェアを開発するための品質向上、工程遵守、コスト削減にとって重要な活動であり、ソフトウェア開発にとって欠かすことのできない活動である。1976年にIBMのMichael Faganによって開発されたソフトウェアインスペクション技法[1-5]の登場から、これまで様々なピアレビューに関する議論がなされてきた。ソフトウェアインスペクション技法とは、ソフトウェア成果物の作成者ではない第三者が成果物に混入した欠陥を除去することを目的とした正規の検証技法である。Faganは、ソフトウェアインスペクションにおけるモデレータなどの役割の定義、プロセスとしてオーバビュー、準備、インスペクション、リワーク、フォローアップの定義、エラータイプを18種類のデザインに関するものと13種類のコーディングに関するものに分類したサマリーレポートを発表した。本研究発表からFaganインスペクションに対する有効性やプロセスの改善など様々な研究が発表された。Gilbは、Faganのソフトウェアインスペクション技法をベースとして、ピアレビュープロセスと導入事例をまとめた[1-6]。Wiegersは、その導入方法や効果を上げるための成功要因を示した[1-4]。一方、Fagan, Gilb, Wiegersの提唱した手法に対して、その効果を疑問視する議論も出ている[1-7][1-8][1-9][1-10][1-11]。このようなピアレビューを取り巻く研究は様々であるが、未だに決着を見ないものであり、ピアレビュー手法とそれを用いた品質評価をこれまで以上に効果的、効率的にする研究が重要となっている。

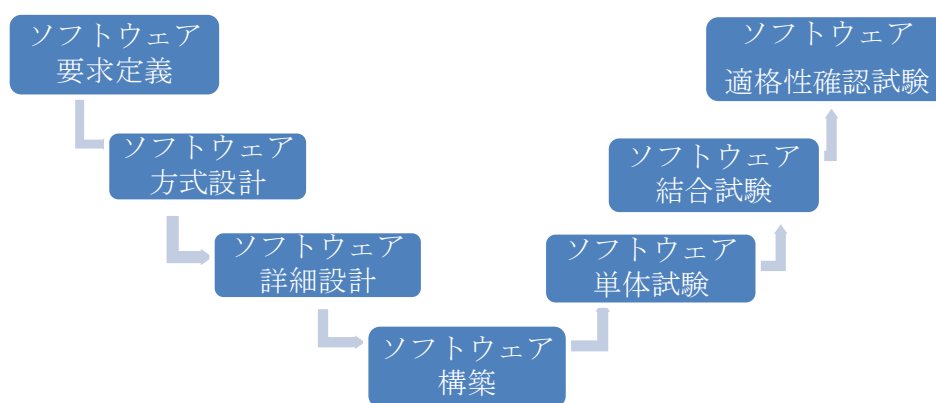


図1 ソフトウェア開発ライフサイクルモデル (V字開発モデル)



## 1.2 ピアレビュー

ピアレビューとは、図 2 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（Fagan のソフトウェアインスペクション技法では、ロギングミーティングと呼ぶ）、編集とフォローアップ、ピアレビュー改善で構成される[1-6]。個々の活動について紹介する。

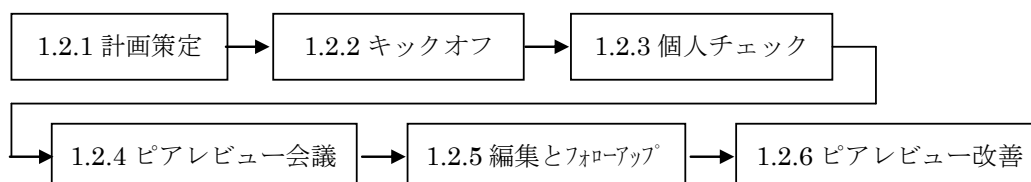


図 2 ピアレビューの流れ

### 1.2.1 計画策定

ピアレビューリーダーは、製品特性、開発工程、開発者スキルに応じて、ピアレビューチーム人数、ピアレビュー工数、ピアレビュー期間、個人チェック速度、欠陥抽出目標件数、ピアレビュー開始・終了基準などを含むピアレビュー計画を策定する。品質向上のためには、ピアレビューチームとして、ピアレビュー対象成果物の上位成果物作成者／利用者、テスト設計者、関連機能の設計者などを含み、多面的に欠陥を抽出できるようにするべきである。またピアレビューでは、チーム内に表 1 のような役割を設定する。さらにピアレビュー開始・終了基準を定量的に明確化しておき、欠陥を後工程に流出しないことが重要である。ただし欠陥抽出目標件数などの単一の指標のみで終了を判断すると、不具合抽出が不十分となることがある。欠陥検出数と工数の組合せで品質を評価するゾーン分析技法[1-12]（詳細は 4.2 章に示す）などを用いて、終了判断を予め計画しておく必要がある。

表 1 ピアレビューチームメンバーの役割

| 役割              | 活動内容  |
|-----------------|---|
| リーダー            | <ul style="list-style-type: none"><li>・ 活動計画、開始・終了のチェック、データの収集、フォローアップ活動を行う。管理者は通常リーダーにならない。</li><li>・ インスペクションの進め方に対する教育を受けた人から選ぶ。</li></ul> |
| ミーティング<br>モデレータ | <ul style="list-style-type: none"><li>・ ミーティングの進行役である。通常リーダーが担当する。</li><li>・ 自身もチェッカとして行動して良い。</li></ul>                                      |
| 書記              | <ul style="list-style-type: none"><li>・ ミーティングにおいて、課題やその他の事項を記録する。</li><li>・ 書記もチェッカとして行動して良い。</li></ul>                                      |

| 役割   | 活動内容  |
|------|---|
| チェッカ | <ul style="list-style-type: none"> <li>・ 特定のチェックタスクを請け負う技術的な専門家。</li> <li>・ 各々のチェッカの役割（表 2）に応じて独立したチェックリストを用意する。</li> </ul> |

### 1.2.2 キックオフ

キックオフの目的は、ピアレビューチームメンバー全員に対し同時に必要な情報を伝えることである。特に、ピアレビューに不慣れなメンバーがいる場合に実施すると良い。主な内容は以下の通りである。

- ・ ピアレビュー計画の周知
- ・ 最新のルールセットやチェックリストの確認。
- ・ ピアレビュー手順を伝える。特に個人チェック速度を説明すると良い。
- ・ リーダが、各チェッカの観点（役割）に応じた作業指示をする
- ・ スケジュールを調整，確認する

### 1.2.3 個人チェック

ピアレビュー会議の前に、各チェッカが個別にレビュー対象成果物を精査する活動である。推奨される（最適な）個人チェック速度を守る必要がある。また、各チェッカに与えられた観点（役割）の範囲を完全に遂行する必要がある。役割の例を表 2 に示す[1-6]。

表 2 チェッカ役割の例

| 観点（役割） | 観点（役割）内容  |
|--------|---|
| ユーザ    | ユーザや顧客の視点からのチェックに専念する。  |
| テスト担当  | テストに関する考慮（テストがしやすいか、テスト要求、テストの順番、並行テストのための開発順序等）に専念する。        |
| システム   | 開発対象ソフトウェアの範囲を越えて、システム全体やそれに関連すること（ハードウェア、文書化、納期、販売形態等）に専念する。 |
| 財務     | コストと収益に関する見積りやリスク、金額に専念する。                                    |
| 品質     | 直接・間接に品質に関わる属性の視点に専念する。                                       |
| サービス保守 | フィールドサービス、インストール等に専念する。                                       |
| バックワード | 資料を最後のページからチェックする。  |
| ルール    | 成果物に対するルールに特別の注意を払う。  |

個人チェックにおける成果物のリーディング技法として、表 3 に挙げられるものがある。

表 3 個人チェック技法の例

| リーディング技法           | リーディング概要                                  |
|--------------------|---|
| アドホックリーディング        | 成果物を個人のスキルを用いて読む技法。<br>熟練者の場合、効果がある[1-13] |
| チェックリストベースドリーディング  | 過去の不具合やノウハウをチェックリストとして纏めて利用する技法[1-13]     |
| パースペクティブベースドリーディング | チェックカに観点（パースペクティブ）を与え、その観点到集中する技法[1-13]   |
| リスクベースドリーディング      | 想定されるリスクを洗出し、リスクが発生しないかという点に集中する技法[1-13]  |
| シナリオベースドリーディング     | 想定されるシナリオ（ユースケースなど）を用いる技法[1-13]           |
| タイムコントロールドリーディング   | 優先順位の高い対象に、より多くの時間を割り当てる技法[1-14]          |
| ペルソナパースペクティブリーディング | ユーザをモデル化したペルソナを利用する方法。<br>[1-15]          |

#### 1.2.4 ピアレビュー会議

ピアレビュー会議の目的は、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。ピアレビュー会議で、個人チェックの結果見つかった、欠陥と思われる項目をチームの記録として収集する。グループでチェックプロセスを実施することで、新たな課題を発見する。また改善提案や、レビュー対象成果物の作者に対する質問も記録する。会議開始にあたっては、レビューア全員の個人チェックが終っていることが前提であり、終わっていない場合にはピアレビュー会議を延期する。これは、ピアレビュー会議の品質を保つための措置である。ただし、実際の開発現場では、ピアレビュー会議に集まれる日時が限られているなどの制約があることが多く、ピアレビュー会議延期などの措置をとれない場合も多い。その場合には、事前にリスクとして考慮し対策を講じておくべきである。

ピアレビューにおいては、表 1 に示したミーティングモデレータ、書記の役割が重要である。特にミーティングモデレータは会議時間内に作業成果物の全体に対す

るチェックを完了できるように会議を制御する必要がある。書記は、ピアレビュー会議における発言を漏れなく記録することに集中し、記録が間に合わない場合には会議を一時中断する。またピアレビュー会議終了時には、記録した内容を読み上げ、ピアレビュー会議で指摘した内容に漏れがないかを合意する。図 3 にピアレビュー会議プロセスを示す。ピアレビュー会議は、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ことである。Gilb は、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、ピアレビュー会議を効果的な活動にできると報告した [1-6]。

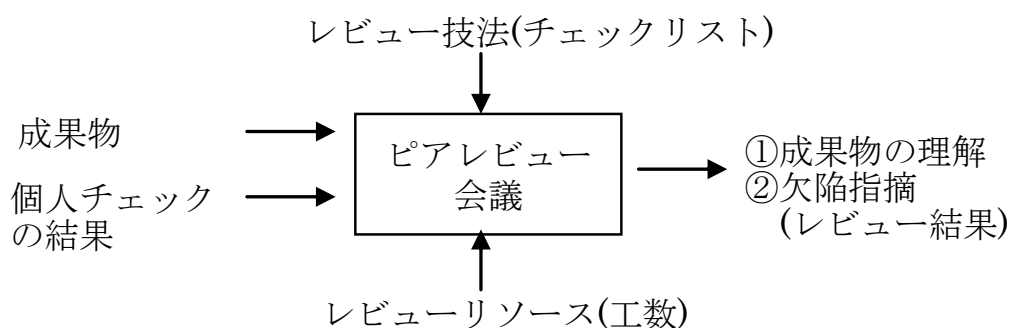


図 3 ピアレビュー会議プロセス

### 1.2.5 編集とフォローアップ

ピアレビュー会議の結果を受けて、レビュー対象成果物の欠陥を修正する活動である。通常、レビュー対象成果物の作成者自身が修正を行う。その際に、新たな欠陥の混入につながるため、機能に磨きをかけたり、新機能の追加をしないことが重要である。また課題が成果物を作成する際の上流文書に起因する場合、上流文書に対する修正要求を発行する。さらに、今後のプロセス改善のため、修正に要した時間と欠陥数と欠陥の重大度を記録しておくことが必要である。

### 1.2.6 ピアレビュー改善

ピアレビュープロセスを改善するために、表 4、表 5 に示すような値を測定分析し、次のピアレビューのプロセスを変更する必要がある [1-12]。

ピアレビュー改善の内容は、計画策定から編集とフォローアップまでを対象とする。計画策定においては、標準的なピアレビューチームメンバからプロジェクトの状況に応じてチームメンバ数や役割を決定する。キックオフにおいては、プロジェクトメンバ全員を招集し、成果物に関する短時間の説明を行うなどの改善も可能である。

個人チェックやピアレビュー会議では、ピアレビューの効果と効率を高めることが求められるので、表4、表5に示した測定量を用いて改善を行うことが重要である。

表4 基本測定量

| 測定量                                 | 説明  |
|-------------------------------------|---|
| ピアレビュー指摘欠陥数                         | ピアレビューで検出できた欠陥数                                       |
| 成果物規模                               | ピアレビュー対象の成果物規模  |
| ピアレビューア毎の個人チェック時間                   | 個人チェックは、レビューで必須な活動であり、どれくらい時間をかければ、有効なレビューとなるのかを把握する。 |
| ピアレビュー会議工数（ピアレビュー会議参加人数×ピアレビュー会議時間） | 会議参加者数は、教育目的などのメンバがいる場合には、カウントの対象外とした方がよい。            |
| ピアレビューで用いたチェックリストやシナリオで検出した欠陥数      | チェックリストやシナリオの改善の為に有効な測定量である。                          |
| 試験へ流出した欠陥内容と件数                      | 試験へ流出した欠陥数と内容は、今後のレビュー改善に有効な情報である。                    |

表5 導出測定量

| 測定量   | 説明   |
|---|--|
| ピアレビュー指摘密度 = $\frac{\text{ピアレビュー指摘欠陥数}}{\text{成果物規模}}$  | レビュー対象の規模に見合った指摘が行われているかを判断する。値がこれまでの傾向と比べ高い・低すぎることを確認する。                        |
| ピアレビュー工数 = $\text{ピアレビュー会議工数} + \text{個人チェック時間の合計}$     | ピアレビューにかけられた全工数である。  |
| ピアレビュー工数密度 = $\frac{\text{ピアレビュー工数}}{\text{成果物規模}}$     | レビュー対象の規模に見合ったレビュー工数をかけられたかを判断する。値がこれまでの傾向と比べて高すぎる・低すぎることを確認する。                  |
| レビュー指摘効率 = $\frac{\text{ピアレビュー指摘欠陥数}}{\text{ピアレビュー工数}}$ | レビュー能力の評価とレビュー結果の評価に利用する。低い時は、対象成果物の品質が良いのか・レビューアの能力が低いのか・レビュー体制が適切であるかを確認する。高い時 |

| 測定量                             | 説明  |
|---------------------------------|---|
|                                 | は、対象成果物の品質が悪くないかを確認する。  |
| 個人チェック速度 = 個人チェック時間 / 成果物規模     | 個人チェック時間と成果物規模の関係を把握し、適切な速度で個人チェックを行ったかを確認する。チェック速度は速いと、見逃した欠陥が多くなる傾向がある。     |
| ピアレビュー会議速度 = ピアレビュー会議工数 / 成果物規模 | ピアレビュー会議工数と成果物規模の関係を把握し、適切な速度でピアレビュー会議を行ったかを確認する。チェック速度は速いと、見逃した欠陥が多くなる傾向がある。 |

### 1.3 ピアレビューの研究

本章では、これまでのピアレビューの研究について説明することで、1.4章における企業の開発現場でレビュー技法を適用する際の課題を浮き彫りにする。

#### 1.3.1 ピアレビューの全般的な研究

これまでピアレビューに関する研究はさまざまである。これまでのピアレビュー（インスペクション）に関する文献では、1982年にFreedmanとWeinberg[1-16]が、ウォークスルーとインスペクションの実践的ノウハウとして、レビュー種類としてウォークスルー、インスペクション、ラウンドロビンレビューなどのレビュー技法や仕様書レビューの実施方法について、質問に対する回答という形式でまとめている。1993年にはGilbらがFaganの体系を発展させソフトウェアインスペクションプロセスを体系化すると共に、実践的ノウハウや事例をまとめている[1-6]。これらの画期的な点は、ソフトウェアインスペクションプロセスの定義において、レビューアの動機づけ方法や導入における抵抗など実際に開発現場で起こり得る問題について、その解決を示唆する知見がまとめられている点である。1992年には堀内[1-17]がデザインレビューに役立つ考え方として「逆を考える」「境界点を考える」「視点の移動」など24種を示し、開発工程別にデザインレビューにおけるポイントをISO/IEC9126の品質特性に関連付けて示した。さらに、2002年には、WieggersがFaganやGilbが示した活動を強化、修正すると共に、その活動が必要な理由など詳細に解説している[1-4]。

森崎は、ソフトウェアレビュー／ソフトウェアインスペクションと欠陥予防[1-13]において、ソフトウェアインスペクションの動向、ソフトウェアインスペクションの効果と効率、上流品質向上に関するソフトウェア評価技術の国際標準動向、上流工程における発注者視点からの品質向上への取り組み、第三者インスペクションによ

る品質検査と欠陥予防，テストエンジニアが参加するアジャイルインスペクションなど，ソフトウェアインスペクションの動向において，IEEE 1028における Inspection の定義，インスペクションの効果として後工程になるほど欠陥修正や欠陥発見のコストが大きくなる欠陥をテストに先立って検出できるということを述べている．また，実施回数の例として，同一グループで1回／同一グループで複数回／複数グループで1回実施する場合のメリット・デメリットを示している．さらに，参加者が同一会議に参加する同期型と参加者が個々に指摘結果を提出する非同期型のメリット・デメリットも示している．最後に今後のインスペクション研究の方向性として，限られた時間内で効率的に欠陥を指摘する支援，インスペクションのテーラリングが重要であると指摘している．

野中は[1-18]，ソフトウェアインスペクションの実務的課題として効果と効率について解説している．インスペクションの効果的かつ効率的に実施する上での課題として，(1)定量的管理の実践，(2)有効性と効率性の改善，(3)アプリケーション領域知識の組織的活動の3点を示している．(1)について，評価メトリクスとして①欠陥除去率，②欠陥除去規模密度，③インスペクション速度，④欠陥指摘工数密度を示し，その目安として表6の値を示した．(2)有効性と効率性の改善については，アドホックよりもシステマティックなリーディング技法を適用した方が効果的であること，リーディング技法により指摘しやすい欠陥の種類に違いがあることを示し，要求インスペクションではユースケースベースドリーディングが効果的であり，派生開発ではテストケースに基づいたリーディング手法が効果的であると言及している．(3)アプリケーション領域知識の組織的活動については，(1)(2)で示したエンジニアリング技法と共にアプリケーション領域知識の共有の必要性を述べている．

表6 インスペクションメトリクス

| メトリクス      | 要求<br>インスペクション | 設計<br>インスペクション | コード<br>インスペクション |
|------------|----------------|----------------|-----------------|
| 欠陥除去率      | 70%            | 70%            | 70%             |
| 欠陥除去規模密度   | 1.5 欠陥／頁       | 1.0 欠陥／頁       | 5.0 欠陥／頁        |
| インスペクション速度 | 2 ページ／時間       | 5 ページ／時間       | 200LOC／時間       |
| 欠陥指摘工数密度   | 0.5 欠陥／人時      | 0.5 欠陥／人時      | 1.0 欠陥／人時       |

込山は[1-19]，上流品質向上に関するソフトウェア評価技術の国際標準動向として ISO/IEC JTC1 SC7 における国際標準の動向として，その体系と上流工程品質向上のための内部／外部品質特性のメトリクスとして，機能性(14個)，信頼性(8個)，使用性(18個)，効率性(9個)，保守性(9個)，移植性(12個)があることを示し，その測

定方法の一部を示している。

このようにピアレビューにおける研究について様々な報告がある。これらの研究は、図4に示すようにピアレビュー管理技法とチェックリストなどのピアレビュー実施技法の研究に分けられる。ピアレビュー管理技法は、ピアレビュープロセスの見直し、ピアレビュー結果の評価方法、ピアレビュー会議の有効性評価がある。ピアレビュー実施技法の研究では、チェックリストの改善、シナリオを用いる手法、ツールによるサポートの提案などがある。

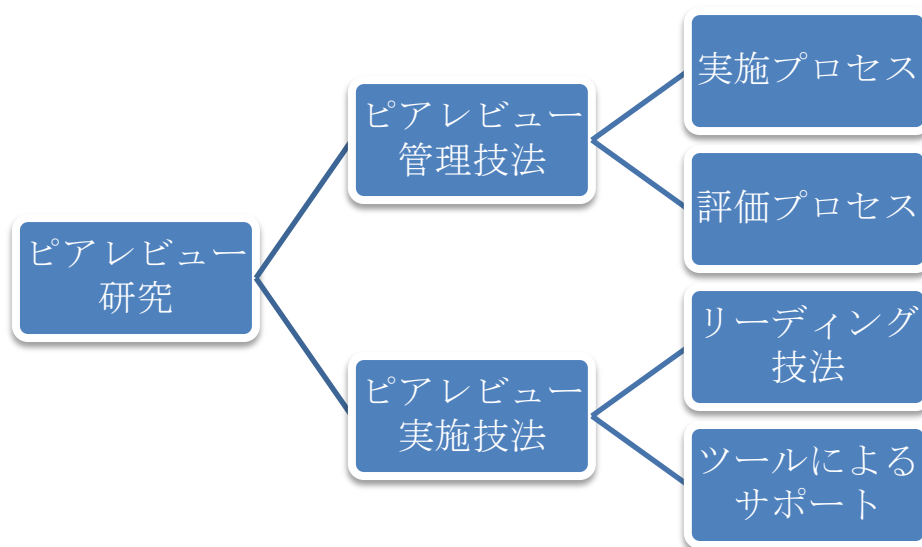


図4 ピアレビュー改善の樹形図

### 1.3.2 これまでのピアレビュー管理技法の研究

Weller[1-20]は、3年間で6700件のコードインスペクションデータと670のインスペクションのデータ、650件のドキュメントに関するインスペクションデータを収集し、ユニットテスト前にコードインスペクションを実施する有効性を示した。またデザインドキュメントやソフトウェアアーキテクチャのインスペクションをしないとコードインスペクションが有効でないことを示した。

Grady[1-21]は、インスペクションの導入における5つのステップとして、①組織目標の定義、②準備、③インスペクションの為にインフラ整備、④現状のプロセスとのベンチマーク、⑤現状プロセスの変更とトレーニングが必要であることを示し、リワークの60%を削減できたと結論付けた。

Eickelmann[1-22]は、Faganインスペクションを改善し、インスペクション時間が8時間以下であれば、かけた時間と検出できる欠陥数の関係が直線的に変化し、インスペクションプロセスの省略や準備時間の削減は、検出できる欠陥数を減少させると



結論付けた。

Biffi [1-23]は、チームサイズとリーディング技法の関係をデータで示した。

Miller [1-24]は、インスペクションにおけるチーム選定に対するコンピュータのサポートにより、Myers-Briggs Type Indicator を用いることで、インスペクションの効果を向上できるという仮説を示した。

Shull [1-25]は、これまでのインスペクションの歴史を振り返り、NASA のインスペクションの拡大期において、プロセス標準・トレーニング・ガイドブック・サポート環境・データによる成功評価と組織的に新しいアイデアを受け入れる風土が重要であるとした。

Hashemitaba [1-26]は、創造的インスペクションとして、検出した欠陥を分析し、ナレッジ化することで、インスペクションの効率を上げることができることを示した。

中野らは、ピアレビュー／ソフトウェアインスペクションと欠陥予防の研究として、レビュー工数とレビュー指摘の関係について、500 プロジェクトのコードレビューデータを用いて評価を行った。その結果、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 [1-27]。

Ferreria らは [1-28]、コードレビューのスピードを制御することでコードインスペクションのパフォーマンスを向上できることを示した。

### 1.3.3 これまでのピアレビュー実施技法の研究

リーディング技法に関しては、Robbins [1-29]らは、成果物を読む際に重要なステークホルダであるユーザ／テスター／デザイナーなどの役割で成果物をレビューする Perspective-Based Reading において、テスターとユーザの観点でレビューを実施し、検出できるタイプの欠陥が異なることを示した。また、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。

Berling [1-30]は、インスペクションにおけるリーディング技法として Perspective-Based Reading (PBR) と Checklist-Based Reading (CBR) の比較を行い、PBRの方が多くの欠陥を少ない工数で検出できたことを示した。

Farchi [1-31]は、コードレビューにおいて、個人チェックを行わずにレビューを行う手法として、Selective Homeworkless という手法を提案し、開発現場で継続的にできるには、個人チェックの負荷を削減する必要があるとした。

Hatton [1-32]は、コードインスペクションにおけるチェックリストの有効性を評価し、過去の欠陥に注目して作成されたチェックリストは、時間当たりの欠陥検出数を向上できるとし、また Capture-Recapture 手法で成果物に含まれる欠陥数を予測で

きる理論を示した。このようにこれまで様々な文献や研究で Fagan インспекションのプロセスを改善し、その効果を測定する試みがなされてきた。

Murphy [1-33]らは、個人チェックを 2 回実施することで、ピアレビュー会議を実施せずに効果的なレビューを可能とし、個人チェックの結果をリアルタイムに表示する Review Windows を用いる手法を提案した。

Johnson[1-10]は、従来のフォーマルテクニカルレビュー手法に対し、顔を合わせた会議形式ではなく、ネットワークなどを利用するなどを 7 つの改善点を示した。

Genuchten [1-34]らは、Fagan のインспекションにおけるピアレビュー会議の方法として、Electronic Meeting System を用いて 14 回の実験を行い、その効率と効果を向上できることを示した。Electronic Meeting System において、個人チェックで検出した欠陥をネットワーク上のパソコンへ登録し、全員で共有しておく。本実施方法の特徴は、全員で共有した欠陥を参照し、各レビューアがピアレビュー会議中に新たな欠陥を検出するという点、各レビューアが対象成果物の別々の部位をレビューするという点である。

## 1.4 ピアレビューの課題

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 [1-6] [1-15] [1-35]。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 [1-13]。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 [1-29] やプロセスを重視し組織力を活用する方法 [1-36] [1-37] [1-38] などが提案・評価されている。

Gilb は、Fagan の先駆的なソフトウェアインспекション（ピアレビュー）の考え方を発展させ、その手法を体系的にまとめた [1-6]。Wiegers は、厳格なピアレビューを導入する際に発生するコストに対する開発現場からの反対に対し、どのように対応すれば良いかなどピアレビューを導入する際に起こる問題点と解決策を導入事例と共に示した [1-4]。これらの取組は、ピアレビューがソフトウェア欠陥を世の中に流出させないための重要なプロセスであるという信念に基づいている。

一方、実際にピアレビューを開発現場に導入しようとする時、ピアレビューを実施しても誤字脱字の欠陥しか検出されず効果がない、動作タイミングなどの欠陥はピアレビューでは検出できない、ピアレビューが工程を遅らせるなど様々な問題点を挙げる反対勢力が現れる [1-4]。テストにおける欠陥の発見・修正コストに比べ、レビューにおける欠陥の発見・修正コストは、1/20 であるという報告例 [1-4] などを示しても、他社と自部門では製品・顧客が異なり費用対効果はないという反対勢力からの意見が示される。これらの反対勢力の意見は、現状を維持したい一部の声の大きい開

発者のものであり、開発チーム全体のものではなく、データに基づくことはほとんどない。あるいは過去に導入しようとして失敗した経験に基づく意見の場合もある。さらに、開発部門と管理部門間の組織の壁という問題もある。開発部門では、ピアレビューが効果的に実施できておらず、レビューへの期待が低く、レビューにかかる時間を少なくしたいという潜在的な思いがある。管理部門では、V字モデルの左側部分における製品品質の向上の主たる活動がピアレビューであると信じ、ピアレビューに投入する時間を多くすることを開発部門へ要求する。

このような組織的な背景において、ピアレビュー実施の効果を向上することと、ピアレビュー管理の効果を向上することが必要である。そのためには、先人の築いた体系や経験を基に、開発現場の実態に即したピアレビューのプロセス定義と実践が重要である。

## 1.5 本論文の概要

本論文では、ピアレビューという活動において、実開発の現場でピアレビューが真に有効な活動となっているか、その有効性を高めることができるかという視点から、以下の3つの課題について、これまで提案されていない測定手法や新しい指標を用いて定量的に課題を見える化し、その解決策を示す。

- ・課題1：ピアレビュー後に不具合が流出することに対する効果的な活動がない
- ・課題2：ピアレビュー会議の有効性について疑問があり決着がつかない
- ・課題3：品質評価技法が不十分である

以下2章では、課題1に対して、ピアレビュー後に不具合が流出する原因について調査し、その一つの要因がピアレビュー会議の実施方法に問題があると仮定し、定量的に測定した。その結果、ピアレビュー会議を不具合抽出に特化するプロセスを定義し、不具合流出を防止できることを示す。

3章では、課題2に対して、ピアレビュー会議の有効性を評価する実験を行い、これまでの研究と異なるデータを示し、開発現場でピアレビュー会議が有効であることを示す。

4章では、課題3に対して、これまでのピアレビュー評価技法であるゾーン分析における課題を明確化し、その解決策としてピアレビュー網羅率という新たな指標を定義し、開発現場で適用し、その有効性を示す。

これらの3つの課題に対する解決策によって、これまで開発の現場で、その効果と適用の効率性が不十分であったピアレビューを真に有効な活動とし、ソフトウェア品質と開発効率を向上することに貢献できた。



# 第2章 ピアレビュー会議の改善に関する研究

## 2.1 ピアレビューにおけるピアレビュー会議の位置づけ

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[2-1].

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平均的に2週間の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 [1-9] [1-10].

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という取組み、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している取組みもある[2-2] [2-3].

上記のいずれの研究においても、ピアレビュー会議の問題については定性的評価であり、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調査しておらず、企業活動で重要な限られた時間内で効率と品質を改善するという点からは不十分である.

本章では、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変更し製品品質を改善できることを示す. 2.2 章において、ピアレビュー会議の問題点を明確化する. 次に、2.3 章においては、レビュー会議を定量的に評価する手法を提案し、2.4 章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. 2.5 章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. 2.6 章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.

## 2.2 従来研究と解決すべき課題

ピアレビュー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研究を本研究における従来研究として述べ、本研究で解決すべき課題を明確にする.

### 2.2.1 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される。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[1-10]。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、発言がないなどアイドル時間があることを挙げている。

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[1-9]。さらにピアレビュー会議開催がプロジェクトの進捗を平均2週間遅らせていると報告している。

### 2.2.2 本研究が扱う課題

三菱電機の開発現場において、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、欠陥抽出を行う。しかし、ピアレビュー会議を完了しても、テスト段階に流出する欠陥が残存する場合が多い。

2.2 章で示した課題や上記の開発現場における課題など様々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、欠陥抽出件数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。

## 2.3 ピアレビュー有効時間比率を用いたピアレビュー会議の改善

前章で述べた問題点を解決することを目的に、ピアレビュー有効時間比率を用いて、ピアレビュー会議を改善する手法を示す。

提案する手法は、レビュー会議における発言内容を測定し、全時間と欠陥検出時間との比率であるピアレビュー有効時間比率と呼ぶ指標を用いてピアレビュー会議を欠陥抽出活動となるように制御するものである。以下、ピアレビュー会議を改善するための新しい指標であるピアレビュー有効時間比率とそれを利用したピアレビュー会議の改善方法について述べる。

### 2.3.1 ピアレビュー会議の定着

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、読み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、まず現場で実施されているピアレビュー会議を定量的に把握するために、ピアレビュー会議の流れに注目し、Gilb ら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）と実施してはいけない活動（内容の読上げ、議論、修正案、無発言）を表 7 のように定義した。

表 7 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   | ピアレビュー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作業成果物以外の議論  |
| 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              |

測定者はピアレビュー会議に出席し、図 5 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毎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 7 の活動に該当するボタンを押下する。その際に発言者の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。

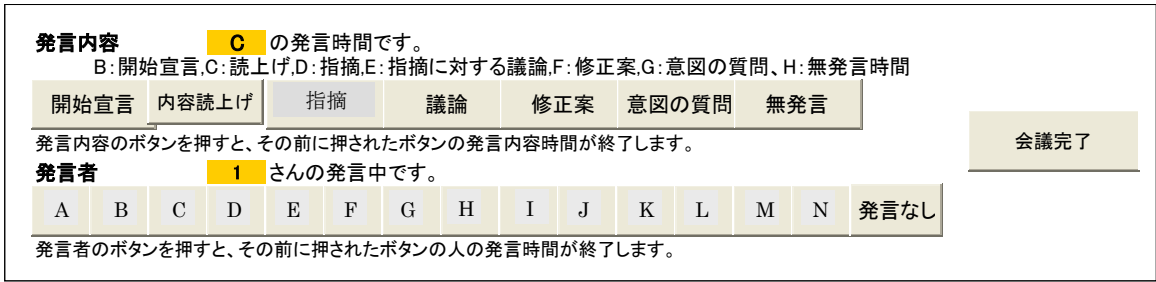


図 5 ピアレビュー会議測定ツール

### 2.3.2 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。なお三菱電機におけるピアレビュー会議時間は、成果物規模により決定されたため、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が必要である。

表 8 ピアレビュー会議時間測定結果



| 会議 | 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、2.3.3章に述べる。

表8より、部門毎に若干の差はあるが、以下の課題があった。

- 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6から図8は、測定結果をグラフ化したものであり、円グラフは発言内容、棒グラフは出席者毎の発言時間である。図6は、表2の会議No.12とNo.13の測定結果である。図6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わかる。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。

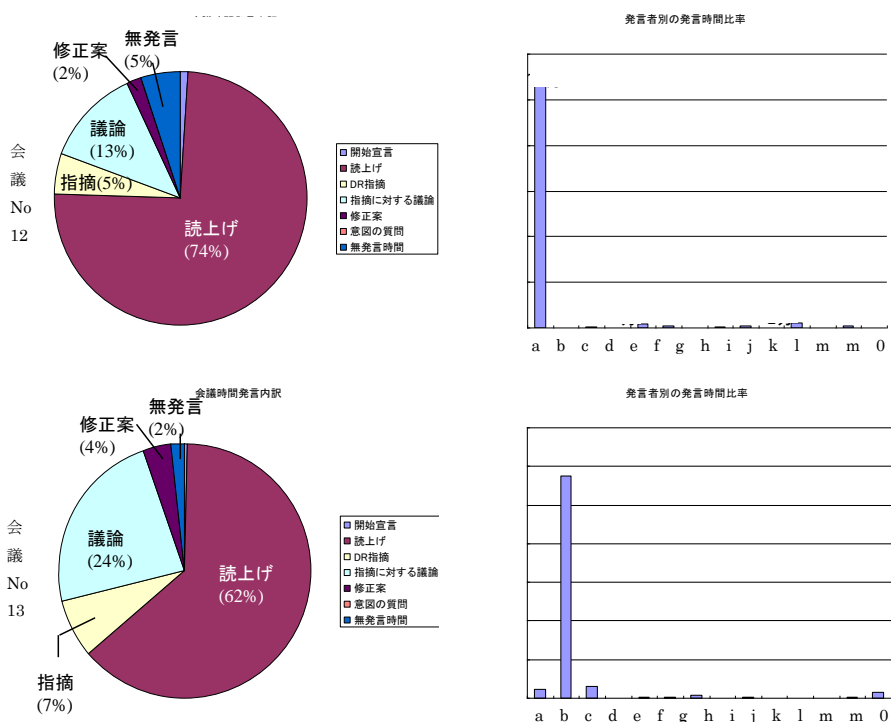


図6 仕様説明を中心としたピアレビュー会議測定結果

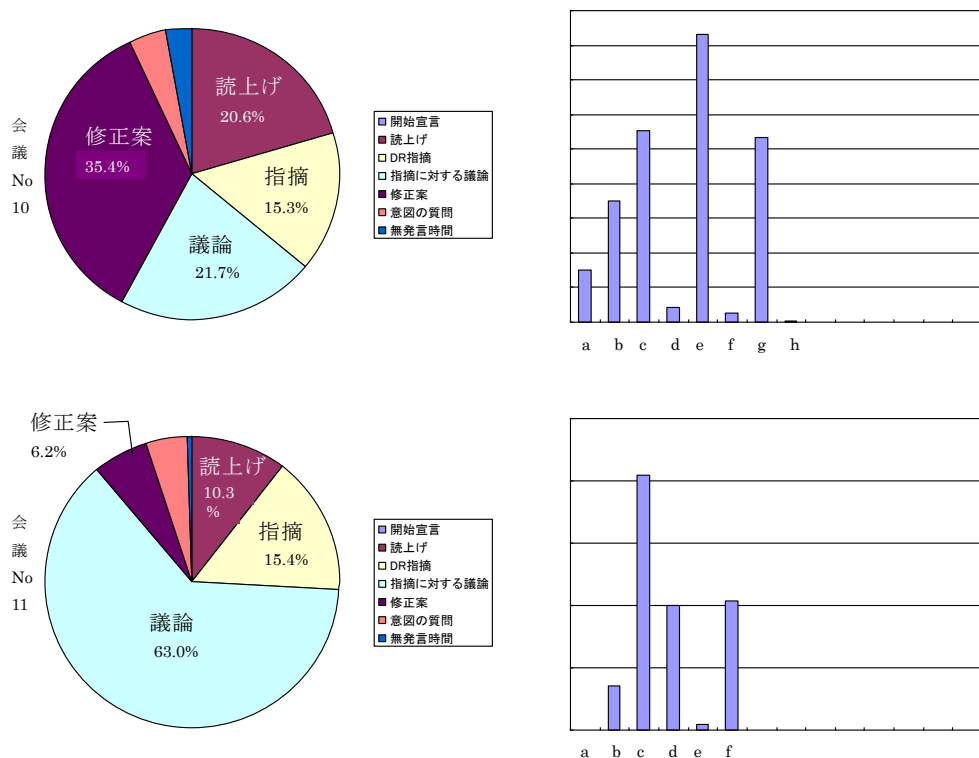


図7 設計活動を中心としたピアレビュー会議測定結果

図7は、表2の会議No.10とNo.11の測定結果である。図7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。

図8は、表2の会議No.9とNo.19の測定結果である。図8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。

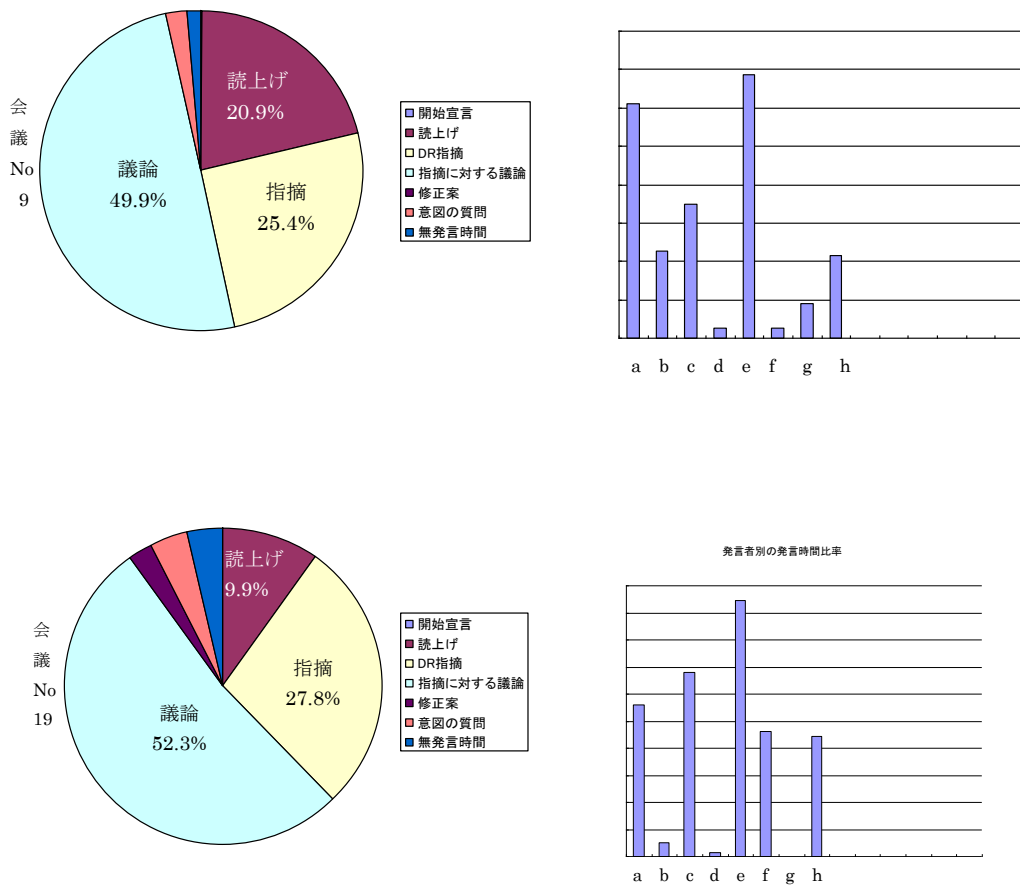


図 8 欠陥抽出を中心としたピアレビュー会議測定結果

### 2.3.3 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

2.3.2 章で示したように、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る。まず、図 6 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、ピアレビュー会議のプロセス定義を行い、プロジェクトのピアレビュー会議に適用した。ピアレビュー会議プロセス定義内容の一部を以下に示す。

- 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。
- 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。
- 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。
- 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。
- ⑪モデレータは，以下に注意し，会議が効果的になるよう制御する。
  - ・資料自体の読上げをしていないか。
  - ・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。
  - ・「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。
  - ・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。
  - ・用語の解釈が出席者間で異なっていないか。

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毎の時間を測定した。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（ $TF$ ）と呼ぶ指標を導入した。ピアレビュー有効指摘比率（ $TF$ ）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であり，以下の式で表わされる。

$$TF = \sum_{n=1}^i(Ti) / T \quad (\text{式 1})$$

ここで， $T_i$ : 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，ピアレビュー有効時間比率（ $TF$ ）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，三菱電機におけるピアレビュー会議時間は，成果物規模により決定されたため，ピアレビュー有効時間比率を向上することは，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。成果物規模に応じてピアレビュー会議時間を決めることにより，ピアレビュー会議のプロセスを一定にする効果があり，一定の工数かけた上で成果物の品質を評価することができる。また，ピアレビュー会議を欠陥抽出中心とした活動とするために，内容読上げ時間を少なくする必要がある。ただし，

内容読上げはレビューアの意識合わせの為には重要であり、新規開発や製品知識が不十分なレビューアがいる場合には、図2で示したキックオフで実施しておくことが望ましい。

## 2.4 適用と評価

### 2.4.1 適用

2.3章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9に示す。表9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わかる。

表9 ピアレビュー会議時間測定結果（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9は、表9の会議No.2の測定結果である。改善後のピアレビュー有効時間比率は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。

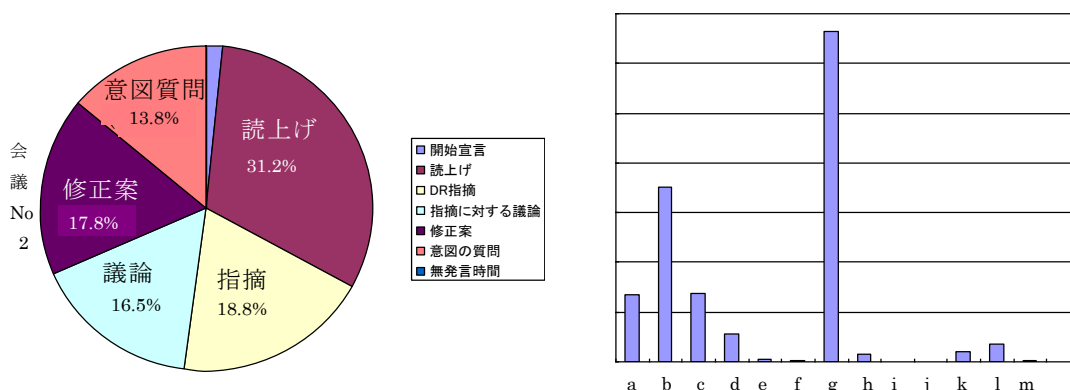


図9 部門Aのピアレビュー会議測定結果（改善後）

なお、意図の質問は表 7 に示す通り、指摘に対するその意図の確認であり、指摘が具体的であれば不要となるものであり、その時間を少なくすることが望ましい。

図 10 は、表 9 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、指摘時間比率が向上、読上げ時間比率の減少し改善効果があった。

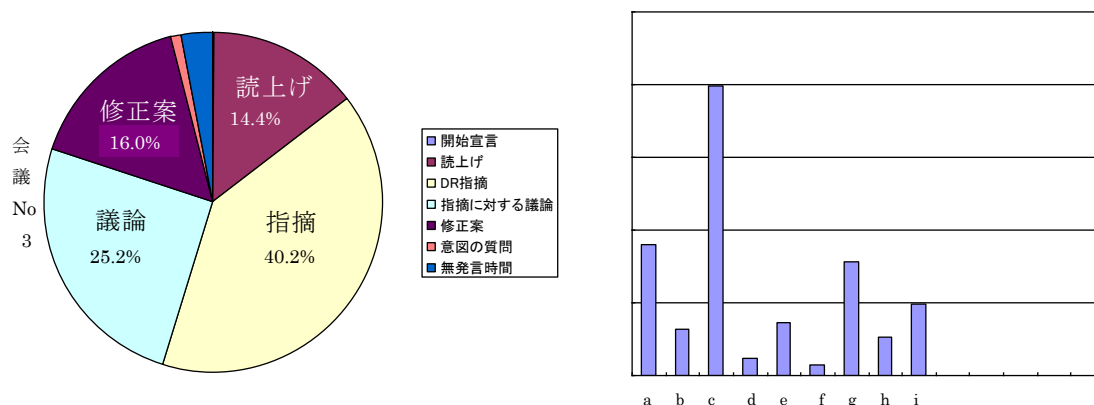


図 10 部門 H のピアレビュー会議測定結果（改善後）

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ても、ピアレビュー会議におけるピアレビュー有効時間比率は、ばらついている。そこで全体的な傾向を把握する為、改善前後のピアレビュー有効時間比率の変化を評価した。

## 2.4.2 評価結果

2.4.1 章の適用結果について、改善前と改善後のピアレビュー有効時間比率の変化について評価し、その有効性を判定した。まず改善前後の 2 群の分散が異なるかどうかを確認するために分散比の検定を行った。表 10 から、P 値は 0.022 となり有意水準 5% で有意差ありとなり、分散が異なるという結果となった。そこで、分散が異なるとして、改善前後の平均値の差の検定結果を表 11 に示す。表 11 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、改善前後の平均値が統計的に有意差ありとなった。

表 10 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  | 36.10  | 12  | 11  | 3.927    | 0.022      | 3.012    |
| 改善後 | 26.36 | 141.77 | 8   | 7   |          |            |          |

表 11 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  | 36.10  | 12  | 9   | -3.673 | 0.003      | 1.833    | 0.005      | 2.262    |
| 改善後 | 26.36 | 141.77 | 8   |     |        |            |          |            |          |

図 11 は、改善前と改善後のピアレビュー有効時間比率の測定結果を箱ひげ図を用いて示したものであり、改善前後のピアレビュー有効時間比率に差異があることが判る。

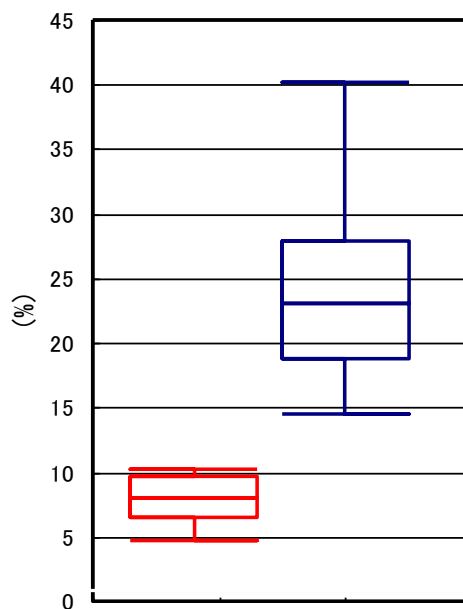


図 11 改善前後のピアレビュー有効時間比率の箱ひげ図



## 2.5 レビュー会議改善による品質改善に対する評価

### 2.5.1 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の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の変化について測定した結果を示す。

表 12 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 12 から、P 値は 0.0001 となり有意水準 5% で有意差ありとなり、分散が異なるという結果となった。そこで、分散が異なるとして、改善前後の単位時間当たりの指摘件数の平均値の差の検定結果を表 13 に示す。表 13 から P 値が 0.0003 と 0.0005 であり、改善前後の平均値が統計的に有意な差があることがわかる。

表 12 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 13 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 12 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、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。

三菱電機においては、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。

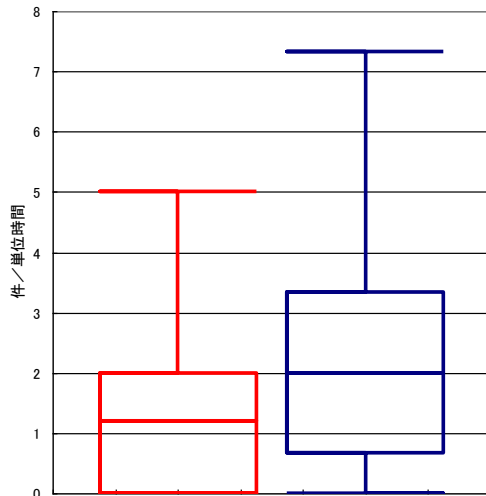


図 12 改善前後の単位時間単位指摘件数の箱ひげ図

### 2.5.2 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 13 は、今回のピアレビュー改善により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。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。

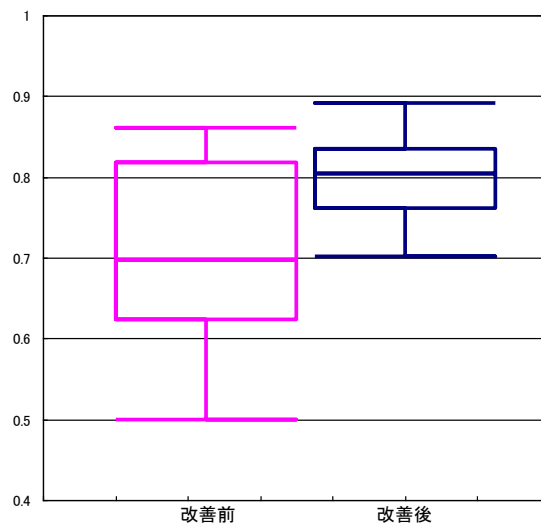


図 13 全検出欠陥に占めるレビュー指摘欠陥割合の変化

### 2.5.3 ピアレビュー参加者による主観的評価および改善活動の展開

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。

- ・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。
- ・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、以下のような課題点も上げられた。

- ・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。
- ・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[2-3]。その結果をガイドラインとして纏め、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。本ピアレビュー定義を付録1に示す。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。

## 2.6 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[1-27]。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、プロジェクト毎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[2-4]では、レビュー工数を式2で定義した。

$$\text{レビュー工数} = \sum \text{各レビューアのレビュー実施時間} \quad (\text{式2})$$

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[2-5]。

$$\text{欠陥指摘工数密度} = \sum N F_i / \sum T_i \quad (\text{式 3})$$

$N F_i$  は欠陥数， $T_i$  は総レビュー時間であり， $i$  はレビュー参加者を示す。

式 3 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している。

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[2-6][2-7][2-8][2-9]。

上記研究においては，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することができ，ピアレビュー評価手法を改善することができると思う。

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[1-29]。

## 2.7 まとめと今後の課題

本章では，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制御することが重要であることがわかった。

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。



# 第3章 レビュー会議の有効性評価に関する一考察

## 3.1 まえがき

要求分析からコーディングに至るソフトウェア開発の上流工程では、設計文書やプログラムを対象としたレビューが主たる検証手段である。Gilb はソフトウェアインスペクションを体系化し、その中心的活動としてレビュー会議（ロギングミーティング）を位置付けた[3-1]。一方 Porter は、レビュー会議を用いるレビュー手法は、レビュー会議を用いない手法に比べて、より効果が大きいとも言えず、また少ないとも言えないと結論付けている[3-2]。

本章では、開発現場で実施している条件でレビュー会議の実験を行い、ソフトウェア開発におけるレビュー会議の有効性を再評価する。

## 3.2 レビュー会議に関する有効性評価

実際のソフトウェア開発現場では欠陥検出のためにレビュー会議を利用する機会が多い。筆者らは、レビュー会議の欠陥検出上有効性について Porter の結論と産業界の評価にギャップがあると認識している。今回の実験では、開発現場の実施条件に合わせた実験を行い、Porter の結果と比較しその結論の再評価を行った。なお、今回の実験の比較対象を Porter の実験とした理由として、3.2.5 章で示す不十分な条件で実開発者にピアレビューを要求することは困難であったことによる。

### 3.2.1 実験条件

Porter らの Maryland 大学実験[3-2]と今回の実験条件を表 14 に示す。今回の実験では家電製品に関する要求分析ドキュメントを用いた。Porter の実験では 2 つのドキュメントを対象にデータを収集しているが、データ数の多い WLMS (Water Level Monitor System) を比較の対象とした。表 15 は、レビュー形態の定義である。

表 14 実験の条件

| 実験条件       | A: Porter 実験の値    | B: 今回の値 |
|------------|-------------------|---------|
| 成果物規模      | 24 頁※             | 12 頁    |
| レビュー形態     | PI, DC, DD の 3 形態 | DC      |
| チーム員数      | 3 名               | 6 名     |
| 個人チェック時間   | 2.5 時間            | 0.5 時間  |
| ピアレビュー会議時間 | 2.5 時間            | 0.5 時間  |

※文献[3-3]参照

表 15 レビュー形態

| レビュー形態                    | 説明   |
|---------------------------|--|
| PI:Preparation-Inspection | 個人チェックは仕様書理解に集中，ピアレビュー会議で欠陥を抽出する。              |
| DC:Detection-Collection   | 個人チェックで仕様書の欠陥を抽出，ピアレビュー会議で欠陥を報告，追加で新たな欠陥を検出する。 |
| DD:Detection-Detection    | ピアレビュー会議を実施せず，ピアレビュー全時間を個人チェック時間に割り当てる。        |

上記 2 つの実験結果を，Votta のレビュー会議の効果尺度 Meeting-Gain（以降，欠陥検出率と呼ぶ）Rdm を用いて評価した[3-3]。以下にその定義を示す。

$$Rdm = Ndm / N \times 100 \quad (\text{式 4})$$

Ndm : レビュー会議で検出した欠陥数

N : 検出した全欠陥数。

欠陥検出率は，個人チェックで検出できなかった欠陥をレビュー会議においてどれだけ検出できたかを評価する尺度である。

### 3.2.2 Porter の実験結果

Porter らは要求分析ドキュメントに対し，3 人で構成する学生とプロ開発者チームを複数編成し，レビュー形態を割り当て比較した。その結果，欠陥検出数は DC 形態より DD 形態の方が多いいことを示した[3-2]。

### 3.2.3 今回の実験結果

今回の実験では，13 チームが参加し要求分析ドキュメントに対しチェックリストを用い DC 形態でレビューを行った。表 16 はその実験結果である。今回の実験の Rdm は平均 21.8%であり，文献[3-2]の Fig. 9 から DC 形態の結果を読取り計算した 9%に比べ 12.8 ポイント高い。



表 16 今回の実験結果

| チーム | N  | N <sub>dm</sub> | R <sub>dm</sub> | チーム                | N  | N <sub>dm</sub> | R <sub>dm</sub> |
|-----|----|-----------------|-----------------|--------------------|----|-----------------|-----------------|
| 1   | 27 | 7               | 25.9            | 8                  | 33 | 1               | 3.0             |
| 2   | 13 | 4               | 14.8            | 9                  | 28 | 4               | 14.3            |
| 3   | 36 | 10              | 37.0            | 10                 | 13 | 7               | 53.9            |
| 4   | 38 | 10              | 37.0            | 11                 | 21 | 6               | 28.6            |
| 5   | 29 | 2               | 7.4             | 12                 | 19 | 4               | 21.1            |
| 6   | 22 | 3               | 11.1            | 13                 | 25 | 3               | 12.0            |
| 7   | 31 | 6               | 19.4            | 平均 R <sub>dm</sub> |    | 21.8            |                 |

### 3.2.4 2つの実験結果の統計的評価

Porter は、DD 形態と DC 形態における検出欠陥数を評価し、DD 形態が DC 形態に比べ高いことを示した。そこで、Porter の DD 形態の結果と今回の実験結果を統計的に評価した結果を表 17 に示す。まず 2 群の分散が同一であると仮定し、分散比の検定を行った。P 値は 0.002 となり、分散が同一であるという仮説は却下された。そこで、平均値が同等であると仮定し有意差 5%で分散比が異なる場合の t 検定を行った。その結果 P 値が 0.079 であり 2 つの平均値が同等であるという仮説を却下できず、Porter の結論とは異なる結果となった。

表 17 t 検定の結果

|                                   | 平均   | 観測数 | P 値   |
|-----------------------------------|------|-----|-------|
| Porter DD 形態結果 (欠陥数を図より読み取った値を利用) | 29.0 | 7   | 0.079 |
| 今回の実験結果                           | 21.8 | 13  |       |

### 3.2.5 2つの実験結果の相違に関する考察

2 つの実験結果が異なった要因について、以下に示す。

#### (1) 個人チェック時間の設定方法

今回の測定では 12 頁の文書を用い、Porter の実験では 24 頁の文書を用いた[3]。成果物規模とレビュー参加人数、レビュー時間の関係から Porter のデータにおけるレビュー工数は、0.6 人時/頁 (5 時間×3 名/24 頁) であり、今回の測定結果 0.5 人時/頁 (1 時間×6 名/12 頁) とほぼ同等である。一方、Porter 実験では、レビュー形態に関わらず Phase1 で同一の方法で個人チェックを実施し、Phase2 で、DC 形態ではレビュー会議を、DD 形態では、DC 形態のレビュー会議と同一時間をかけて個人

チェックを行った。その結果、式 5 で示すピアレビュー欠陥検出率  $R$  は、DC 形態では 0.21、DD 形態では 0.43 であった[3-2]。

$$R = \text{ピアレビューで検出した欠陥数} / \text{全欠陥数} \quad (\text{式 5})$$

この式は、次のように表すことができる。

$$\begin{aligned} R^{xx} &= (\text{xx 形態の Phase1 で検出した欠陥数} + \\ &\text{xx 形態の Phase2 で検出した欠陥数}) / \text{全欠陥数} \\ &= R_1^{xx} + R_2^{xx} \end{aligned} \quad (\text{式 6})$$

ここで  $xx$  はピアレビュー形態 (DC または DD)、 $R^{xx}$  は  $xx$  形態による全体のピアレビュー欠陥検出率、 $R_1^{xx}$  は  $xx$  形態による Phase1 のピアレビュー欠陥検出率、 $R_2^{xx}$  は Phase2 のピアレビュー欠陥検出率である。

式 7、式 8 に、DC、DD 形態のピアレビュー欠陥検出率を示す。

$$R^{DC} = R_1^{DC} + R_2^{DC} = 0.21 \quad (\text{式 7})$$

$$R^{DD} = R_1^{DD} + R_2^{DD} = 0.43 \quad (\text{式 8})$$

DD 形態の Phase1 の個人チェックの方法は DC 形態の Phase1 と同様であり、ピアレビュー欠陥検出率は等しいと考えられるので、

$$R_1^{DD} = R_1^{DC} \quad (\text{式 9})$$

これを式 7 に代入して変形すると、 $R_1^{DD} = 0.21 - R_2^{DC}$  が得られるので、

$$R_1^{DD} < 0.21 \quad (\text{式 10})$$

式 8 と式 10 から、

$$R_2^{DD} = 0.43 - R_1^{DD} > 0.22 \quad (\text{式 11})$$

成果物に内在する欠陥数が減れば、かけた労力に対して欠陥の検出量は減少するはずで、DD 形態の場合、Phase2 のピアレビュー欠陥検出率が小さくなることが予想されるが、 $R_1^{DD} < R_2^{DD}$  となっている。これは Phase1 では成果物の理解に時間を要し、欠陥の検出を行うには時間が不十分だったことが原因と推測される。今回の実験では、実験前に 2 名の開発者に依頼し、現場で個人チェックを行うのと同様に、成果物をチェックする予備実験を行っており、30 分で全体に対する指摘ができることを確認した。さらに受講後のアンケートを実施し、13 チーム×6 名

全員が個人チェック時間で一通りの確認を行い時間に不足はないとの意見を得ている。

ソフトウェア開発の現場では工期と掛けられる工数制約の中で最適なレビューを実施する。その為に個人チェックで十分に欠陥を検出してからレビュー会議を実施する。個人チェックが不十分な状況でレビュー会議を実施しても、その有効性が低いことはGilbが主張する通りである[3-1]。

#### (2) チーム編成及びレビュー会議運営方法

Porter 実験では学生とプロ開発者が参加し、今回の実験はプロ開発者のみである。この条件による差異について、文献[2]のFig.7のDetection RatioのデータからDC形態では学生チーム平均 24%、プロ平均 19.25%、DD形態では学生チーム平均 48.6%、プロ平均 42.8%であり学生とプロ開発者チームによる相違はレビュー形態の違いに比べて無視できると判断できる。

次に、レビュー会議では、進行役、書記、読み手、レビューアの4つの役割が必要である[3-1]。Porter 実験では3人1チームであった為、各々の役割を十分果たせず、レビュー会議で新たな欠陥を検出する活動を適切に実施できなかった可能性がある。今回の筆者の実験では、進行役が議論や修正案、無発言の時間を少なくするなど欠陥検出以外の活動を削減する改善を実施済みのレビュー会議であった為[3-4]、欠陥検出効率が向上していることも一因であると考えられる。

#### (3) 成果物の違いによる要因

今回の実験に用いた成果物は、家電製品の要求分析ドキュメントであり、被験者は、開発対象としてドメインの知識を有しており成果物を理解することが容易であった。一方PorterのWLMSは、製品仕様の理解が難しかった可能性がある。

#### (4) 各レビューアのチェック観点設定

今回の実験においては、レビューアにチェックする際の観点を設定してから個人チェックを開始した。各レビューアのチェック観点は、1.2.3の表2に示すものを参考に、以下の4つの観点を利用した。

- ①成果物の作成情報を与える立場（顧客）
- ②成果物を直接利用する立場（設計者）
- ③成果物からプログラムを作成するプログラマで
- ④要求を検証・妥当性を確認する試験者の立場

Porterの実験では、レビューアは3名であり、1名が複数の観点で成物をチェッ

クするする必要があり、成果物の個人チェックが不十分であった可能性が高いと判断できる。

(5) その他の妥当性の脅威に関する要因

今回の実験と Porter の実験結果の相違は、上記以外に以下のような要因が考えられる。

①被験者のスキル

レビューにおける欠陥検出は、実験に参加した被験者のスキルに依存する場合がある。今回の実験と Porter の実験では、被験者のスキルの違いによる結果への影響を評価しておらず、スキルの相違による要因が考えられる。

②利用したチェックリスト

今回の実験と Porter の実験におけるチェックリストは異なるものであり、その違いによる結果への影響を評価しておらず、チェックリストの相違による要因が考えられる。

③欠陥の種類および数

成果物に含まれる欠陥の種類やその数は、成果物ごとに大きく変わることが考えられる。これが実験結果に影響を与えた可能性がある。

### 3.3 まとめと今後の課題

レビュー会議に関する我々の測定結果では、ピアレビュー会議の欠陥検出率は21.8%であり、ピアレビュー会議の有効性は低いという Porter の結論と異なる結果となった。ただし、どのような条件であればピアレビュー会議が有効となるか、欠陥検出率を向上するための条件などを評価できていないため、今後データを蓄積し、レビュー会議の有効性を高めるパラメータを明確化する予定である。

# 第4章 ピアレビュー網羅率を用いた品質評価技法に関する研究

## 4.1 まえがき

品質管理手法は、ピアレビューの結果を評価し成果物に残存する欠陥を計画目標内に抑えるために定量的なデータを用いて行う管理方法であり [4-1]、その管理精度向上のために種々の提案がなされている [2-5] [2-6] [2-7]。しかし、従来の品質管理手法はレビュー効果のバラツキを扱う方法を十分に確立できていない。実際、テスト段階で検出される欠陥の中でレビュー時に検出可能なものが多く、その原因はレビューの実施結果を適切に評価できていないことにある [4-3]。

一般に、各開発工程の品質管理者は、成果物に対するレビュー実施結果を見て、成果物が要求品質を達成しているか、再レビュー等の対策が必要であるかを判断することが多い。この判断は、主に成果物の成果物品質（及びそれによって想定される後工程のリスク）に基づいて行われる。

レビュー結果に基づく成果物品質の判断に用いる方法として、検出欠陥数に加えてレビュー密度の実績から品質を判断するゾーン分析法 [4-1] [4-2] や検出欠陥数の実績から残存欠陥数を推定する欠陥数推定法 [4-3] [4-4] [4-5] [4-6] [4-7] がある。

ゾーン分析法や欠陥数推定法は、ともにレビュー結果に基づく定量的な成果物の品質評価技法であり、レビューが成果物に対してムラなく実施されていることを前提としている。しかし、現実には、成果物に対して、その前半部分でレビュー時間を費やし後半までレビューできない場合が多い。この状況は潜在欠陥が多い成果物を対象としたときに生じやすい。こうした場合に上記の品質評価技法を単純に適用すると、収集されたデータが平均値で評価されてしまうため、残存欠陥を多く含む部分があるにもかかわらず、品質上問題なしと判断を誤ることがある。これが、設計工程から後工程に欠陥が流出する大きな要因となっている。

この問題を解決するために、我々は、実施したピアレビューが対象成果物をどの程度網羅しているかを評価するためにピアレビュー網羅率という新しい指標を導入することで、従来の品質評価技法の弱点を補完する品質評価技法を開発した。実プロジェクトに本技法を適用し、テスト段階への流出欠陥数が減少する効果を確認した。

本章では、4.2章において、従来研究としてゾーン分析について特長と課題を述べ、その適用上の問題を明確にすることで、本章で解決すべき課題を明確にする。次に、4.3章においては、レビュー実施の網羅性を考慮する手法を提案し、4.4章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。4.5章では関連研究について述べ、本提案技法との関係を明確にする。

## 4.2 従来研究と解決すべき課題

レビュー結果に基づく成果物品質の判断を行う方法の一つであるゾーン分析法を、本研究の従来研究として述べ、本研究で解決すべき課題を明確にする。

### 4.2.1 ゾーン分析法

ある開発工程で、検出欠陥数の実績に基づき残存欠陥数を推定しようとするとき、レビューによる検出欠陥密度（欠陥数を対象規模で割り正規化したもの）が過去の実績に比較して高い場合に、二つの相反する解釈が成り立つ。一つは、レビュー対象に元々含まれる欠陥数が多い、つまり残存欠陥数は多いという解釈、もう一つは、レビューが効果的であり欠陥を取りきった、つまり残存欠陥数は少ない、という解釈である[4-1][4-2]。

ゾーン分析法 [4-1][4-2]は、この相反する解釈の問題を解決するために、品質データとして検出欠陥数に加えてレビューにかけた工数を用いる。欠陥数推定法のように直接残存欠陥数を推定するのではなく、図 14 に示すようなレビュー密度と検出欠陥密度の 2 軸平面上で、9 つのゾーンに分け、レビュー対象毎に、計画値に対する実績値を、品質状態がどのゾーン入るかで評価する[4-1][4-2]。

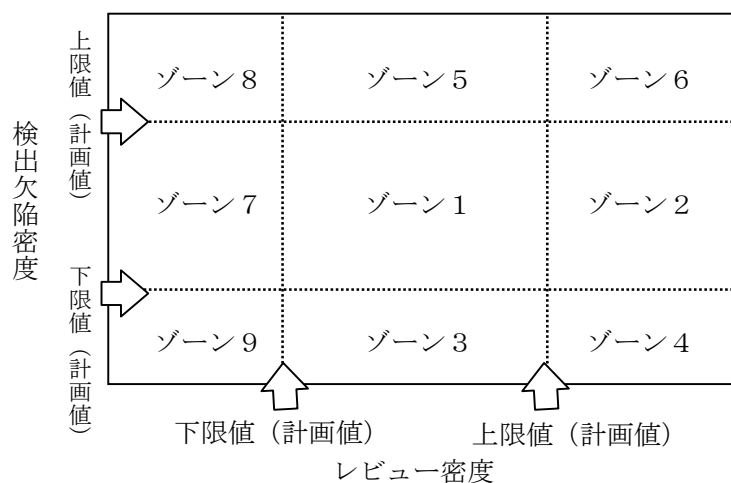


図 14 ゾーン分析法

ここで、計画値は、過去の実績から導出したレビュー密度と検出欠陥密度における上限値と下限値である。レビュー密度  $Dr$  と検出欠陥密度  $Df$  は、検出欠陥数  $Nf$ 、レビュー工数  $Tr$ 、及び成果物規模  $Nw$  から、以下の式を使って計算する。

$$Dr = Tr / Nw \quad (\text{式 12})$$

$$Df = Nf / Nw \quad (\text{式 13})$$

表 18 に、各ゾーンにおける品質評価、及び判断と対策を示す。

表 18 ゾーンの意味

| ゾーン   | 品質評価             | 判断と対策                               |
|-------|------------------|-------------------------------------|
| ゾーン 1 | 良い               | 問題なし                                |
| ゾーン 2 | 良い               | レビュー方法（指摘内容及びレビュー人数や回数）を確認する        |
| ゾーン 3 | 良い               | レビュー方法を確認後、追加レビューを検討する              |
| ゾーン 4 | 良い               | レビュー方法の妥当性を確認後、追加レビューを検討する          |
| ゾーン 5 | 悪い               | 指摘内容を確認し、追加レビューを検討する                |
| ゾーン 6 | 悪い               | 成果物の記載内容を再確認し、記載レベルが低くなければ追加レビューを行う |
| ゾーン 7 | 良い               | 指摘内容を確認し、追加レビューを検討する                |
| ゾーン 8 | 悪い               | 成果物の記載レベルが低ければ、記載の見直しを行う            |
| ゾーン 9 | 判断保留（追加レビュー後に判断） | 追加レビューを行う                           |

ゾーン分析法は、レビュー結果に基づく成果物品質の判断を行う方法として、以下の特長をもつ。

- ① レビュー会議で通常収集されるデータ（レビューにかけた延べ時間と、それによって検出された欠陥数）だけから評価を行うことができる。
- ② 同一平面上で複数のレビュー対象を比較して見ることで、同一プロジェクト内の均質な条件下で、同一作成担当者での複数対象及び作成担当者間の対象の相対比較により、特異点を見出すことができる。
- ③ 同一レビュー対象に対する複数回のレビューを必須としない。
- ④ 複数のレビュー結果を統合して評価することや、累積で経時変化をプロットして評価することもできる。

#### 4.2.2 ゾーン分析法の課題

ゾーン分析法には、次の二つの課題がある。

課題① ゾーンの設定の仕方： ゾーンを決定するのは、レビュー密度と検出欠

陥密度の計画値である。計画値として、組織実績値（あるいは過去の類似プロジェクトの実績値）を使うことが多い。そのため、過去のデータの集積が少ない場合には、あるいは対象プロジェクトの特殊性（例えば、開発体制・要員や流用率の変化）がある場合には、品質判断の信頼性が低くなってしまふ。

課題② 個々のレビュー実施に対する考慮：ゾーン上にプロットされる点は、個々のレビューの実績値を表す。実績値は規模で正規化されるため、対象に対する均一なレビューの実施が前提とされている。しかし、現実には、大量の成果物を対象にしていること、人とチームに依存して実施の仕方が異なることにより、実施方法と欠陥にバラツキが生じており、それに対する考慮がない。

課題①に対しては、過去の実績値が蓄積することで徐々に軽減されてくる。また、過去の実績がなくても、特長②を用いて実施の特異点を見つけることや定性的なデータを併用することで判断を補強し利用することができる。しかし、課題②に関しては、レビュー実施のバラツキを防止するために、品質管理者が直接レビューに参加し状況を把握することが対策として考えられるが、現実にはリソース制約上網羅的に実施することは困難である。そのため、ゾーン分析による品質評価の有効性を損なうという意味で、課題②がより重要な問題となっている。

#### 4.2.3 本研究が扱う課題

本研究は、ゾーン分析の課題②に関して、検出効率のバラツキ以外の原因を明らかにし、その解決方法を提案するものである。本研究で扱う課題について以下に記述する。課題②が生じる例を以下で詳しく述べる。図15は、ゾーン上に4つの機能仕様書のレビュー密度（工数／ページ）と検出欠陥密度（件／ページ）の実績値をプロットしたものである。

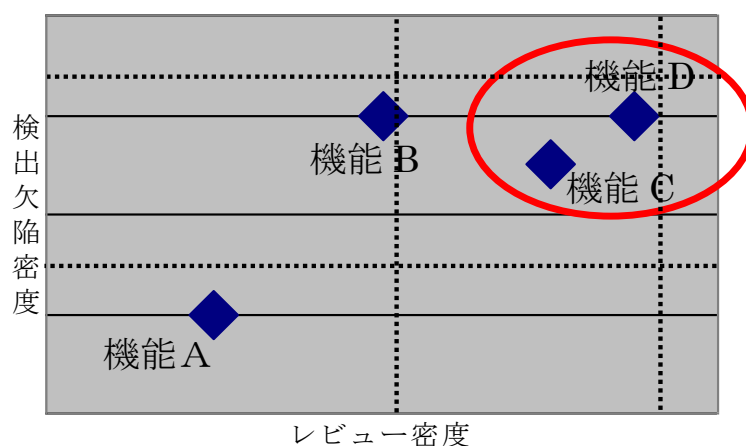


図15 従来のゾーン分析手法の適用实例



図 15 では、機能仕様書 A に対しては、ゾーン 9（レビュー密度及び検出欠陥密度が低い）に入るため、追加レビューを行うという対策を実施する。仕様書 B に関しては、ゾーン 7（レビュー密度が低いにも関わらず欠陥検出密度が高い）に入るため、指摘内容の点検を行うという対策をとる。

問題は、機能仕様書 C と D である。これらは、ゾーン 1（レビュー密度と欠陥検出密度が目標値を満たしている）に入るため問題なしと判断されるが、実際にはテスト段階で検出された欠陥が多く、その中に明らかにレビュー検出漏れであると判断されるものが多く含まれているケースがあった[2-3]。

レビュー検出漏れが生じる原因の一つとしては、レビューの検出効率のバラツキがある。これは、レビュー活動中に成果物の説明や設計修正自体を実施し、欠陥検出の時間を掛けていないことが原因であった。この問題を解決するため、レビュー会議の実施方法を見直すことで解決を図ってきたが、改善後においても、レビュー検出漏れを要因とするテスト段階で検出される欠陥数は減少しない場合があった[2-3]。

### 4.3 ピアレビュー評価手法を用いた品質評価技法の提案

前章で述べた問題点を解決することを目的に、成果物のピアレビューにおいて、レビュー実施漏れを防止し、テスト段階への流出欠陥数を削減するためのピアレビュー評価技法を提案する。

提案する品質評価技法は、ピアレビュー網羅度と呼ぶ指標を導入することで、ゾーン分析を実施する際に、実施したピアレビューが対象成果物をどの程度網羅しているかを定量的に評価できるようにするものである。以下、提案する新しい指標とそれを利用した品質評価技法について述べる。

#### 4.3.1 事前調査

##### 4.3.1.1 レビューでの欠陥検出パターン

図 15 のゾーン分析の結果において、品質良好と判断された機能のうち、テスト段階で欠陥が多発したものと、そうでなかったものを、レビュー記録の確認することで、比較分析を行った。レビュー記録は、指摘毎に、①ページ番号、②指摘内容、③指摘者、④重要度（大/中/小）、⑤回答、⑥回答者、⑦確認者を記載している。

図 16 と図 17 は、単一の成果物に対して、複数レビュー者がピアレビューを行った結果から、ページ毎の指摘件数を集計して示したグラフである。図 16 はテスト段階でレビュー実施漏れの欠陥が多かった機能部分、図 17 は欠陥が少なかった機能部分に関するものである。同一指摘は欠陥数 1 件としてカウントした。図 16 は、成果物の前半部分に多くの指摘があるが、後半部分ではレビュー指摘件数が 0 件のページが

多いことを示している。これに対して、図 17 は前半から後半まで満遍なく指摘が上  
がっていることを示している。

この分析結果から、欠陥流出の原因は、レビュー対象の後半部分に欠陥検出数が少  
ない場合に多く、この場合後半部分に欠陥が多く残存している可能性が高いことがわ  
かった。対象成果物に混入した欠陥は成果物全体に均質に存在すると仮定すると、レ  
ビュー対象の後半部分に欠陥検出数が少ない原因は、レビュー実施が不足しているか、  
検出効率が落ちているかのどちらかである。しかし、前半と後半部分に対するレビュー  
は読み進む順序以外に差異が見当たらない。以上のことから、我々は、成果物の後  
半部分が前半部分に比して検出件数が少ない原因が、前半で多くのレビュー時間を費  
やしたため後半のレビューにかかる時間が不足したことにあるのではないかと推定し  
た。

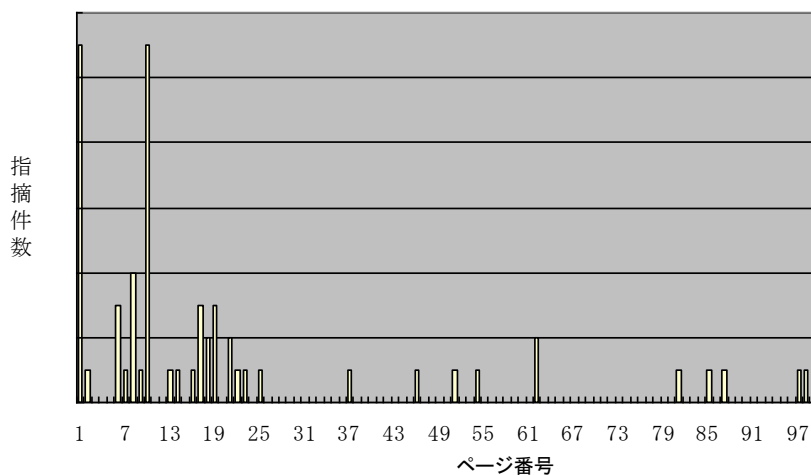


図 16 ページ番号と指摘件数の関係（流出欠陥数の多かった機能）

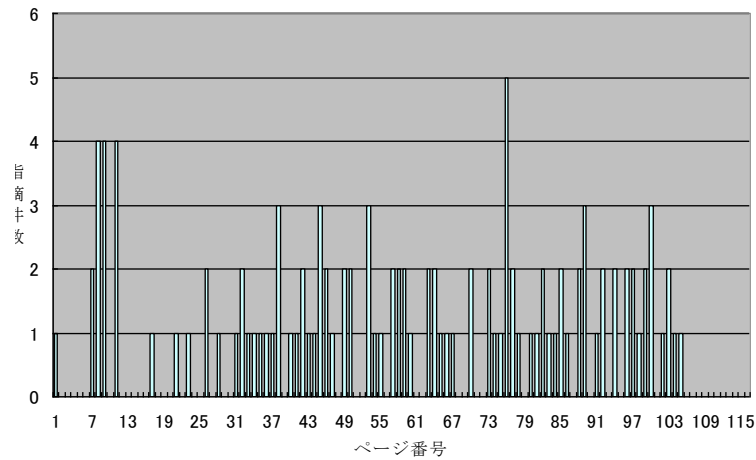


図 17 ページ番号と指摘件数の関係（流出欠陥数の少なかった機能）

上記の問題点は単純なゾーン分析法による評価では顕在化できない。実際、図 16 と図 17 の 2 つのケースにおいて、欠陥の多かった機能及び欠陥の少なかった機能のレビュー指摘密度、検出欠陥密度はそれぞれ下限値上限値の範囲内であり、両機能とも問題なしと判断されている。

このことは、ゾーン分析法が成果物全体に対するレビュー密度という平均化された情報でのみ評価を行っていることに原因がある。例えば、特定ページで多く時間を費やし多くの指摘を行い、その他のページでレビューがされなかった場合、全体的には欠陥検出密度が目標通りという評価となってしまふからである。そして、これは非熟練者による作成された成果物に多く見られるパターンであり、品質管理上見逃してはいけないケースである。本研究はこの問題点を解決する技法を提案する。

#### 4.3.2 作業成果物への混入欠陥の分布

図 18 は、テスト段階への欠陥流出の少なかった上位 13 の機能を選択し、その設計成果物を対象に、設計段階のレビューにおいて、成果物上のどの位置で欠陥が検出されたかを示すグラフである。横軸は、成果物全体の頁を 100%とした際の欠陥のある頁の相対位置である。縦軸は、相対頁 10%刻み毎に成果物の検出欠陥数を足し合わせたものを、相対件数で表している。

図 18 に示すように、テスト段階への欠陥流出の少ない場合において、検出欠陥が成果物の前半と後半に偏在せず成果物全体にわたっている。このことから、対象とするプロジェクト組織においては、設計成果物の混入欠陥が、レビュー対象の頁全体にわたり存在し、かつ前後半に偏在せず分布する可能性が高いことがわかる。

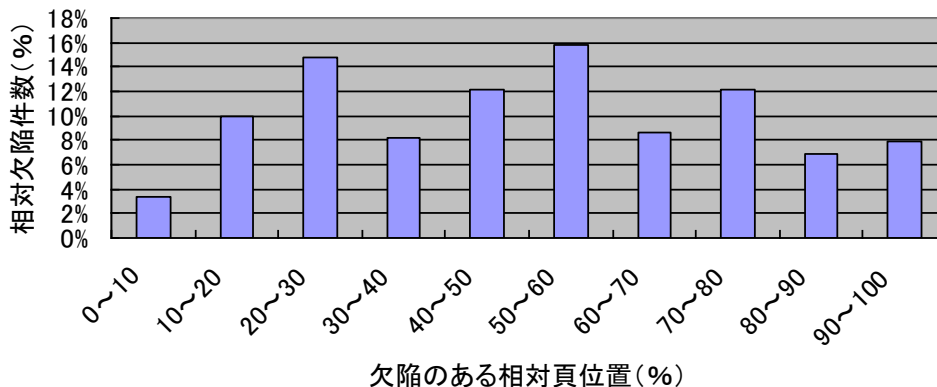


図 18 レビュー指摘の成果物

## 4.4 提案技法

### 4.4.1 ピアレビュー網羅率

一般には、ピアレビューが対象物に対して万遍なく実施されたかどうかを測定することは難しい。レビュー対象部分（例えばページ）に対して直接レビューが行われた時間を測定できれば直観的であるが、実際のレビューでは対象を頻繁に変更させることが多く、どこを対象としているかを明確に区別できないため、その時間計測は困難である。

そこで、我々は、（ピアレビュー結果である）個々の指摘情報から取得可能なデータを使って計算するピアレビュー網羅度と呼ぶ指標を、レビュー実施の網羅性を表す代替特性として採用した。ピアレビュー網羅率は、ページ毎の指摘件数の有無から、対象成果物に対するピアレビュー実施のバラツキを評価する。定義を次式に示す。

$$D = \sum (D_i) / N \quad (\text{式 14})$$

ここで、 $D_i$ : ページ  $i$  の指摘有無（有は 1、無は 0 で示す）

$i$ : ページ番号、 $N$ : 総ページ数

上式で示すように、ピアレビュー対象成果物のページ毎の指摘が有なら 1、無なら 0 を割り当てる。これによりそのページがレビューされたかどうかを判断している。この値を対象文書の全ページにわたり総和を取り、総ページ数で割ったものを、成果物のピアレビュー網羅率とする。例えば、10 ページに 3 件の指摘があった場合において、あるページに 3 件の指摘が集中していれば網羅率は 0.1 となり、指摘が 3 ページに分散していれば網羅率は 0.3 となる。

#### 4.4.2 ピアレビュー網羅率の分割

ピアレビュー網羅率は、ピアレビューが対象成果物の全体に対してレビュー実施の網羅性を表す代替特性として定義した。さらに、ピアレビュー網羅率を成果物の前半部分のページに対するものと後半部分のページに対するもので分離することで作業成果物後半におけるレビューの網羅性を評価するために2つの指標、ピアレビュー網羅率（前半） $Df$  及び ピアレビュー網羅率（後半） $Db$  を以下のように定義する。

$$Df = \sum (Di) / (M) \quad i=1 \dots N/2 \quad (\text{式 } 15)$$

$$Db = \sum (Di) / (M) \quad i=N/2+1 \dots N \quad (\text{式 } 16)$$

ここで、 $N$ : 総ページ数

#### 4.4.3 ゾーン分析とピアレビュー網羅率を組合せた品質評価技法

我々が提案する品質評価技法は、ピアレビュー網羅率（後半）の情報をゾーン分析法のグラフに付加することで、対象の網羅度に対する品質判断の観点を補うものである。図 19 は、提案技法におけるレビュー結果の評価に用いるグラフの表示例である。図 19 において、レビュー対象は（従来のゾーン分析法と同様に）レビュー密度と検出欠陥密度のゾーン上にプロットされ、各プロットはその大きさがピアレビュー網羅率（後半）を表す。ここで、ピアレビュー網羅率（前半）とピアレビュー網羅率（後半）の差ではなく、ピアレビュー網羅率（後半）の絶対値を用いたのは、成果物の後半部分で時間切れによりピアレビューが不十分となる可能性が高いという仮定に基づいている。

個々のレビュー結果は、従来のゾーン分析法の評価にレビュー実施の網羅性の観点からの評価を加えて品質判断される。図 19 の例において、矢印の成果物は、ゾーン分析法ではゾーン 1 に入るため品質が良いと判断されるが、ピアレビュー網羅率（後半）が 0.3 であることから、後半部分でのレビューが不十分であることがわかり、追加レビューが必要であるとの品質判断を下すことができる。なおピアレビュー網羅率の閾値 0.3 は、検出欠陥密度の組織実績値を参考に算出した。

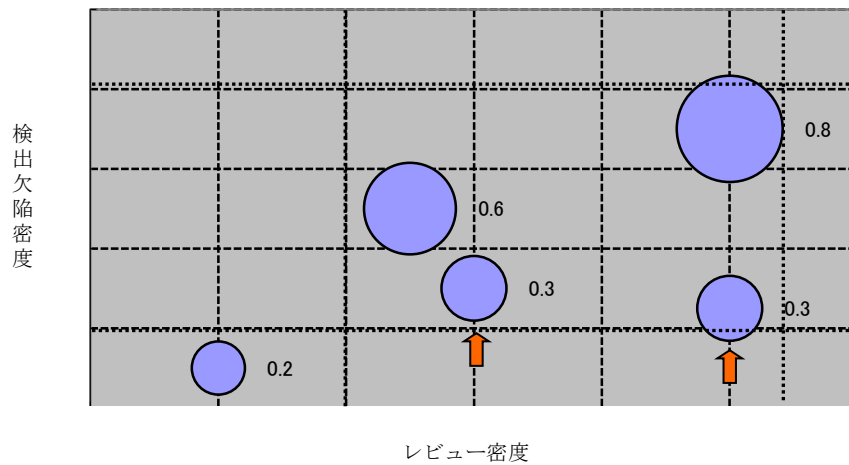


図 19 ピアレビュー網羅率を用いたゾーン分析

## 4.5 適用と評価

提案技法を実サブプロジェクトに適用し評価を実施した。レビュー対象成果物はソフトウェア要求仕様書（総ページ数 500 超），レビュー者は 10 名であった。

### 4.5.1 ピアレビュー網羅率の評価

過去のレビュー結果から，レビュー対象成果物別に算出したピアレビュー網羅率を表 19 に示す。なお，表 19 では総ページ数が 2 ページ以下のものは含めていない。

表 19 から，ピアレビュー網羅率に関して以下のことがわかる。  
 総ページ数が少ない場合，時間切れによるレビュー実施漏れの発生は少ない。  
 成果物の総ページ数が多くなると，網羅率に差異が表れ，レビュー実施漏れとの相関を検討できる。

表 19 過去の成果物に対するレビュー網羅率

| 成果物<br>番号 | 総ページ数<br>( $N$ ) | 指摘件数 | 網羅率<br>( $D$ ) |
|-----------|------------------|------|----------------|
| 1         | 106              | 91   | 0.53           |
| 2         | 106              | 139  | 0.50           |
| 3         | 61               | 76   | 0.46           |
| 4         | 40               | 51   | 0.58           |
| 5         | 24               | 39   | 0.96           |
| 6         | 4                | 9    | 0.75           |

ピアレビュー網羅率とレビュー実施漏れとの相関を見るために，成果物の総ページ数とレビュー網羅率の相関を分析した結果を図 20 に示す。サンプル数は少ないが，

図 20 から、総ページ数が多くなるほどピアレビュー網羅率は低下し、レビュー実施漏れになり易いという可能性が考えられる。

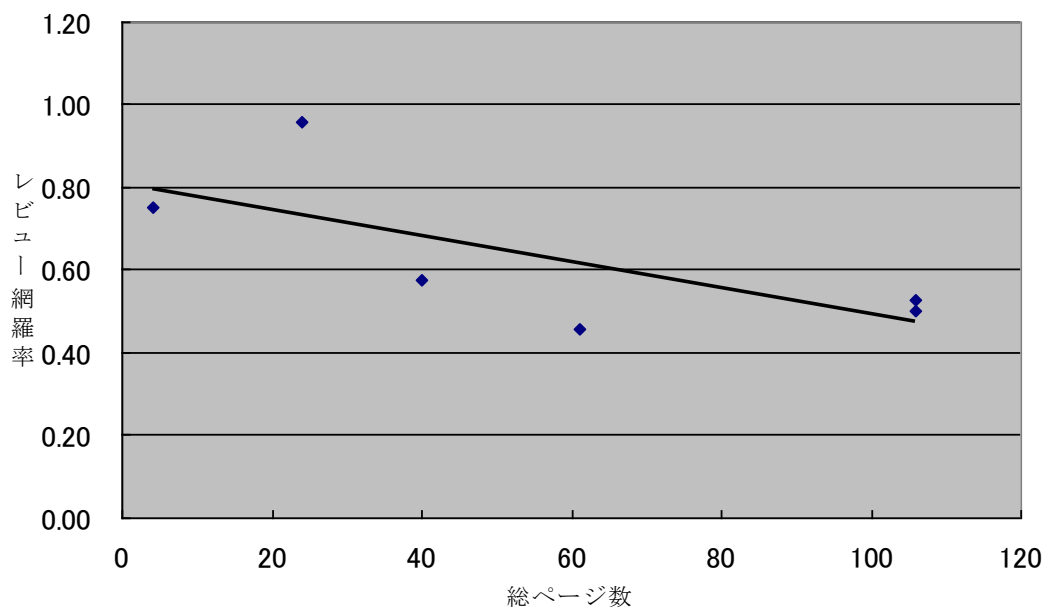


図 20 総ページ数とレビュー網羅率 ( $D$ ) の関係

#### 4.5.2 分割したピアレビュー網羅率の評価

総ページ数の多い成果物 1, 2, 3, 4, 5 に対して計算した結果を表 20 に示す。

表 20 から、成果物全体でのピアレビュー網羅率  $D$  より、 $Df$  と  $Db$  の関係より、成果物の後半部分でのピアレビュー網羅率が低く、後半での漏れが発生し易い状況であることがわかる。

表 20 成果物 1, 2, 3, 4, 5 に対するピアレビュー網羅率

| 成果物番号 | 総ページ数 ( $N$ ) | 指摘件数 | 網羅率 ( $D$ ) | 網羅率 (前半) ( $Df$ ) | 網羅率 (後半) ( $Db$ ) |
|-------|---------------|------|-------------|-------------------|-------------------|
| 1     | 106           | 91   | 0.53        | 0.32              | 0.21              |
| 2     | 106           | 139  | 0.50        | 0.35              | 0.15              |
| 3     | 61            | 76   | 0.46        | 0.33              | 0.13              |
| 4     | 40            | 51   | 0.58        | 0.43              | 0.15              |
| 5     | 24            | 39   | 0.96        | 0.50              | 0.46              |

また、図 21 は、成果物総ページ数と後半網羅率  $Db$  の関係を示したものである。

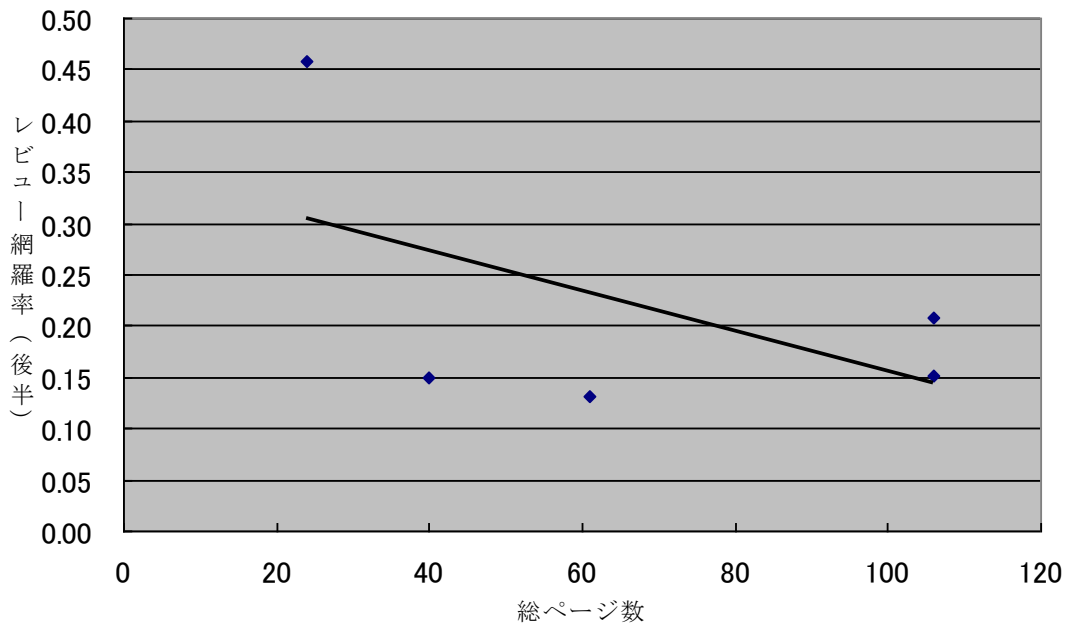


図 21 総ページ数と後半レビュー網羅率 ( $Db$ ) の関係

図 21 から、成果物の総ページ数が一定規模を超えるとピアレビュー網羅率（後半）の値が小さくなることがわかる。このことから、前半にレビュー時間を多くかけ、後半のレビューが十分に時間をかけなかった可能性がある。

#### 4.5.3 ゾーン分析とピアレビュー網羅率を組合せた品質評価技法の評価

レビュー結果の品質評価は、ゾーン分析を基本として、本技法により品質判断を補うことで実施した。具体的には、以下のようにピアレビュー網羅率を利用した品質判断と対策を行った。

レビュー工数が目標に満たない場合、ピアレビュー網羅率を確認する。

ピアレビュー網羅率の前半と後半で差があれば、成果物の後半部分でレビューが不足していると判断する。

後半部分に対して、短時間の追加レビューを行う。

追加の指摘件数を加えてレビュー網羅率を評価する。

##### (1) 適用結果と評価

図 22 は、1つのプロジェクトにおける成果物のレビュー実施結果に対するゾーン分析結果である。対象としたプロジェクトは、14個のサブプロジェクトで構成され、1つのサブプロジェクトは、2名から5名程度の独立した開発者により開発を実施



している。1つの丸が1つのサブプロジェクトの作業成果物、丸の中の数値は、ピアレビュー網羅率（後半）を示す。図中の数値の付いている部分は、従来のゾーン分析では品質が良いと判断される領域にあり問題なしと判断される。しかしピアレビュー網羅率（後半）が0.27と低いため、後半のページ部分でレビュー実施漏れが発生していると判断され、当該機能部分に対して追加レビューを実施した。その結果、新たに欠陥が発見された。結果として図22の矢印の位置へ移動し、ピアレビュー網羅率が0.27（レビュー密度：0.36，欠陥検出密度：0.43）からピアレビュー密度0.60（レビュー密度：0.65，欠陥検出密度：0.71）へ向上した。また別成果物において、レビュー網羅率（後半）が0.1（レビュー密度：0.67，欠陥検出密度：0.53）から0.36（レビュー密度：1.29，欠陥検出密度：0.73）へ向上した。

このように、提案する品質評価技法の適用により、ピアレビュー網羅率（後半）の低い作業成果物を検出することにより、後半部分に対するレビュー実施漏れのある成果物を見逃さず、追加レビューを実施させる管理が可能となった。

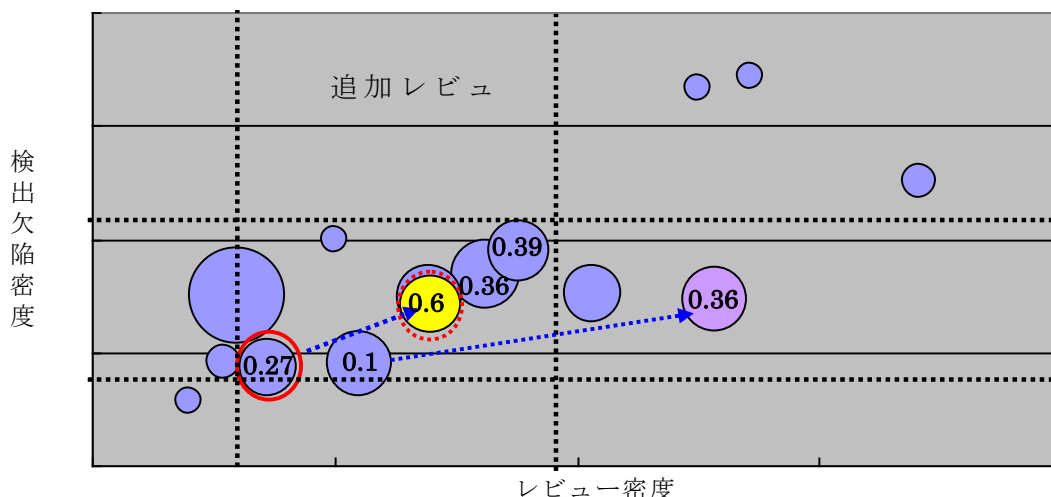


図22 ピアレビュー網羅率（後半）付ゾーン分析結果

結果として、適用サブプロジェクトのテスト段階での欠陥検出数を過去実績比で34%削減できた。ここで、比較対象の過去プロジェクトは、本技法の適用を除き、同様な開発体制、スタイル、及び品質基準で開発を実施している。このため、コーディング差異などによる欠陥の防止効果などの影響は小さいと考えている。

また、開発者へのヒアリング結果から、本技法を用いることで、ピアレビュー網羅率を用いて追加レビューを行った場合、後半部分に重点的なピアレビューを行うことができ、欠陥検出を効果的に実施できる点が効果として挙げられた。

## (2) 見出された課題

適用プロジェクトにおいて、ケース数は少ないが、ピアレビュー網羅率が低い  
ため追加レビューを実施したのに新たに欠陥が検出できない場合があった。これは、  
ピアレビュー網羅率が、成果物の潜在欠陥数が元々少ない場合でも、低い値をとるこ  
とが原因であった。

ピアレビュー網羅率を用いて追加レビューの要否を判断する際には、その絶対値  
が基準値を満たすか満たさないかで単純に評価せず、ピアレビュー網羅度の前半対後  
半比が高いかどうかなどを判断材料として加えることが必要であると考える。

## (3) 技法の適用範囲と限界

本技法は、1回のレビューにおいて対象成果物を前から後ろへ読み進むことを前提  
としている。そのため、部分に分けて均等に時間配分を行うやり方や、レビュー対象  
をサンプリングするやり方のレビューには適用できない。

なお、本技法は、上記前提を満たせば、レビューを成果物の途中から読み始める  
場合にも適用可能である。ピアレビュー網羅率の計算では、開始ページ番号を  $m$ 、レ  
ビュー終了ページ番号を  $n$  とし、(式 14)から(式 16)の  $N$  に以下を当てはめればよ  
い。

$$N = n - m \quad (\text{式 17})$$

また、本技法は、レビュー対象成果物に対して均質に欠陥が分布していることを  
前提としている。このため、例えば、前半部分に欠陥が偏重する場合においてもピア  
レビュー網羅率が前半でより高くなってしまい、追加レビューの判断を誤る可能性が  
ある。

この前提については、4.3.2 章において、この開発現場において過去の実績が欠陥  
の前後半での均質さを裏付けていることを示した。しかし、レビュー対象成果物の部  
分毎に、成果物の複雑度が異なる場合、作成担当者あるいは掛けた設計工数が異なる  
場合などは有意差が現れる可能性があるため、技法適用時に考慮する必要がある。

## 4.6 関連研究

ピアレビューの実施結果から品質を評価する技法として、4.2 章で述べたゾーン分  
析法以外に、検出欠陥数の実績から残存欠陥数を推定する欠陥数推定法がある。以下  
に、欠陥数推定法のアプローチ及びレビューの品質評価への適用上の特徴と制約をま  
とめ、提案する技法との関係について考察する。

### 4.6.1 欠陥数推定法とレビュー評価への適用上の特徴と制約

#### (1) Capture-Recapture モデル [4-3] [4-8]

同一対象・同一観点での複数人による独立したレビューの結果から残存欠陥数を

推定する技法である。ある地域での野生動物の生態数を求める統計手法を適用し、地域をレビュー対象に野生動物を欠陥に置き換えて、レビュー対象に含まれる総欠陥数を推定する。レビュー者 1 とレビュー者 2 により検出された欠陥をそれぞれ  $n_1$ ,  $n_2$  とする、そのうち共通の欠陥を  $m$  とすると総欠陥数は次式で計算できる。

$$N = n_1 \cdot n_2 / m \quad (\text{式 18})$$

残存欠陥数は次式で計算する。

$$R = N - (n_1 + n_2 - m) \quad (\text{式 19})$$

Capture-Recapture モデルには、レビュー者の検出能率を可変と考える推定法 (Mt)、欠陥の検出確率を可変と考える推定法 (Mh) などのバリエーションがある。Vander Wiel ら [4-4] はシミュレーション結果から、Mt が Mh に比較してよい推定結果を得ること、Mt が欠陥クラスを分類することにより推定結果に改善がみられることを示した。Briand ら [4-3] は実適用のデータに基づき全バリエーションを評価し、以下の結果を報告している。

- ① 4 以上の独立レビューを実施しないと良い推定値を得られない。
- ② 推定値 (中央値) が低く出る傾向がある。これを調整するバリエーションは安定性が悪い (極端に高い推定値を出すことがある)。

現場での実運用を考慮すると、①を常に前提にすることは適用上難しい。②は残存欠陥数を過少評価することになり、再レビューの必要性を判断する目的には合わない。

## (2) 信頼度成長モデルによる外挿法

レビュー中に検出した欠陥のタイムスタンプを使い、レビュー時間に対する累積欠陥検出数をプロットし信頼度成長曲線にフィッティングすることで残存欠陥数を推定する技法である。Goel らは、信頼度成長曲線として Goel-Okumoto のモデル [4-5] を採用している。Goel-Okumoto のモデルは、レビュー時間  $t$  での累計検出欠陥数を  $D(t)$ 、検出能力を  $k$  とすると、次式となる。

$$D(t) = N(1 - e^{-kt}) \quad (\text{式 20})$$

総欠陥数  $N$  と検出能力  $k$  は実データを使って推定する。Goel-Okumoto のモデルは、レビュー者グループの検出能力は一定と仮定し、レビュー時間を経るとレビュー対象成果物は残存欠陥数が減少するため検出効率 (件/h) は指数関数的に減少していくものとしている。

レビューにおいて、検出能力を時間にわたって一定とする前提は適切ではない。実際のレビューは、複数人が並行して進める事前準備段階と、レビュー者が集まり会

議形式で行う段階からなることが多いが、事前準備段階は正確な実施時間を計画・把握することはできないため、タイムスタンプをとることで検出能力一定の前提をおくことができない。また、会議形式においてもレビュー参加者が変化することがあるためこの前提は崩れており、この技法をレビューの品質評価に用いる場合には、こうした実施状況に注意してデータを取り扱う必要がある。

なお、山田は[4-8]は、テストプロセスでの検出能力の時間的変化を考慮した信頼度成長モデルとして習熟S字形モデルを提案しているが、レビューへの適用は報告されていない。

### (3) 欠陥プロファイル推定法[4-7]

複数レビュー者が同一対象を独立してレビューする。検出された欠陥を、その欠陥を検出したレビュー者の数で値づけし、その値の多い順にソートする。

図 23 に示すように、指数関数にフィッティングし、閾値 0.5 を設定してそれ以下になった欠陥番号が総欠陥数の推定値とする方法である。

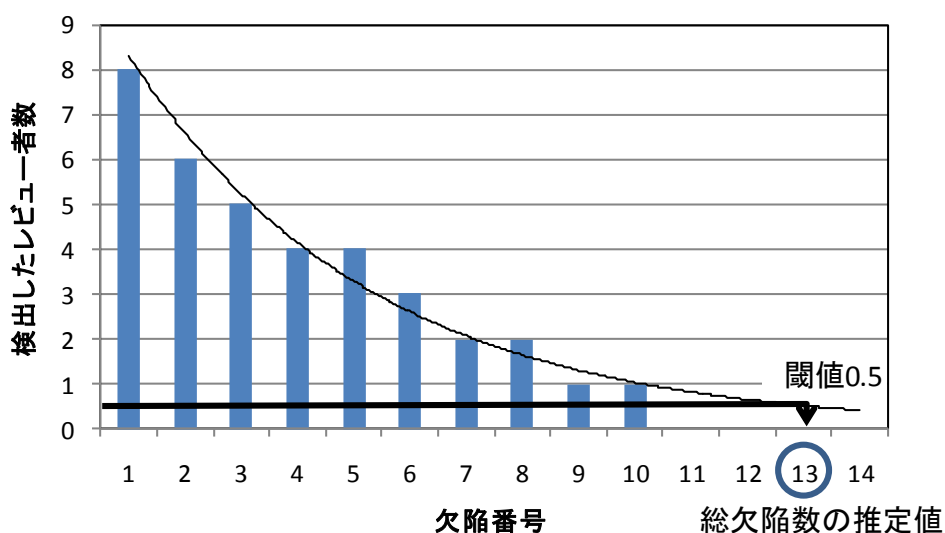


図 23 欠陥プロファイル推定法

Biffi ら[4-7]は、欠陥プロファイル推定法が、リーディング法の影響も受けにくく、また平均誤差及び標準偏差で安定した結果を得ることができたと報告している。しかし、この推定法では、指数関数へのフィッティングを用いているが、指数関数を選択する理論的根拠は示されていない。その点が信頼度成長曲線の場合と大きく異なる。

#### 4.6.2 欠陥数推移法と提案技法との関係

4.6.1 で述べたように、Capture-Recapture モデルは種々の制約から追加レビューの必要性を判断する目的には利用しにくい。これに対して、信頼度成長モデルによる外挿法及び欠陥プロファイル推定法は、制約事項はあるものの、この目的に利用することが可能である。

しかしながら、これらの欠陥数推移法は、ゾーン分析法と同様に、成果物全体に対する平均化された情報でのみ評価を行っている。このため、レビュー実施の不均一さを見逃してしまう可能性がある。我々が提案するピアレビュー網羅率を用いることで、これらの欠陥数推移法の収束判断を補完できると考える。特に、信頼度成長モデルによる外挿法は、レビューの進捗に伴う基準値付近での推定残存欠陥数の変化（つまり収束具合）を判断に用いるが、これにピアレビュー網羅率（及び前半対後半比）の変化を見ることで、レビュー実施の不均一さがなく見えることができると考える。

### 4.7 まとめと今後の課題

本章では、ピアレビューの結果に基づく従来の品質評価技法が、成果物全体に対する平均化された情報のみに基づいているため、レビューが均質に実施できていない場合に欠陥を多く見逃してしまうという問題を明確にした。この問題を解決するために、レビュー実施の均質さを測定するための指標として、ピアレビュー網羅率を定義し、その指標を用いて従来技法を補完する品質評価技法を提案した。本指標を、レビュー網羅率（前半）とレビュー網羅率（後半）に分離することで、後半部分のレビュー実施漏れを推定することができる。さらに、提案技法によるレビューの品質管理を実プロジェクトへ適用し、レビュー実施漏れを防止しテスト段階への欠陥流出を減少する効果を確認し、提案技法の有効性を示した。今後は、ピアレビュー網羅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、ピアレビュー網羅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。



# 第5章 むすび

## 5.1 まとめ

本論文では、開発現場でピアレビューを実施しているが、ピアレビューを完了してもテスト段階に流出する欠陥が多いという課題を解決する目的で、ピアレビューに関する以下の技法を提案した。

1. ピアレビュー会議の有効性を高める改善手法
2. レビュー会議の有効性評価技法
3. ピアレビュー網羅率を用いた品質評価技法

1. については、実際に開発現場で実施しているピアレビューの実態の調査を行った。調査に当たり、ピアレビュー会議で実施している活動を調査する測定ツールを開発し、発言内容、発言者を記録し分析した。その結果、ピアレビュー会議が仕様説明を中心とした活動・設計活動を中心とした活動・欠陥抽出を中心とした活動の3種類に分類できることを特定し、さらに“ピアレビュー有効時間比率”と呼ぶ新たな指標とテスト段階へ流出する欠陥数を削減するピアレビュー会議の実施方法を定義し、実開発で適用し品質向上の効果を得た。また、これらの新たな技法の導入効果を評価するため、技法適用前後のピアレビュー有効時間比率・単位時間当たりの指摘件数を統計的に分析し、2つの活動に統計的な差異があることを示した。

2. については、これまでピアレビュー会議の有効性について疑問視する論文が発表されてきたことから、その解決を目的に研究を行った。Porterらは、ピアレビュー会議において検出できる欠陥数はわずかであり、FaganやGilbらの主張してきたピアレビュー会議が有効であるという結論と反する実験結果を示した。我々は、Porterらの主張が産業界における評価との間にギャップがあるという認識であり、実験を通じてピアレビュー会議を評価した。その結果、我々の実験におけるMeeting-Gainが21.8%であり、Porterの結果である9%に比べ高いことを示した。2つの実験の結果の差異の原因として、(1)個人チェック時間の設定方法、(2)チーム編成及びレビュー会議運営方法、(3)成果物の違いによる要因、(4)各レビューアのチェック観点設定であると結論付けた。特に(1)の個人チェック時間が成果物全体をチェックするのに十分でない場合、ピアレビュー会議を実施しても個人チェックで検出した欠陥以外の欠陥を検出することは困難であると評価した。

3. については、ピアレビュー会議を適切に実施し、ゾーン分析法で適切に結果を評価しても、テストへ欠陥が流出する可能性があることから、その原因を特定しテストへの欠陥流出を防止することを目的に研究を行った。ゾーン分析法は、ピアレビュー

に投入した工数と検出した欠陥数の関係から、ピアレビュープロセスの適切さを評価し、その結果として成果物の品質を判断する手法である。ゾーン分析法の課題は、ゾーンの決定の難しさとピアレビュー実施に対する考慮の2点である。ゾーンの決定に関しては、実績を蓄積することで、その製品や組織に最適なゾーンを決定することができる。一方、ピアレビュー実施において、成果物に渡り、どのようなピアレビューを実施したかを把握することは困難であり、その結果として成果物の一部分でピアレビューができていなかったにも関わらず、成果物全体に対する投入工数と欠陥数が計画値（ゾーンとして表現される）を満たしていると判断することがある。特に問題となるのが、成果物の一部分をレビューするのに時間がかかり、全体をレビューできないまま、実施しているレビューアがそれに気付かずに、ピアレビューを完了してしまう場合である。このような問題を解決するために、ピアレビューが網羅的に成果物全体をチェックしたかを評価する“ピアレビュー網羅率”という新たな指標を導入した。その結果、従来のゾーン分析法に“ピアレビュー網羅率”を付加し、新たなピアレビュー評価技法を定義し、テスト段階への欠陥流出を抑制することができた。

このようなピアレビューの効果と効率を高めることの必要性は、次の通りである。今後の高信頼化ソフトウェアでは、最終検査工程だけで品質要求を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するため、ピアレビューが重要である。さらに検証段階では、要求定義書や設計書に記載された事項が正しく実現できているかという視点で検証を行うため、要求定義や設計における“暗黙の了解”、“行間”と呼ばれる文書化されていない事項を発見するためにもピアレビューが有効であり、ピアレビューなくして欠陥のない製品を作ることは困難である。

さらにフロントローディングにより、手戻り工数を削減するためにも、検証段階へ流出する欠陥数を減らすことが重要であり、その主たる活動がピアレビューである。テストに関しては、テスト戦略、テスト計画、テスト設計、テストツール、同値分割法、境界値分析法、異常値・特異点分析法、テストカバレッジ、直交表、All-Pair法、HAYST法、状態遷移テスト、探針テストなど様々な手法が提案され実践されてきた。一方、レビューに関する技法については、レビュー戦略、レビュー設計などは開発現場では検討しておらず、研究としても不十分である。仕様検討から設計、試験まで、全ての作業成果物に対して、必ず実施すべきレビューについて、管理技法、実施手法を強化し、高信頼なソフトウェアを効率的に開発できるようにすべきである。

本論文で示した内容は、高信頼なソフトウェアを効率的に開発できるレビュー技法について、実開発の現場と共に検討し、試行錯誤し、プロセスを変更し、その効果を確認したものである。つまり、一部の製品における品質向上と開発効率化に対しては、本研究による事業への貢献を実現できた。一方、今後製品をグローバル展開し、事業として発展するという視点では1つの製品への貢献だけでは不足している。すな



わち、実践し有効性を確認できた技法をどのように全社的に展開するかが企業発展の分かれ目である。その視点から、本論文で示した技法を全社的なレビュー標準プロセスとして纏めた上で、全社的な教育である『ソフトウェアプロジェクトリーダー育成コース』[5-1]のテキストとして整理し、技法の全社展開に活用している。『ソフトウェアプロジェクトリーダー育成コース』の受講生は、様々な製品を開発する製作所から選抜されたメンバであり、講座で習得した技法・標準プロセスを利用し、製品特性に合わせてプロセスに変更し活用している。今後も、レビューに関する技法を改善し、その結果を全社的に展開・活用することで、企業の発展に貢献するつもりである。

## 5.2 今後の研究方針

近年レビューチームの構成により、指摘件数が変化するという研究成果が発表された。Goswami [5-2]は、個人の学習スタイルとして8つの Learning-Style を変数として、ピアレビューの効果と効率を測定した。その結果、Active-Sensing-Sequential な Learning-Style を有するレビューチームが最も効果的・効率的に欠陥を検出できることを示した。本研究成果を利用し、製品特性毎のピアレビューチーム構成メンバの最適化やピアレビューチーム構成メンバの教育に活用できる。今後、本 Learning-Style を利用した研究に取り組み、より効果的なピアレビューを実施できるような研究を実施する予定である。

また、ピアレビュー会議時間をどれくらい確保すれば良いかという課題については、これまでも議論されてきた。Seaman は、コードレビューにおいて、ピアレビュー会議時間と検出できる欠陥数の関係を組織構造から議論し、ピアレビュー会議の出席者同士が良く知っている場合と、地域的に離れている場合では、検出できる欠陥数に差異があると指摘した[5-3]。今後、このような心理的な問題もピアレビューの効果を上げる重要な課題であると考えられる。

さらに、欠陥の関連を図示した欠陥連鎖チャートを用いたレビュー方法の提案[5-4]、医療における問診を参考に、設計者に対し問診により重大欠陥の検出に集中するレビュー法[5-5]、レビューリーダーが過去のレビュー結果を収集し、その知見をまとめ、以降のレビューに反映する手法[5-6]など様々な手法が提案されている。これらの手法に関しては、これまでのレビューチェックリストやシナリオを用いたレビュー技法の延長線上にあり、その効果と適用するための条件などを検討する必要がある。今後の研究の対象と考えている。



## 参考文献

- [1-1] 情報処理推進機構（IPA）ソフトウェア高信頼化センター（SEC），高信頼化ソフトウェアのための開発手法ガイドブック，pp.3-6，2010.
- [1-2] S.Kan, Metrics and Models in Software Quality Engineering, Addison-Wesley, 2002.
- [1-3] 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.
- [1-4] K.Wiegers, Peer Reviews in Software:A Practical Guide, Addison-Wesley, 2002.
- [1-5] M.Fagan, Design and Code inspections to reduce errors in program, IBM SYSTEMS JOURNAL, Vol.15, No.3, pp.182-211, 1976.
- [1-6] T. Gilb and D. Graham, Software Inspection, Addison-Wesley, 1993
- [1-7] A. Porter, L. Votta, V. Basili, Comparing Detection Methods for Software Requirements Inspections: A Replicated Experiment, IEEE TSE, Vol.21, No. 6, pp.563-575, 1995.
- [1-8] A.Porter, L.Votta, What Makes Inspectiobs Work?, Software IEEE, Vol. 14, No. 6, pp.100-102, 1997.
- [1-9] R.Glass, Inspection - Some Surprising Findings, COMMUNICATION OF THE ACM, Vol.42, No. 4, pp.17-19, 1999.
- [1-10] P.Johnson, Reengineering Inspection, COMMUNICATION OF THE ACM, Vol.41, No. 2, pp.49-52, 1998.
- [1-11] T.Hall, D.Wilson, N.Baddoo, Towards Implementing Successful Software Inspection, Proceedings of the Software Methods and Tools(SMT), pp.127-136, Wollongong, NSW, 2000.
- [1-12] 情報処理推進機構（IPA）ソフトウェア高信頼化センター（SEC），定量的品質予測のススメ，pp.17-18，2008.
- [1-13] 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.377-384, 2009.
- [1-14] K.Petersen, K.Ronkko, C.Wohlin, The Impact of Time Controlled Reaing on Software Inspection Effectiveness and Efficiency :A Controlled Experiment, Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement, pp.139-148, New York, 2008.

- [1-15] 森下月菜, 青山幹雄, IPSJ SIG Technical Report Vol.2014-SE-183, No. 7, pp. 1-7, 2014.
- [1-16] D.Freedman, G.Weinberg, Handbook of Walkthrough, Inspections, and Technical Reviews 3<sup>rd</sup> edition, Little Brown, 1982.
- [1-17] 堀内純孝, 役に立つデザインレビュー, 日科技連, 1992.
- [1-18] 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, No.118, pp.25-31, 2004.
- [1-19] 込山俊博, 上流品質向上の関するソフトウェア評価技術の国際標準化同行, 情報処理 Vol. 50, No. 5, pp.391-399, 2009.
- [1-20] E.Weller, Lessons from Ththree Years of Inspection Data, IEEE SOFTWARE, Vol. 10, No. 5, SEPTEMBER, pp.38-45, 1993.
- [1-21] R.Grady, T.Slack, Key Lessopns In Achieving Widespread Inspection Use, IEEE SOFTWARE, Vol.11, No.5, JULY, pp.46-57, 1994.
- [1-22] N.Eickelmann, F.Ruffolo, J.Baik, A.Anant, An Empirical Study of Modifying the Fagan Inspection Process and the Resulting Main Effects and Interaction Effects Among Defects Found, Effort Required, Rate of Preparation and Inspection, Number of Team Members and Product 1<sup>st</sup> Pass Quality, Proceedings of the 27<sup>th</sup> Annual NASA goddard/IEEE software Engineering Workshop, 2003.
- [1-23] S.Biffli, W.Gutjahr, Influence of Team Size and Defect Detection Technique on Inspection Effectiveness, Proceedings of Seventh International Software Metrics Symposium, pp.63-75, 2001.
- [1-24] J.Miller, Z.Yin, Adding Diversity to Software Inspections, Proceedings of the Second IEEE international Conference on Cognitive Infomatics, pp. 81-86, 2003.
- [1-25] F.Shull, Inspecting the History of Inspectios:An Example of Evidence-Based Technology Diffusion, IEEE SOFAWARE, Vol.25, No.1, pp.88-90, 2008.
- [1-26] N.Hashemitaba, S.Ow, Generativw Inspection:An Intelligent Model to Detect and Remove Software Defects, IEEE SOFAWARE, pp.688-691, 2012.
- [1-27] 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.
- [1-28] A.Ferreria, R.Machado, L.Costa, J.Silva, R.Batista, M.Paulk, An Ap- porach to Improving Software Inspections Performance, IEEE International Conference on Maintenance, pp.1-8, Timisoara, Romania, 2010.
- [1-29] B.Robbins, J.Carver, Cognitive Factors in Perspective-Based Reading (PBR):A Protocol Analysis Study, Proceedings of the IEEE Third Interna-

- tional Symposium on Empirical Software Engineering and Measurement, pp.145-155, Lake Buena Vista, FL, 2009.
- [1-30] T.Berling, T.Theelin, A Case Study of Reading Techniques in a Software Company, Proceedings of the 2004 international Symposium on Empirical Software Engineering, pp.229-238, 2004.
- [1-31] E.Farchi, S.Ur, Selective Homeworkless Review, Proceedings of the 2008 International Conference on Software Testing, Verification, and Validation, pp.404-413, 2008.
- [1-32] L.Hatton, Testing the Value of Checklists in Code Inspections, IEEE SOFAWARE, Vol.25, No.4, pp.82-88, 2008.
- [1-33] P.Murphy, J.Miller, A Process for Asynchronous Software Inspection, Software Technology and Engineering Practice, 1997. Proceedings of 8th IEEE International Workshop on [incorporating Computer Aided Software Engineering], pp.96-104, London, UK, 1997.
- [1-34] M.Genuchten, W.Comelissen, C.Dijk, Supporting Inspections With an Electric Meeting System, Proceedings of the Hawaii International Conference on System Sceinces, pp.405-411, 1997.
- [1-35] 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.
- [1-36] 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.
- [1-37] 森崎修司, 間違いだらけの設計レビュー, 日経 BP, 2013.
- [1-38] 久野倫義, 丹羽友光, 前川隆昭, デザインレビューの効果的な実施方法, 第26回ソフトウェア品質シンポジウム, Tokyo, Japan, 2006.
- [2-1] T.Gilb, D.Graham, Software Inspection (邦訳)ソフトウェアインスペクション, 構造計画研究所, pp.83-95, 1999.
- [2-2] 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 No.4, pp.6-15, 2005.
- [2-3] 久野倫義, 丹羽友光, 前川隆昭, デザインレビューの効果的实施及び評価方法, 三電技法 Vol.83, No.5, p.18, 2009.
- [2-4] 情報処理推進機構 (IPA) 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, p.35, 2008.
- [2-5] 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.

- [2-6] 猪野仁, 効率的なレビュープロセスの設計方法について, 情報処理学会研究報告ソフトウェア工学研究会報告 96(112), pp.17-24, 1996.
- [2-7] 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.
- [2-8] 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.
- [2-9] P. McCarthy, A. Porter, H. Siy, L. Votta, An Experiment to Assess Cost-Benefits of Inspection Meetings and their Alternatives A Pilot Study, IEEE Proceedings of METRICS, pp.100-111, Berlin, Germany, 1996.
- [3-1] T. Gilb, D. Graham, Software Inspection (邦訳)ソフトウェアインスペクション, 構造計画研究所, pp.179-194, 1999.
- [3-2] A. Porter, P. Johnson, Assessing Software Review Meetings: Results of a Comparative Analysis of Two Experimental Studies, IEEE TSE, Vol.23, No.3, pp.129-146, 1997.
- [3-3] A. Porter, L. Votta, V. Basili, Comparing Detection Methods for Software Requirements Inspections: A Replicated Experiment, IEEE TSE, Vol.21, No.6, pp.563-575, 1995.
- [3-4] N. Kuno, T. Nakajima, M. Matsushita, K. Inoue, 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, SEC journal, No36, pp.16-23, 2014.
- [4-1] 情報処理推進機構 (IPA) 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, p.37, 2008.
- [4-2] 情報処理推進機構 (IPA) ソフトウェア・エンジニアリングセンター, 続定量的品質予測のススメ, pp.111-115, 2011.
- [4-3] L. Briand, K. Eman, B. Freimut, O. Laitenberger, Quantitative Evaluation of Capture-Recapture Models to Control Software Inspections, Proceedings of the 8th International Symposium on Software Reliability Engineering (ISSRE), pp.234-245, 1997.
- [4-4] S. Wiel, L. Votta, Assessing Software Design using Capture-Recapture Models, IEEE Transactions on Software Engineering, Vol.19, No.11, pp.1045-1054, 1993.
- [4-5] A. Goel, Software Reliability Models: Assumptions, Limitations, and Applicability, IEEE Transactions on Software Engineering, Vol.11, No.12,

pp. 1411-1423, 1985.

[4-6] C. Wohlin, P. Runeson, Defect Content Estimations from Review Data, Proceedings of the 20th International Conference on Software Engineering, pp. 400-409, 1998.

[4-7] S. Biffl, M. Halling, Investigating Reinspection Decision Accuracy Reading Product-Quality and Cost-Benefit Estimates, Proceedings of the 25th Annual International Computer Software and Applications Conference (COMPSAC'01), pp. 87-96, 2001.

[4-8] 山田茂, ソフトウェア信頼性モデル, 日科技連, 1994.

[5-1] 藤岡卓, 田村直樹, 中島毅, 久野倫義, 真野哲也, ソフトウェア技術者エントリー層に対する職能教育コースの設計と実装, 公益社団法人日本工学教育協会 工学教育第 62 巻第 1 号, pp. 46-51, 2014.

[5-2] A. Goswami, G. Walia, An Empirical Study of the Effect of Learning Style on the Faults found during the Software Requirements Inspection, Proceedings of The 24th IEEE International Symposium on Software Reliability Engineering (ISSRE), pp. 330-339, 2014.

[5-3] C. Seaman, V. Basili, Communication and Organization: An Empirical Study of Discussion in Inspection Meetings, IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, Vol. 24, No. 6, JULY, 1998.

[5-4] 井田達也, 欠陥連鎖チャートを用いたレビュー方法の提案-欠陥知識の有効利用によるレビュー, ソフトウェア品質シンポジウム, 2014.

[5-5] 篠崎悦郎, 重大欠陥検出に集中する問診に基づくレビュー法 (IBR 法) の提案-レビュー前に成果物作成状況の問診によるレビューポイントの導出, ソフトウェア品質シンポジウム, 2014.

[5-6] 北地敏隆, 効率的・効果的なレビュー実施のための新規役割「ハーベスタ」の提案-知見分析表を用いた欠陥傾向分析によりレビューの質を向上, ソフトウェア品質シンポジウム, 2014.





## 付録 1 ピアレビュー定義

本章では、開発現場でピアレビューを導入する際に開発者へ説明するためのピアレビュープロセス定義を示す。

### 付録 1.1 ピアレビューの種類とその目的

プロダクトの品質向上を目的に実施するレビューでは、関係者との合意形成、プロセスの実施有無のレビュー、QCDR の課題を抽出する目的でのレビューは対象とせず、成果物の品質向上を目的とする。

すなわちレビューとは、作成者、同僚、営業、客先等の利害関係者が、品質向上を目的に作業成果物の内容をチェックする活動である。以下にレビューの種類を示す。

表 21 レビューの種類

|     | 名称        | 概要  |
|-----|-----------|---|
| 公式  | チームレビュー   | 複数のメンバによる計画的、手順化されたレビュー。                                      |
|     | インスペクション  | 最も厳格で、体系的に実施されるレビュー。欠陥除去能力が高く、最も効果的なレビュー。成果物に潜む不具合を発見することに集中。 |
| 非公式 | アドホックレビュー | 必要に応じて、非計画的に行われるレビュー。記録されないことが多い。                             |
|     | パスアラウンド   | 作業成果物のコピーを複数人に配付し、複数コメントを獲得。                                  |
|     | ペアレビュー    | 作成者とレビューア 1 名で行われるレビュー。                                       |
|     | ウォークスルー   | 自らが対象成果物内容を説明することで、自らが欠陥に気づく。                                 |

上記の一般的なレビュー手法は、様々な文献等に詳細に記載されている。ただし、各アクティビティに関しては記載されているが、アクティビティ間の関係については記載されていない事が多い。以下の章で、レビュー手法におけるアクティビティ間の関連を示す。

## 付録 1.2 ピアレビューの主要アクティビティ

ピアレビューの主要アクティビティを以下に示す。

表 22 ピアレビューの主要アクティビティ

| 主要アクティビティ  | 説明  |
|------------|---|
| 計画         | 適切な頁数に分割，査読者役割設定，事前査読スピード，チェックリスト準備，ソース文書・規則の用意 |
| 作業成果物の品質確認 | ピアレビューリーダーによる成果物の確認。ピアレビューに値するかをチェックする。         |
| 作業成果物の事前配布 | 成果物を配布。必要に応じて説明会を開催                             |
| 作業成果物の事前査読 | 査読スピードを守り，チェックリスト，役割に応じて査読を行う。査読時間，指摘数を記録する。    |
| ロギングミーティング | 事前チェックされた内容を収集し，追加指摘を抽出する。                      |
| 編集とフォローアップ | ピアレビューリーダーが，指摘全てに対して，終結までフォローを行う。               |

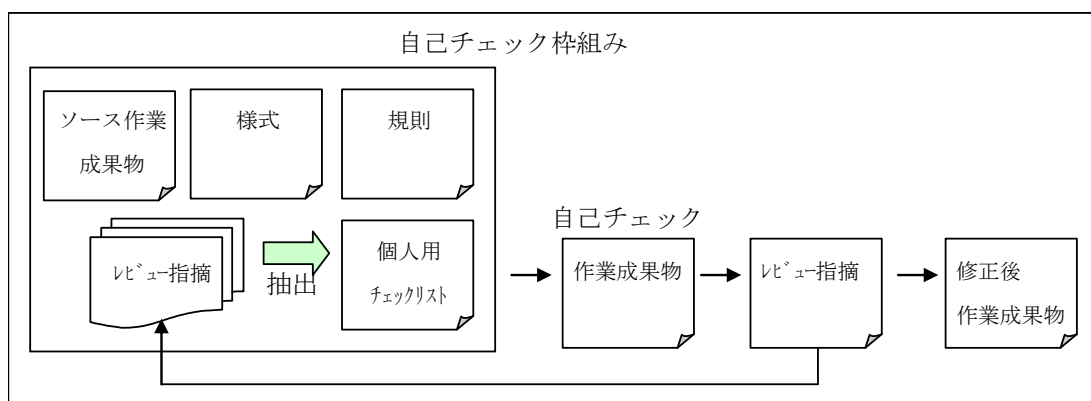
## 付録 1.3 ピアレビューに必要な作業成果物

ピアレビューを行う場合には，以下の作業成果物が必要である。

- ① 作業成果物（レビュー対象）を印刷
- ② ソース文書等（レビュー対象作業成果物を作成する為に使った，作業成果物）を印刷
- ③ 規則/様式（レビュー対象作業成果物を作成する為に使った規則や様式）
- ④ チェックリスト（自己チェックリスト，事前査読用チェックリスト）

## 付録 1.4 ピアレビュー枠組み

以下は，ピアレビューに関する一般的な枠組みである。チェックリスト作成においては，テスト不具合情報からの反映も重要であるが，まずは，レビュー記録からの作成を推奨する。



↓徹底的な自己レビュー完了後にチームでレビュー開始する。

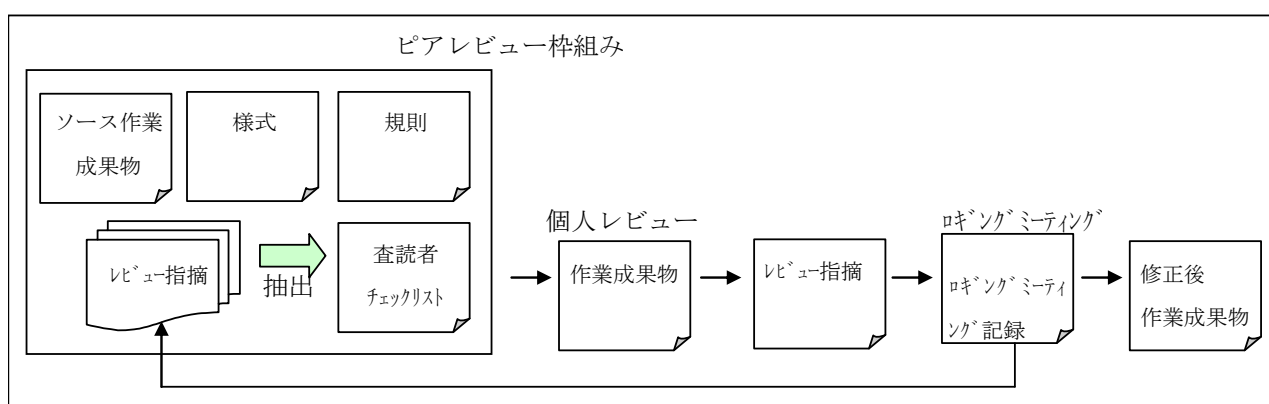


図 24 ピアレビューの枠組み

### 付録 1.5 自己チェックリスト作成方法例

自己チェックリストの作成方法を以下に示す。

- ① レビュー指摘を分析し、欠陥型をまず 10 個以下になるように設定する。
- ② 欠陥型に対応したパレート分析により、最も多い欠陥型を明確化する。
- ③ 最も多い欠陥型に対応した、チェック項目を検討する。欠陥内容に応じて、欠陥型 1 つに対して、複数のチェック項目を作成する。
- ④ 作成した自己チェックリストを使った結果をフィードバックし、チェックリストを更新し続ける。

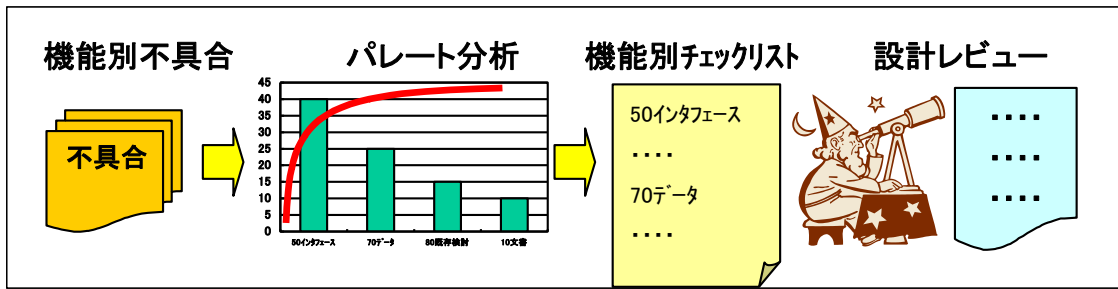


図 25 個人用チェックリスト作成フロー概要

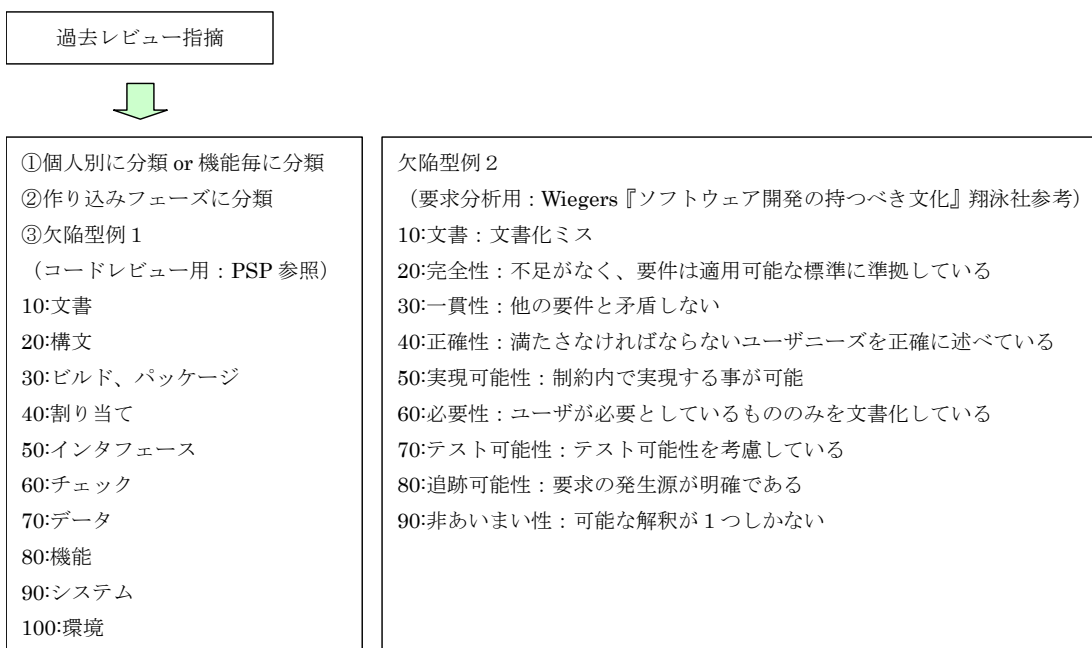


図 26 欠陥型例

### 付録 1.5.1 査読者用チェックリストの作成方法

自己チェックリストは、個人の作り込み易い欠陥を過去のレビュー指摘より抽出するが、査読者用のチェックリストは、シナリオ（Scenario-Based Reading）や観点（Perspective-Based Reading）を考慮して作成する。

(1) SBR：具体的なフローに基づいて成果物を読む事ができる。ただし、シナリオ作成の手間とシナリオの出来に依存する。

(2) PBR：利用者，設計者，テスト実施者などの立場でレビューする。焦点を絞って読む事ができる。

### 付録 1.5.2 SBR の作成観点

- ① ユースケースがあれば、ユースケースに基づいて、シナリオを作成する。
- ② ユースケースがなければ、関係者と役割や機能をマトリクスで表し、各マトリクスセルを網羅するシナリオを作成する。
- ③ シナリオの網羅性を検討する。

### 付録 1.5.3 PBR の作成観点

- ① 関係者を列挙し、各関係者毎に品質特性から必要な観点を取捨選択する。
- ② 取捨選択した観点に対して、具体的な観点を作成する。
- ③ 具体的な観点が、モレなく重複ないことを確認する。

### 付録 1.6 チェックリストの一般的な使用方法

チェックリストは、各観点やシナリオ 1 項目毎に、全てのドキュメントやソースコードをチェックする。全てのチェックが完了した後に、チェック完了とする。

### 付録 1.7 ロギングミーティングの運営方法

ロギングミーティングは、レビューやインスペクションで複数の視点で成果物をチェックするし、また自己チェックや個人チェックでは見落とされた欠陥を抽出できる場である。ロギングミーティングを効果的効率的に運用することで、効果的なレビューとする事が出来る反面、モデレータの力量で、非効率なミーティングになる。本章は、ロギングミーティングの効果的な運営方法について以下に示す。

- 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召集する。
- ④ 事前査読時間と指摘件数を全員に報告させ、事前査読が不十分なら、延期も検討する。
- ⑤ 会議内容を測定し、発言内容毎／出席者毎の時間を測定する。
- ⑥ 会議内容を記録。（指摘内容ではなく、各発言で気になる点※を記録する：改善推進者が実施）
- ⑦ 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。
- ⑧ 全員が発言するように促す。
- ⑨ モデレータは、参加者の発言内容、動作も観察し、納得していないようなら

発言を促す。

- ⑩ 目的が達成できたかを，終了時に確認する。
- ⑪ 指摘をミーティングの最後に再確認し，誤り，抜けやモレがないことを確認する。

#### 付録 1.8 留意すべき発言とその解釈例

- ① 指摘に対しては，その場で回答をしようとして，ディスカッションになる。効率が悪い。
- ② 事前レビューを実施し，資料として纏めているので，資料自体の読上げはなく，指摘の読上げ及び指摘理解を行っていた。
- ③ 議論が多く，特に仕様書に記載されていない事項に関してのものが多い。「以前に説明したように x x」という話から，今までも口頭での説明及び記録されない前提が多いと判断する。今後も，同様な議論が発生する可能性が高い。
- ④ 議論は，仕様書に記述されるか？（この場にはいない人には伝わらない。）
- ⑤ 議論したが，相手は今ひとつ納得してないものがある。今後の手戻りの火種となる懸念あり。
- ⑥ 製品そのものの理解するための打合せというように感じた。議論した内容を仕様書に記述しないと，試験時に，多くの不具合が発生する可能性が高い。
- ⑦ 関係者間で用語の理解が異なり，大きなすれ違いが起こる可能性が高い。
- ⑧ 仕様書自体の記述内容が，不明瞭な部分が多いように感じた。
- ⑨ 前提条件が記載されていないため，実装後に手戻りが発生する可能性が高い。
- ⑩ GUI は実機で確認したい。GUI 部のみ早急にデモする必要あるのではないか。
- ⑪ 対象仕様書に記述されるべき事と，かかれなことが不明。どこまで見れば良いか疑問。S/W 設計のレベルと製品レベルの記述が混在してるようである。
- ⑫ 事前レビューは 2 人日ということだが，1/3 は誤記/先祖がえりということであり，深い指摘ができたか不明。

#### 付録 1.9 レビューの評価

レビューを評価する指標は，①欠陥除去率，②欠陥密度，③欠陥検出率，④レビュー率，⑤ゾーン分析，⑥レビュー網羅率などがある。

- ① 欠陥除去率：レビューで検出できる欠陥の割合。テスト完了しないと評価できない
- ② 欠陥密度：単位規模当たりの欠陥数

- ③ 欠陥検出率：レビュー1時間当たりの検出欠陥数
- ④ レビュー率：レビュー1時間当たりレビューする成果物規模
- ⑤ ゾーン分析：X 軸レビュー工数，Y 軸レビュー検出率でプロットし，プロセスを評価
- ⑥ ピアレビュー網羅率：成果物全体にわたってピアレビューを実施できたかを判定する。

ただし，レビュー方法に応じて，何を測定するかを明確に決めておく必要がある。自己チェック，個人レビュー，ロギングミーティング毎に測定するのか，ロギングミーティングのみの情報を測定するのか等。