JAIST Repository
https://dspace.jaist.ac.jp/

Title	要件定義書の閲読時の視線情報に基づく閲読能力の特徴付け
Author(s)	斉藤, 功樹
Citation	
Issue Date	2021-06
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/17491
Rights	
Description	Supervisor:日高昇平, 先端科学技術研究科, 博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

博士論文

要件定義書の閲読時の視線情報に基づく閲読能力の特徴付け

斉藤　功樹

主指導教員　日髙　昇平

北陸先端科学技術大学院大学

先端科学技術研究科［知識科学］

令和３年６月

Abstract

A software development review is a method to improve the quality of deliverables, and it contributes to quality improvement. In software development, the waterfall development model is often used, and the process takes place from the upstream to downstream. Therefore, the quality of the requirement definition document (RDD), which is a deliverable of the upstream process, is very important. Although many review methods have been developed to improve the quality of deliverables, there is no single best method for all situations, and individual differences are more important than differences in review methods. Furthermore, existing indices, such as defect detection rate and review efficiency, cannot sufficiently evaluate the review quality. In addition, the process of the review is a black box, and it is difficult to determine how the review was conducted.

This study was performed to develop tasks to evaluate the review performance of the RDD in the upstream process of software development and characterise factors affecting review performance by using gaze patterns. The review process consists of the following three parts: reading, understanding the structure, and detecting and fixing defects. Understanding the structure, detecting and fixing defects are specific processes in the review, the gaze patterns that affect these two processes were characterised.

In Chapter 2, the sensitivity of defect detection was defined as an index of review quality to replace the defect detection rate, and the relationship between sensitivity and gaze patterns was discussed. In the experiment, RDDs used in actual software developments introduced defects absent in the original RDDs, and the sensitivity to detect those defects was calculated using signal detection theory. As a result, there was a significant correlation between the sensitivity to detect defects and the blink rate, indicating that blink was a major feature of characterising a review performance. Related to the review process, it was suggested that reviewers with a low defect-detection sensitivity may remain in the reading process.

In Chapter 3, the software was symbolised by geometric patterns, and an experiment was conducted to create a pair of client requirements and RDD. In the review, the reviewer matches the deliverables from the previous and current processes. However, client requirements are not explicitly defined in the RDD review. Therefore, the client requirements were described as geometric patterns by symbolising the software with geometric patterns. By using these pairs of client requirements and RDD, it is possible to evaluate the review performance to determine whether a RDD is necessary and sufficient to meet client requirements. In Chapter 4, the relationship between review performance and gaze patterns was analysed. As a result, the reviewer with a high correct rate tended to have mydriasis in their pupils, it was likely that they had greater cognitive effort and better concentration on the task. Furthermore, the fixation tended to be focussed on a certain part of the sentences when the review was

successful. This result indicated that the reviewer with good review performance may strategically allocate cognitive resources according to the amount of information in each sentence.

In summary, it was suggested that the blink rate could be used to determine whether the reviewer remained in the reading process, and the distribution of pupil diameter and the fixation for each sentence could be used to evaluate review performance in processes specific to the review (understanding the structure, detecting and fixing defects). Therefore, gaze patterns are appropriate for characterising review performance, and review quality can be predicted in real time using these findings.

Keywords
gaze, software review, requirement definition document, review performance, machine learning

目次

第 1 章　序論 10
1.1　背景と課題 10
1.1.1　ソフトウェア開発の品質 10
1.1.2　閲読手法と個人差 12
1.1.3　閲読品質と評価指標 13
1.2　目的とアプローチ 13
1.2.1　閲読とは 14
1.2.2　閲読能力と経験の関係 15
1.2.3　閲読能力と視線の関係 16
1.2.4　認知と視線の関係 16
1.2.4.1　Fixation と認知の関係 17
1.2.4.2　Saccade と認知の関係 18
1.2.4.3　瞬目と認知の関係 19
1.2.4.4　瞳孔と認知の関係 22
1.3　本論文の仮説 23
1.4　本論文の構成 24
第 2 章　実験 1：欠陥の検出感度と視線の関係 26
2.1　実験の目的 26
2.2　実験の概要 27
2.3　実験手続き 28
2.3.1　被験者 28
2.3.2　閲読対象文書 28
2.3.3　実験手続きと装置 30
2.4　結果 31
2.4.1　閲読品質 32
2.5　分析 33
2.5.1　分析 1：経験と閲読品質の関係 34
2.5.2　分析 2：瞬目率と閲読品質の関係 36
2.5.3　分析 3：閲読品質のモデルベース分析 40
2.5.3.1　特徴量の抽出 40
2.5.3.2　閲読品質予測モデル 41
2.5.3.3　d-prime 正負分類モデル 43
2.5.4　分析 4: 意図しない欠陥を考慮した閲読品質と瞬目率の関係 44
2.5.4.1　意図しない欠陥の分類 44
2.5.4.2　経験と意図しない欠陥を考慮した閲読品質の関係 45
2.5.4.3　瞬目率と意図しない欠陥を考慮した閲読品質の関係 46
2.6　考察 48

2.6.1 瞬目率と閲読品質 48
2.6.2 本実験の限界 49
第 3 章 実験 2：要件の必要十分性判別能力課題作成 51
3.1 実験の目的 51
3.2 実験手続き 56
3.2.1 被験者 56
3.2.2 手続き 56
3.2.3 課題難易度の調整 58
3.2.3.1 正例の作成規定 58
3.2.3.2 顧客要求文の言語的記述規定 60
3.3 結果 62
3.3.1 Client グループ 62
3.3.2 Developer グループ 64
3.4 考察 66
第 4 章 実験 3：要件の必要十分性判別能力と視線の関係 68
4.1 実験の目的 68
4.2 実験手続き 69
4.2.1 被験者 69
4.2.2 課題 69
4.2.3 手続き 70
4.3 結果と分析 73
4.3.1 分析 1：正答率の分析 73
4.3.1.1 経験と正答率の関係 73
4.3.1.2 タイプごとの正答率と回答時間 74
4.3.1.3 タイプ N と NS の正答率の関係 76
4.3.2 分析 2：正答率のモデルベース分析 77
4.3.2.1 特徴量の抽出 77
4.3.2.2 正答率予測モデル 79
4.3.2.3 瞳孔径と平均正答率の関係 79
4.3.3 分析 3：正答／誤答の分析 85
4.3.3.1 正答／誤答と回答時間 85
4.3.3.2 正答／誤答と fixation 86
4.3.3.3 回答時間の予測モデル構築 89
4.3.4 分析 4：誤認識を考慮した正答／誤答の分析 92
4.3.4.1 正答／誤答と fixation 92
4.3.4.2 回答時間の予測モデル構築 95
4.3.5 分析 5：正答／誤答のモデルベース分析 98
4.3.5.1 誤答データのオーバーサンプリング 98
4.3.5.2 特徴量の重要度 99
4.3.5.3 正答／誤答分類モデル 99

4.3.6　分析 6：情報量と視線の関係 99
4.3.6.1　情報量の定義と分析手法 100
4.3.6.2　特徴量と顧客要求文との相関関係 100
4.3.6.3　文集合のもつ情報に関連する特徴量と fixation の関係 103
4.4　考察 105
4.4.1　本課題の有用性 105
4.4.2　閲読品質が高い人の特徴 105
4.4.3　実際の閲読との比較 107
4.4.4　本実験の限界 108
第 5 章　総合議論 109
5.1　閲読能力と視線情報の関係 109
5.2　本研究の貢献 111
5.3　今後の課題 113
5.4　おわりに 114
参考文献 116
付録 124
付録 A 実験 1 のデータ 124
付録 A.1 被験者ごとの d-prime と属性情報 124
付録 A.2 47 個の特徴量 125
付録 A.3 被験者のノイズに対する指摘欠陥分類と件数 126
付録 B 実験 3 のデータ 127
付録 B.1 回答時間の予測モデル構築における 24 個の特徴量の重要度 127
付録 B.2 誤認識を考慮した回答時間の予測モデル構築における 24 個の特徴量の重要度 128
付録 B.3 正答／誤答分類モデル構築における 24 個の特徴量の重要度 129
研究業績 130
謝辞 131

図 1.1　ウォーターフォールモデル.....11
図 1.2　著者が文章を読んでいる際の fixation と saccade の様子をアイトラッカの分析ソフトウェアにて出力した結果（文章の出典：平成 28 年度 秋期 IT ステラジスト試験区分 午後 I 問 1）..... 17
図 1.3　会話・休憩・読書における年齢ごとの瞬目率(Bentivoglio et al., 1997)..... 22
図 1.4　閲読における三つの認知プロセス 24
図 2.1　被験者の要件定義書の閲読経験分布..... 28
図 2.2　実験時の画面サンプル（社外秘情報が含まれるため本文を黒塗りとする） 30
図 2.3　(左)実験の様子,(右)アイトラッカ（モニタ下部にアイトラッカを設置） 31
図 2.4　頁ごとの有効な fixation の割合のヒストグラム..... 32
図 2.5　要件定義書の閲読経験数と文書閲読の頻度の散布図..... 35
図 2.6　要件定義書の閲読経験数と d-prime の散布図 35
図 2.7　文書の閲読頻度と d-prime の散布図 36
図 2.8　d-prime と瞬目率の散布図（青線は瞬目率の平均値を示す） 38
図 2.9　d-prime とノイズへの反応数の散布図..... 39
図 2.10　F 値と瞬目率の散布図（青線は瞬目率の平均値を示す） 39
図 2.11　d-prime の実測値と予測値（SVR の結果）の散布図 43
図 2.12　要件定義書の閲読経験数と意図しない欠陥を考慮した d-prime の散布図..... 46
図 2.13　意図しない欠陥を考慮した d-prime と瞬目率の散布図（青線は瞬目率の平均値を示す） 47
図 2.14　意図しない欠陥を考慮した F 値と瞬目率の散布図（青線は瞬目率の平均値を示す） 48
図 3.1　ソフトウェア開発での閲読と CD ゲームの関係 53
図 3.2　閲読者によるソフトウェアの内包的定義と外延的定義の突合の例..... 54
図 3.3　集合 R（要件定義書）と集合 P（ソフトウェアプロダクト）の間の 4 種類の関係 54
図 3.4　client-developer ゲームにおける全体概要図..... 56
図 3.5　（a，左）正例，（a，右）負例，（b）顧客要求文..... 58
図 3.6　20 種類の単位図形..... 59
図 3.7　同値な図形の例（青枠内の図形が同値グループ） 60
図 3.8　各被験者のラウンドごとの顧客要求文の数推移（赤線は実験者が事前に設定した必要な顧客要求の文数） 64
図 3.9　client グループ 2 における顧客要求文のラウンドによる変化例..... 64
図 4.1　実験時の画面イメージと顧客要求，正例，負例それぞれの拡大図..... 72
図 4.2　問題ごとの有効な fixation の割合のヒストグラム 73
図 4.3　要件定義書の閲読経験数と正答率の散布図 74
図 4.4　タイプごとの回答時間の箱ひげ図（$*: \boldsymbol{p < .05}, ** : \boldsymbol{p < .01}, *** : \boldsymbol{p < .005}$） 76
図 4.5　タイプ N とタイプ NS の正答率の散布図..... 77
図 4.6　瞳孔径（z スコア）の最小値と正答率の散布図..... 82
図 4.7　瞳孔径（z スコア）の歪度と正答率の散布図 82

図 4.8　瞳孔径（z スコア）の尖度と正答率の散布図 83
図 4.9　瞳孔径の平均値の増減率と正答率の散布図 83
図 4.10　瞳孔径の標準偏差の増減率と正答率の散布図 84
図 4.11　KL 情報量と正答率の散布図 84
図 4.12　安静時と課題実施時の瞳孔径のヒストグラム（上段：正答率 0.847 の被験者，下段：正答率 0.538 の被験者） 85
図 4.13　タイプごとの正答／誤答の回答時間の箱ひげ図（$*: p < .05, ** : p < .01, *** : p < .005$） 86
図 4.14　タイプ N における顧客要求の文番号ごとの fixation 数（エラーバーは標準誤差を示す，$*: p < .05, ** : p < .01$） 87
図 4.15　タイプ N における顧客要求の文番号ごとの単位時間あたりの fixation 数（エラーバーは標準誤差を示す，$*: p < .05, ** : p < .01$） 87
図 4.16　タイプ NS における顧客要求の文番号ごとの fixation 数（エラーバーは標準誤差を示す，タイプ NS の問題では文番号 11 が最大値である） 88
図 4.17　タイプ NS における顧客要求の文番号ごとの単位時間当たりの fixation 数（エラーバーは標準誤差を示す，タイプ NS の問題では文番号 11 が最大値である） ... 88
図 4.18　6 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図 91
図 4.19　7 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図 91
図 4.20　重回帰分析による予測回答時間と回答時間の散布図（6 番目と 7 番目の顧客要求文における単位時間当たりの fixation 数を説明変数として重回帰モデルを構築） 92
図 4.21　タイプ N における顧客要求の文番号ごとの fixation 数（エラーバーは標準誤差を示す，$*: p < .05$） 93
図 4.22　タイプ N における顧客要求の文番号ごとの単位時間あたりの fixation 数（エラーバーは標準誤差を示す，$*: p < .05, ** : p < .01$） 94
図 4.23　タイプ NS における顧客要求の文番号ごとの fixation 数（エラーバーは標準誤差を示す） 94
図 4.24　タイプ NS における顧客要求の文番号ごとの単位時間あたりの fixation 数.... 95
図 4.25　7 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図 97
図 4.26　3 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図 97
図 4.27　9 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図 98
図 4.28　顧客要求の文番号ごとの要素数の箱ひげ図 102
図 4.29　顧客要求の文番号ごとの減少数の箱ひげ図 102
図 4.30　顧客要求の文番号ごとの減少率の箱ひげ図 103
図 4.31　減少数ごとの単位時間当たりの fixation 数のヒストグラム 104

図 5.1　閲読のプロセスと閲読能力における視線の特徴 111

表 1.1　瞬目に影響する要因(田多他, 1991)..... 21
表 2.1　実験で使用した閲読対象文書の詳細情報..... 29
表 2.2　信号検出理論における反応の分類 33
表 2.3　d-prime と瞬目率の相関係数..... 38
表 2.4　d-prime とノイズへの反応数の相関係数 39
表 2.5　RF により抽出された特徴量と重要度 41
表 2.6　アルゴリズムごとのライブラリとパラメータ..... 42
表 2.7　アルゴリズムごとの閲読品質予測モデルの性能 42
表 2.8　d-prime の実測値と予測値の相関係数..... 42
表 2.9　混合行列 44
表 2.10　d-prime の正負分類モデルの accuracy..... 44
表 2.11　ノイズに含まれる欠陥分類と件数（抜け漏れ：定義すべき要件の抜け，曖昧：二つ以上の解釈が可能，誤ったセクション：記載すべきセクションの誤り，標準化：用語が未定義または用語が未統一，校正：文章表現の修正や誤字脱字，確認／提案：要件に対する確認または要件の修正提案，冗長：複数の箇所にて同じ要件の繰返し）..... 45
表 2.12　意図しない欠陥を考慮した d-prime と瞬目率の相関係数 47
表 3.1　作成可能な図形の組み合わせ数..... 59
表 3.2　A は B である構文において作文可能な組み合わせ数 62
表 3.3　集合 R と P の関係の推移（タイプ N: 集合 R は集合 P の必要条件（necessary）であるが十分条件（sufficient）でない，タイプ S: 集合 R は集合 P の十分条件であるが必要条件でない，タイプ U: 集合 R は集合 P の必要条件でも十分条件でもない，タイプ NS: 集合 R は集合 P の必要十分条件である）..... 63
表 3.4　Developer グループ作図事例の正例類似度 66
表 3.5　Developer グループの作図事例の正例カバー率 66
表 4.1　13 問の課題（太字の顧客要求を用いた）..... 70
表 4.2　顧客要求のタイプごとの正答率（タイプ N: 集合 R は集合 P の必要条件（necessary）であるが十分条件（sufficient）でない，タイプ S: 集合 R は集合 P の十分条件であるが必要条件でない，タイプ U: 集合 R は集合 P の必要条件でも十分条件でもない，タイプ NS: 集合 R は集合 P の必要十分条件である）..... 75
表 4.3　RF により抽出された特徴量と重要度 78
表 4.4　アルゴリズムごとの正答率予測モデルの誤差（太字が最も誤差が小さい値を示す）..... 79
表 4.5　RF による重要度上位 3 位の特徴量..... 90
表 4.6　RF による重要度上位 5 位の特徴量..... 96
表 4.7　RF による重要度上位 3 位の特徴量..... 99
表 4.8　顧客要求の文番号ごとの三つの情報量の平均／標準偏差の推移 101
表 4.9　問題単位での顧客要求の文番号ごとの減少数の推移..... 104
表 4.10　分析 1～6 にて得られた結果の整理 107

第1章　序論

ソフトウェアはいまや生活の基盤となり，日常に欠かせないものとなっている．さらにソフトウェアが高度化しており，ひとたび障害が発生するとその影響範囲は多大である．2020 年 10 月 1 日に東証で発生した障害は，要件とソフトウェアの仕様が異なっていたことにより発生し，終日全銘柄の売買取引が停止した．この障害により約 3 兆円の売買機会が損失したと想定されており，及ぼす影響は非常に大きい．そのようにソフトウェアがインフラ化した現在では，その品質は重要であり，短期間で高品質の製品が求められている．ソフトウェア開発の品質を担保する一般的な方法はレビュー（以降，閲読） であり，含まれる欠陥を検出することで品質を高める．しかし，十分な閲読がなされているのにもかかわらず，品質の低いソフトウェア開発につながることが多い．その原因の一つとして，閲読の目的において欠陥検出に主眼が置かれ，顧客が求めるソフトウェアを表現する要件定義書や設計書になっているかどうかを評価できていないことが挙げられる．そのため，閲読がどのようなプロセスでなされているのかを議論し，品質の良い閲読がなされているかを判別する指標を開発することが求められる．さらに品質の良い閲読者がもつ特徴を明らかにすることも重要である．

本章では，ソフトウェア開発の最上流工程の要件定義書を対象とし，閲読とは何かを議論し，それを基に閲読能力を定義する．その後，先行研究を基に閲読能力に影響を及ぼす要因を議論し，閲読能力を特徴づける要因の候補を明らかにする．

1.1　背景と課題

ソフトウェア開発において品質は重要であり，閲読によってその品質を高めていく．しかし，閲読を行ったとしても必ずしも品質の良いソフトウェアが開発できるとは限らない．その原因として，閲読品質が正確に測れていない可能性が考えられる．本節では，ソフトウェア開発の閲読における課題について議論する．

1.1.1　ソフトウェア開発の品質

ソフトウェア開発において，ウォーターフォールモデル(Royce, 1970)が最も広く用いられている開発手法であり(独立行政法人情報処理推進機構（IPA）社会基盤センター, 2018)，その名の通り上から下へと一方向で開発が進んでいく．ウォーターフォールモデルでは，上流工程である要件定義や設計から始まり，下流工程である開発, テスト及び導入へと一方向で進むモデルである（図 1.1）．それぞれの工程で成果物が作成され，それが次工程への入力となり，開発が進んでいく．各工程での詳細な計画に基づき，後工程が進んでいくため，安定性や確実性が高い開発が実現できる(Barry Boehm, 2004)．そのため，ソフトウェアの要件が比較的流動的でない大規模な開発に向いている．近年は，アジャ

イルモデル(Dybå & Dingsøyr, 2008)も用いられるようになっているものの，日本企業においてはその導入はほとんど進んでおらず，97%以上のプロジェクトでウォーターフォールモデルが採用されている(独立行政法人情報処理推進機構（IPA）社会基盤センター, 2018)．アジャイルモデルでは，開発対象の機能を細かく分割し，図 1.1 で示したフローを一つの反復として，短期間で機能の開発を目指す開発手法である.短期間で動くソフトウェアを開発していくため，小規模な開発に適しており，社会インフラを担うような大規模な開発には不向きである(片岡他, 2017)．

ウォーターフォールモデルでは，各工程での成果物の品質が不十分であった場合,後工程からの手戻りが発生し,開発コストの増加や納期遅れが発生する．成果物に含まれる欠陥を後工程で検出できなかった場合は，導入後にバグが発生し，ソフトウェアの品質低下を招く．近年のソフトウェア開発では，避けられる手戻りに 40–50%の労力を費やしていると報告されている(B Boehm & Basili, 2001)．そのため，各工程での成果物の品質を十分に担保し，手戻りにかかるコストを減らすことが必要である．避けられる手戻りの主要な原因と一つとして，十分な時間を費やせずに定義された要件定義書であると報告されており(B Boehm & Basili, 2001)，品質の悪い要件定義書は後工程すべてに悪影響を与える．欠陥が検出されることなく，ソフトウェアが顧客に導入された場合，その後多くの障害が発生し，その対応のメンテナンスのコストが増大し，さらに顧客からの信頼を失うリスクもある．

したがって，短い期間で高品質のソフトウェアを開発するためには，開発の前段階にあたる要件定義工程で，要件定義書の品質を担保することが重要である．品質が十分でない要件定義書や設計書によって後工程で修正が発生した場合，上流工程で欠陥を修正した場合と比べて 100 倍のコストが発生する(B Boehm & Basili, 2001)．

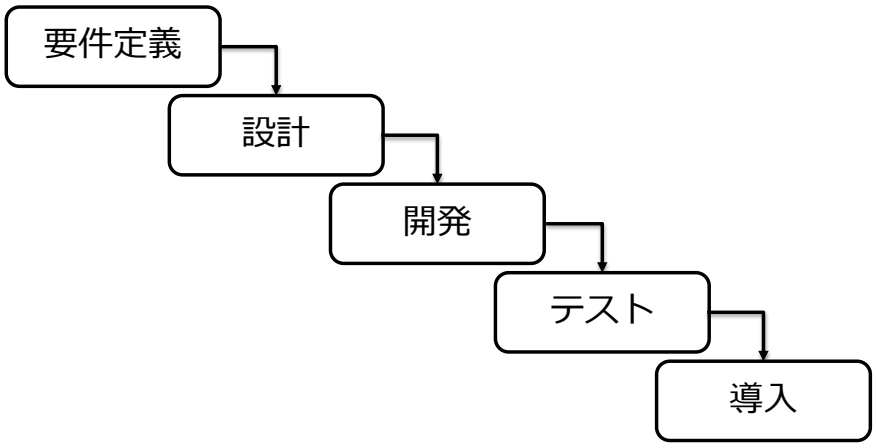


図 1.1　ウォーターフォールモデル

1.1.2 閲読手法と個人差

一般的に上流工程の成果物の品質を担保する方法は閲読であり，その目的は欠陥を検出することである(Ciolkowski et al., 2003)．閲読を行うことで，欠陥の 31-93%（中央値 60%）の欠陥を検出できる(B Boehm & Basili, 2001)．上流工程での欠陥は，omission（抜け漏れ）と commission（誤り）の二つに大別される(Porter & Votta, 1998)．抜け漏れは要求定義において必要な情報が抜けていることである．誤りは要求定義において誤った情報が含まれていることであり，曖昧な情報，矛盾した情報，誤ったまたは余分な機能，および誤ったセクションの四つのタイプに分けられる．

様々な閲読手法が開発され，手法ごとの違いや有効性について比較研究が数多くなされている(Bernardez et al., 2004; Cantone et al., 2003; Thelin et al., 2003, 2004; 松川, 2002; 野中誠, 2004)． Ad-hoc reading（Ad-Hoc）と check-list based reading（CBR）が広く用いられている手法であり，次いで scenario based reading（SBR）や perspective based reading（PBR）である．Ad-Hoc では手順や方法などの制約がなく閲読者が自由に閲読を行うため，最も閲読者の能力に依存する．CBR ではチェックリストを基に閲読を行う手法であり，Ad-Hoc と同程度の有効性であることが報告されている(上野他, 2005)．SBR はシナリオを基に閲読を，PBR は顧客やプログラマなどの異なる観点ごとに閲読を行う手法であり，CBR と比較して有効性が高いという報告がなされている．しかし，適用状況などによって逆の結果になることが示唆されている(Halling et al., 2001)ことから，必ずしもすべての状況で SBR や PBR の有効性が高いとはいえない．

様々な閲読手法が開発されているものの，欠陥を検出する最も優れた手法は確立されておらず，さらに手法による影響よりも個人差の方が大きいと報告されている(Uwano et al., 2006)．Uwano et al. (2007)は，Thelin et al. (2003)の結果に対して，SBR の一種である usage based reading（UBR）は CBR の 1.25-1.5 倍の欠陥検出率を示すものの，同じ手法における個人の効率の違いのほうが大きいと報告している．つまり，要件定義書の品質に与える影響は閲読手法よりも，閲読者の能力に依存すると考えられる．さらに，同一人物であっても，時間的な制約や集中度合いなどにより，閲読品質が異なるため，それらを考慮することも重要である．特にソフトウェア開発の現場においては，開発期間や工数が有限で，十分な時間を閲読に費やすことができるとは限らない．良い閲読者であっても，環境や状況によっては基本的な欠陥を検出できないことも多い．実際の開発現場においてはそれが顕著であり，単純な数値の誤りや誤字脱字が見落とされることもしばしば発生する．

多くの研究で，個人差がソフトウェア開発の品質に与える影響は大きいと報告されているものの，個人差の要因について調査した研究は少ない(Uwano, 2011)．

1.1.3 閲読品質と評価指標

閲読品質は，主に欠陥検出率や閲読効率で評価される(Thelin et al., 2003; Uwano, 2011)．先に述べた閲読手法の比較においても，その二つの指標を用いて評価されることが多い．欠陥検出率は，総欠陥数に対して閲読によって検出できた欠陥数であり，閲読手法の比較研究ではあらかじめ欠陥の総数がわかっているため，正確な欠陥検出率が算出できる．欠陥検出率で評価する理由は，閲読対象に含まれる欠陥数が異なるため，それぞれの閲読を一律に評価できるようにするためである．実際のソフトウェア開発においては，欠陥の総数は開発が完了するまではわからないため，欠陥密度が代わりに用いられることが多い．欠陥密度は，検出した欠陥数を開発規模で割ったものであり，欠陥の総数の代わりに開発規模を用いる．ほかの開発における欠陥密度と比較して，経験的にその品質を評価している．閲読効率は，単位時間当たりに検出した欠陥数である．評価指標に閲読効率が用いられている理由は，実際のソフトウェア開発では開発工数が有限であるため，効率的に欠陥を検出できることも重要だからである．

しかし，欠陥検出率や閲読効率の指標だけでは閲読品質を正確に判断できない．欠陥検出率は，閲読対象によって欠陥の総数が異なるため，それらの閲読品質を一律評価できるように導入された指標である.欠陥の総数が異なるため，欠陥検出率は閲読対象の品質に依存する．例えば，欠陥検出率が低い場合に，閲読能力の低さ／閲読対象文書の品質の高さのどちらに起因するものであるのか考慮する必要がある．合わせて，どのような閲読が行われたのかはブラックボックスとなっており，欠陥が検出されなかった個所についても十分に閲読が行われたかどうかは欠陥検出率からはわからない．そのため，閲読手法や既存の指標によらずに，個人の閲読能力を定量的に評価することが重要である．閲読効率は，閲読時間に依存するため，短い時間で閲読された場合にはその品質が高いのかは判別が難しい．

そのため，閲読品質を正確に評価するための新たな指標が求められる．さらに，閲読品質に影響を及ぼす個人の閲読能力を定量化することが重要である．能力の高／低に影響を及ぼす個人要因を明らかにすることで高品質のソフトウェア開発に繋がる．

1.2 目的とアプローチ

本研究の目的は，一つは閲読能力を測る指標を開発することであり，もう一つはその能力の高い／低い閲読者の特徴を視線情報によって解明することである．1.1 で述べたように，閲読品質は閲読実施率や欠陥検出率によって測られてきたものの，指標としては不十分であり，結果として品質の低いソフトウェア開発につながってしまうことがあった．そこで，本論文では閲読を成果物同士の突合であると考え，閲読品質を測る新たな指標を提案する．

閲読能力に影響を及ぼす個人要因として視線情報に着目し，どのような視線情報が閲読能力に影響を及ぼすのかを明らかにする．閲読時の視線情報を用いて，閲読能力を測る指標との関連を分析することにより，閲読能力の高低に影響を及ぼす視線の特徴を特定する．閲読能力の高い人の視線の特徴を分析することで，閲読者の教育にも有益である．本研究で得られた知見を活用することで，閲読者の能力向上を図れると考える．合わせて，閲読実施時の視線を計測することで，閲読品質をリアルタイムで予測できるようになる．さらに，閲読能力に影響を及ぼす視線の特徴は，人がどのように文章を読み理解しているのかの手かがりになると考える．

1.2.1 閲読とは

閲読の目的は欠陥を見つけることであることはすでに述べたが，閲読自体はどのような認知プロセスで行われるのだろうか．閲読品質を評価する指標を導入するためには，閲読自体の認知プロセスを明らかにする必要がある．

Uwano et al. (2007)は”In the software review, a reviewer reads the document, understands the structure and/or functions of the system, then detects and fixes defects if any.”と閲読を三つのプロセスに分けており，閲読では(1)読みに加えて,(2)構造の理解と(3)欠陥の検出／修正の認知プロセスがある．読みにおいては，要件定義書や設計書においては通常の文章を読む際の認知プロセスと同様であると考えられる．そのため，構造を理解し，欠陥を検出／修正することが閲読特有の認知プロセスであり，この二つのプロセスを明らかにする必要がある．

ウォーターフォールモデルにおいては,各工程にて前工程での成果物を基に,現工程の成果物の閲読が行われる．Thelin et al. (2003)は，設計書閲読における手法ごとの評価にて，実験参加者に要件定義書と設計書の両方を提示し，閲読を行って手法ごとの評価を行っている．設計工程以降では，前工程の成果物が存在するため，それらを基に，抜け漏れや誤りなどの欠陥が存在しないかを閲読できる．

したがって，閲読では，前工程の成果物と現工程の成果物の突合を行っている．例えば，プログラムの閲読の場合，設計書に記述された機能を読んで理解し，それが必要十分にプログラムとして実装されているかを確認している．Uwano (2011)は，設計書の閲読時には前工程の成果物である要件定義書に，プログラムの閲読時には前工程の成果物である要件定義書と設計書に集中するほうが，品質が良いと報告している．プログラムの閲読の場合は，プログラムそのものが実行可能かという観点でも閲読できるものの，それだけでは品質の良い閲読とはいえない．設計書においても，設計書内での曖昧さや矛盾などの欠陥は検出できるもの,同様に品質が低くなる.つまり,前工程の成果物を基に,現工程の成果物の閲読をすることが品質の良い閲読につながる．したがって，

閲読における構造の理解と欠陥の検出／修正は，二つの成果物の突合を行うプロセスである．

合わせて，要件定義工程においては，突合対象の前工程の成果物は存在しないため，設計工程以降の閲読と比較して異なる認知プロセスを示すと考えられる．設計工程以降は，前工程の成果物を基にして，現工程の成果物を作成しており，閲読における突合対象は，作成にあたり基にする成果物である．要件定義工程は，顧客要求を基にソフトウェアの要件を定義する工程であるため，作成の基となる顧客要求が前工程の成果物にあたると考えられる．ただし，顧客要求は明確に文章化されておらず，顧客が暗黙的もしくは経験的にそれらを持っている．例えば，ある業務をソフトウェアとして開発したいという場合には，業務フローが顧客の持っている暗黙知に該当する．したがって，要件定義工程では，暗黙的な顧客要求を基に，要件定義書として形式知化している．そのため，要件定義書の閲読においても，暗黙知である顧客要求を形式知化し，それを前工程の成果物とみなし，突合していると考えられる．設計工程以降では，形式知化された前工程の成果物を基にしていたが，要件定義工程では，暗黙知を基に形式知化した後に，両者を突合していると考えられる．

つまり，要件定義工程では，前工程の成果物が明文化されていないため，構造を理解する際に暗黙的な顧客要求を基に形式知化し，その後は設計工程同様に突合していると考えられる．設計工程以降では，前工程の成果物をより集中した場合に，より品質が高い閲読ができると報告されている(Uwano, 2011)．その事実を基にすると，要件定義の閲読においても，前工程の成果物に相当する顧客要求に集中する方が，閲読品質が高いと想定される．

したがって，要件定義工程での品質の良い閲読には，単に二つの対象（文書やプログラム）を突合する能力だけではなく，暗黙知を形式知化する能力も求められる．既存の欠陥検出率では，そのような能力を測ることはできない．そのため，要件定義書の閲読において，欠陥の検出だけではなく暗黙知を基に形式知化する能力を測る新たな課題も必要である．

1.2.2 閲読能力と経験の関係

品質の良い閲読能力を特徴づけるためには，どのような要因が考えられるのだろうか．ソフトウェア開発の現場では，閲読能力を直接的に測ることは難しく，多くの場合において要件定義工程の実務経験が重要視される傾向がある．特に開発規模が大きければ大きいほど，その傾向は顕著である．

しかし，閲読経験の豊富な人の閲読能力が高いとは限らない．Wong (2003)は,”the experience (i.e. knowledge and skills) of reviewers is the most significant input influencing software review performance.”と述べており，閲読者の経験は閲読品質へ影響を与える要因である．ただし，就業経験は閲読品質に良い影響を与えるものの，閲読経験は品質に大きな影響を与えなかった

結果が示されており(Wong, 2009)，閲読経験が豊富な閲読者が必ずしも良い閲読者とは限らない．閲読経験が豊富だとしても，開発したソフトウェア自体の品質が高くなければその閲読能力は高いとはいえない．さらに，ソフトウェア開発での閲読経験があったとしても，欠陥が少ないソフトウェアが開発できたのかも重要である．しかし，閲読経験にはソフトウェア開発が成功／失敗したという属性情報などは基本的には付与されない．成功したソフトウェア開発に導いたであろう閲読者の経験は，品質良い影響を与える一方で，失敗したソフトウェア開発であればむしろ閲読能力が低い可能性がある．

さらに，閲読経験では同一人物での個々の閲読品質の違いは評価できないという問題もある．ソフトウェア開発の現場では，閲読に十分な時間を費やすことができるとは限らない．そのため，十分な閲読経験を持っている場合でも閲読ごとに品質のばらつきがあると考えられる．閲読経験だけでは，状況や環境の変化に応じた閲読品質の差までは評価できない．

1.2.3 閲読能力と視線の関係

本論文では，閲読能力を特徴づける要因として，視線情報に着目した．近年，視線情報はソフトウェアエンジニアリング（software engineering：SE）にて，様々な種類のタスクにおける認知プロセスの解明に使用されている(Sharafi et al., 2015)．SE に限らず，認知と視線の関係は古くから研究がなされ，特に文章の読みや理解との関連が深いため(Augereau et al., 2016; Campbell & Maglio, 2001; Marcel A Just & Carpenter, 1980; Okoso et al., 2015)，閲読能力を特徴づける要因として適している．

1.2.4 認知と視線の関係

視線情報は以下の四つに大別される．それぞれの特徴は異なっており，異なる認知プロセスと紐づいている．文章を読んでいる際の fixation と saccade の様子を図 1.2 に示す．図 1.2 において円が fixation を示し，円の大きさが fixation の持続時間を示す．Fixation 間をつなぐ線が saccade を示す．

① fixation（固視）：1 箇所を注視している視線の集まり
② saccade（跳躍）：fixation 間の素早い目の動き
③ blink（瞬目）：瞬目の有無
④ pupil（瞳孔）：瞳孔径の大きさ

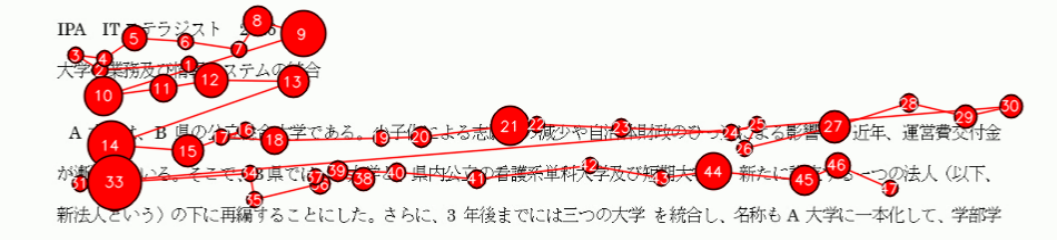


図 1.2　著者が文章を読んでいる際の fixation と saccade の様子をアイトラッカの分析ソフトウェアにて出力した結果（文章の出典：平成 28 年度 秋期 IT ステラジスト試験区分 午後 I 問 1）

1.2.4.1　Fixation と認知の関係

Fixation は特定の個所を注視している視点の集まりであり，主に注意の指標として用いられることが多い. SE での fixation は二つのグループに分けられ，一つは fixation 数に基づく特徴量であり，もう一つは，fixation の持続時間に基づく特徴量である(Sharafi, Shaffer, et al., 2015)．fixation 数に基づく特徴量は，特定の領域（area of interest : AOI）における fixation の総数（fixation 数）や，全体の領域（area of glance : AOG）における fixation に対する割合（fixation 率）などである．Fixation の持続時間に基づく特徴量は，AOI における fixation の持続時間の総時間や平均持続時間などである．

Fixation と理解の関係は，二つの仮定に基づいて解釈されている(Marcel A Just & Carpenter, 1980)．一つは immediacy assumption で，人は名詞や動詞などの内容語に出合うたびに，即座にその解釈を試みるという仮定である．二つ目は，eye-mind assumption で，単語の理解などの処理が完了するまでは，それを注視し続けるという仮定である．もちろん，マインドワンダリング(D'Mello et al., 2016)という「心ここにあらず」といった注意散漫な状態になることもあるものの，基本的には二つの仮定に基づき認知処理が行われていると考えられている．人が何かに注視している場合，その対象に対して注目し，何らかの処理が行われている状態である．

Fixation は，前述したとおり注意の指標として捉えられており，fixation 数が大きければ大きいほどより，対象の AOI に注意関心が向いている(Crosby et al., 2002)．SE の統一モデリング言語（Unified Modeling Language : UML）やコードの閲読において，Fixation 数が大きい場合は，欠陥を検出するために多くの視覚的な努力を費やしていることを示す(Sharif et al., 2013; Sharif & Maletic, 2010).

Fixation 率は，全体における特定の AOI における fixation の割合であり，大きいほど，その AOI に興味を示しているもしくは，複雑で理解するのに困難であることを意味している(Binkley et al., 2013; Poole & Ball, 2006)．検索タスクにおいては，低い fixation 率は，検索における効率の低さを示している

(Goldberg & Kotval, 1999)．SE における UML においては，対象の AOI における fixation 率が高い場合は，それに関連するクラスなどの情報を探すための視覚的な努力が低いことと関連し，検索効率が高いことを示す(Porras & Guéhéneuc, 2010)．Fixation 率はタスクの種類によって，その解釈が異なる．

Fixation の持続時間は，興味の対象に対しての処理時間を示しており，基本的に fixation 数と同様の解釈である．対象の AOI に対する fixation の総持続時間が長ければ，視覚的努力を費やしていることを示している(Bednarik, 2012)．もしくは，被験者がその対象について重要であると考えていると解釈される(Crosby et al., 2002)．

注意の指標と関連して，fixation は認知負荷との関連も報告されており，fixation の数や持続時間が大きいほど認知負荷が高い(Chen et al., 2011; Wang et al., 2014)．前述したように，注視している間は何らかの処理が行われている状態とすると，注視が集中することは，処理の負荷が高い．

まとめると，fixation は基本的には注意の指標として捉えられ，タスクによってその注意の解釈が異なる．理解のタスクでは注意がむけられている対象は，理解しづらいもしくは理解において重要であると解釈できる．その場合は，同時に認知負荷も高い．このような fixation の特性を利用して，難易度の高い単語の推定(大社他，2017)や TOEIC のスコアの推定する試みもなされている(Augereau et al., 2016; Yoshimura et al., 2015)．

1.2.4.2 Saccade と認知の関係

Saccade は fixation 間の素早い目の動き（30–120msec）であり(Sibert & Jacob, 2000)，視線の軌跡を示す．基本的には，視線計測装置を用いた場合は fixation が算出され，saccade は連続する fixation の情報をもとに計算する．saccade の数，持続時間，角度などの特徴量がある．

検索タスクにおいて，saccade の数が多い場合には，検索の量が多いことを示し，インタフェースの評価指標となると報告されている(Goldberg & Kotval, 1999)．他には，Saccade の数や持続時間は，メンタルワークロードと関連していると報告されている(Fritz et al., 2014)．

Saccade は数や持続時間以外にも，文章の読みにおいては戻り読み（regression）の発生を検知できる．文章の読みにおいては，理解が難しい場合には，fixation 数や持続時間の増加に伴い，戻り読みが多くなる(Li et al., 2016; Sanches et al., 2017)．コードの読みにおいても同様であり，戻り読みの割合が少ないほど，よい読み手である傾向が高い(Busjahn et al., 2015)．マインドワンダリングが発生する際には，戻り読みが良く発生するため(Bixler & D'Mello, 2015)，注意散漫な状態であると考えられる．

視線の移動の軌跡を示したものに scanpath があり，fixation だけではわからない，読みや検索における視線の軌跡を示している．図 1.2 の fixation を示

す赤い円には番号が記述されており，文章を読む際の scanpath の一例である．Scanpath が良く使われているのは，マーケティングの分野であり，ホームページなどの広告がどのような順番でみられているのかの分析に用いられている(Eraslan et al., 2016)．それ以外にもコードの閲読においても活用されている．コードの閲読では一つのコード行に対する fixation 数だけでは，良い閲読者の特徴を議論することは難しい．Uwano et al. (2006)は，複数の行をまたがる視線パターンを分析している．コードの閲読ではスキャンと呼ばれるコード全体を眺める行為があり，その時間が短いほど欠陥の検出時間が短くなる傾向がみられた．合わせて，特定の行に集中する場合に欠陥をうまく検出できない傾向がみられた．

1.2.4.3 瞬目と認知の関係

瞬目は大きく随意性瞬目，反射性瞬目，自発性瞬目の三つに分けられる(田多 et al., 1991)．随意性瞬目はウィンクなど意図的に行う瞬目，反射性瞬目は光や音などの外部刺激に対して反射的に行われる瞬目，自発性瞬目は無意識に行われる周期的な瞬目である．本論文では，閲読時の視線情報を取得対象としているため，自発性瞬目に絞り認知の関係を論じる．

瞬目は人の内的な状態を反映していて，その回数が増減するものの，成人以降では年齢による影響は少ない．瞬目は，単位時間当たりの瞬目率によって評価されることが多い．平常時の成人の場合，1 分間の瞬目の回数は約 20 回であり(Bentivoglio et al., 1997; Ponder & Kennedy, 1927; Records, 1979; Tada, 1986)，成人までは加齢ともに変化するものの，成人以降は大きく変化しない(Cruz et al., 2011)．瞬目に影響を与える要因と影響を表 1.1 に示す．瞬目は，環境的な要因だけではなく，身体や心理的状況によっても変化する．多くの研究にて，視覚的な注意が要求される場合には，瞬目が減少することが報告されている(Ledger, 2013; 田多他, 1991)．

計算課題や記憶課題などで心的負荷が高い場合には，瞬目率が高くなるという研究が報告されている一方で，低くなるという報告もある(田多他, 1991)．心的負荷や課題によって瞬目が影響を受けることは示されているものの，課題の種類などを考慮する必要がある．例えば，視覚課題なのか非視覚課題なのかによって影響が変わってくる．田多 (1986)は，視覚課題と聴覚課題における心的負荷と瞬目率の関係を分析したところ，聴覚刺激では負荷の増大とともに瞬目率が増大した一方で，視覚課題では明確な一定の傾向は得られなかった．その原因として，視覚課題では，課題が困難であるほど視覚的な注意を要求されることになるため，注意の集中は一般的に瞬目が低下する(Ledger, 2013)．負荷の増大による瞬目の増大を引き起こすものの，結果として両者によって相殺されていると考えられている(田多他, 1991)．視覚課題において記憶負荷を変化させた場合には，安静時と比べて課題時の瞬目は減少したものの，記憶負荷に応

じて増大した(田多他, 1991).

先行研究の結果を整理すると，視覚課題において注意が要求される場合には基本的に瞬目は減少し，課題による心的負荷の増大に伴い瞬目は増大傾向にある．

瞬目のこれらの増減については，内的注意と外的注意の 2 種類で整理されており，比較的再現性の高いロバストな結果が得られている(Tecce, 1989)．視覚的な注意が必要な課題では，外部に注意が向いているといえ，その場合には瞬目が減少する．例えば，読書など外部の情報を取り込む課題では外的な注意が活性化され，単位時間当たりの瞬目が減少する(Cho et al., 2000; Karson et al., 1981)．心的負荷や記憶負荷などは，内的に注意が向いているといえ，その場合は瞬目が増加する．例えば，暗算や連想などの課題では内的な注意が活性化され，単位時間当たり瞬目が増加する(Cho et al., 2000; Karson et al., 1981)．内的／外的注意による，瞬目へ与える影響は比較的再現性が高く，ロバストな結果である．内的／外的注意による瞬目の違いは,「会話・休憩・読書」の 3 パターンにて綺麗に分かれることが知られている（図 1.3）(Bentivoglio et al., 1997; Karson et al., 1981; Knorr, 1928).

注意の水準以外にも，瞬目を用いて人の集中度合いを定量化する研究も報告されている．JINS は眼電位センサを内蔵した眼鏡型デバイス（JINS MEME）を開発し，瞬目を基に集中度合いを可視化するアルゴリズムを開発している(Uema & Inoue, 2017)．瞬目率の他に，瞬目の強さと顔の位置を基に，独自のアルゴリズムにて集中度合いを定量化している．JINS MEME は日常的に使用する眼鏡として開発され，仕事中や運動中の集中力を可視化できる．

瞬目を用いるメリットは fixation や saccade とは異なり，その取得が比較的簡便な点である．JINS MEME のような眼電位センサであれば，装着者に負担がなく日常的に取得ができる．さらに，fixation を取得するために使用する視線計測装置を用いることなく，パソコンやスマートフォンのカメラでも瞬目は取得できる．

表 1.1　瞬目に影響する要因(田多他, 1991)

	要　因	条　件	瞬目率への影響
視環境	視認性	視覚呈示文字の視認性の悪さ	↑
	照度・輝度	強い照度・輝度	↑
	視野・視角	広い視角	↑
物理環境	聴環境	大きな騒音	↑
	温・湿度	高温低湿度	↑
	空気流速	速い空気流速	↑
眼球恒常変数	眼裂長	大きい眼裂長	↑
	眼球突出度	突出した眼球	↑
眼球状態変数	視力	弱い視力	↑
	コンタクトレンズ着用	コンタクトレンズの着用	↑
	眼科的疾患	眼科的疾患の存在	↑
	目の疲労	目の疲労の増加	↑
	眼球運動	サッケードの多発	↑
		垂直運動の増加	↑
	眼球位置	上方視	↑
身体状態	痛みの知覚	外傷，疾病の存在	↑
	体温	高い体温	↑
	覚醒水準	高い覚醒水準	↑
		わずかな眠気	↑
		強い眠気	↓
心理的変数	状態変数	緊張	↑
		怒り・恐怖	↑
		社会的対人場面	↑
		判断の終了	↑
		高興味刺激の提示	↓
	人格特性	強い神経症傾向，高不安	↑
課題	感覚モダリティ	視覚課題の課題中	↓
	視覚への課題要求	閉眼を抑制する教示	↓
		凝視点の使用	↓
	課題の連続性	不連続・分離試行課題	↑

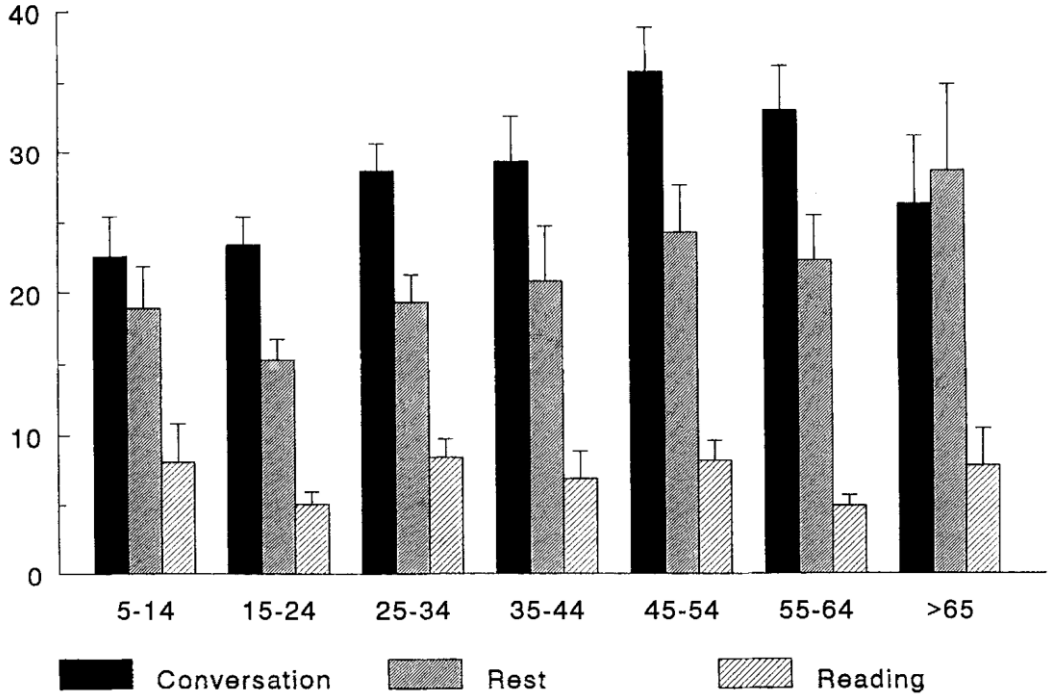


図 1.3　会話・休憩・読書における年齢ごとの瞬目率(Bentivoglio et al., 1997)

1.2.4.4　瞳孔と認知の関係

瞳孔は光に対して大きさを変えることはよく知られており，一般的に 2–8mm のレンジで変化する(宮尾他, 1992; 西山, 2011)．明るいところでは縮瞳により瞳孔径は小さくなり，暗いところでは散瞳により瞳孔径は大きくなる．瞳孔はカメラの絞りのような機能を果たし，瞳孔径が小さいほど焦点深度が大きく，ボケを減少させている(西山, 2011)．

次によく知られている瞳孔の反応は，興味関心がある場合におこる散瞳である(Hess & Polt, 1960)．人が興味関心を持っているものを見る場合には，瞳孔が大きくなり，興味関心がない場合には瞳孔が小さくなると報告されている．

次に，瞳孔径は，認知負荷，ワーキングメモリの負荷，及び心的負荷と関連が報告されている．ワーキングメモリの負荷の増加に伴い，瞳孔径が大きくなる(Kahneman & Beatty, 1966)．読んでいる文章の複雑さと関連して，瞳孔径が大きくなることも報告されている(M A Just & Carpenter, 1993)．SE のコード理解のタスクにおいても同様で，理解が難しく認知負荷がある状態化では散瞳すると報告されている(Fritz et al., 2014)

瞳孔は注意やマインドワンダリングと関連して，縮瞳する(Grandchamp et al., 2014; Kang et al., 2014; Konishi et al., 2017)．Grandchamp et al. (2014) はマインドワンダリング中では，タスク実施時と比較して瞳孔径が有意に減少し，平均値が小さくなると報告している．Konishi et al. (2017)はタスクを実施

していない状態では,タスク時よりも有意に瞳孔径が減少すると報告している.さらに，眠気を感じる前には単調に縮瞳し，眠気を自覚すると大きな揺らぎが発生することが報告されている(西山他, 2008).

したがって，タスクによって認知負荷のように負荷がかかっている状態では散瞳し，負荷がなくなるもしくはタスクから離れるマインドワンダリングのような状態では，縮瞳する傾向がある.

SE の分野では,多くの研究で fixation や saccade が用いられることが多く，瞬目や瞳孔を用いて分析している研究は少ない(Sharafi, Guéhéneuc, et al., 2015; Sharafi, Shaffer, et al., 2015).

近年は，四つの視線情報をもとに算出できる特徴量を複数用いて，機械学習によって認知との関連を報告している研究も多い(Augereau et al., 2016; Bixler et al., 2015; Lima et al., 2018; Okoso et al., 2015; Sanches et al., 2017). Fixation，saccade，瞬目，瞳孔それぞれと認知の関係研究は古くから行われており，様々な知見が得られている．近年は安価なアイトラッカが登場し(Funke et al., 2016)，視線を取得するコストが下がり，機械学習や深層学習が身近になったことで,データドリブンの分析が増えている.複数の視線の特徴量を基に，データドリブンで分析することで，先行研究で明らかにされていなかった新たな知見を得られる可能性が高い.

1.3　本論文の仮説

本研究では，閲読の三つのプロセスと閲読能力の間に関係があると考え，それぞれのプロセスごとの閲読能力を定量化し，視線情報によってその能力の高低に影響を及ぼす要因を明らかにする.

まず，閲読特有のプロセスである構造の理解と欠陥の検出／修正に至っているか否かを視線情報により特徴付ける．実際のソフトウェア開発の現場においては，時間の問題などで十分な閲読がなされるとは限らない．その場合，単なる読みに終始し，閲読特有のプロセスである構造の理解と欠陥の検出／修正に至っておらず，閲読の品質も低いと考えられる．読みにとどまっているのか，構造の理解以降のプロセスの処理がされているのかは瞬目を用いることで推定できる可能性がある．この仮説の検証は第 2 章の実験 1 にて行う.

次に，閲読特有のプロセスである構造の理解と欠陥の検出／修正における閲読能力の違いを定量化し，視線情報を用いてその違いを特徴付ける．そのために，まずそれぞれのプロセスにおいて求められる能力を議論する.

閲読は，前工程と現工程の成果物の突合であることは述べたが，突合をするためには両者が突合可能な状態となっている必要がある．閲読の三つのプロセスの構造の理解において，突合可能な状態へ変換処理を行っている可能性がある（図 1.4）．二つの対象を突合するためには，両者が比較可能な状態に変換さ

れていなければならない．機械には，自然言語で記述された設計書とプログラムの突合は基本的にはできず，突合するためには形式仕様のような形で記述されている必要がある．人はそのような状況においては，設計書またはプログラム，もしくは両者を突合可能な状態に変換している．例えば，設計書を基に頭の中でプログラムに変換し，閲読対象のプログラムと突合するなどしている．したがって，閲読における構造の理解では，前工程の成果物と現工程の成果物を突合可能な状態へ変換していると考えられる．さらに要件定義工程では，前工程の成果物は存在しないため，暗黙知である顧客要求を形式知化し，突合していると考えられる．したがって，暗黙知を形式知化したものと要件定義書を基に，突合可能な状態に変換して，突合により欠陥の検出／修正を行っていると考えられる．突合においては，要件定義書が顧客要求を必要十分に満たすかを判別していると考えられる．まとめると，閲読特有のプロセスにおいて求められる能力は，二つの閲読対象を突合可能な状態に変換し，一方が他方を必要十分に満たすかを判別する能力である．

しかし，この二つの能力を定量的に測る指標は存在しないため，要件定義工程を模したゲームを開発することにより閲読能力を測る課題を作成し，閲読能力と視線の関係を分析する．課題の作成は第 3 章の実験 2 にて，課題のスコアと視線の関係の分析を第 4 章の実験 3 にて行う．

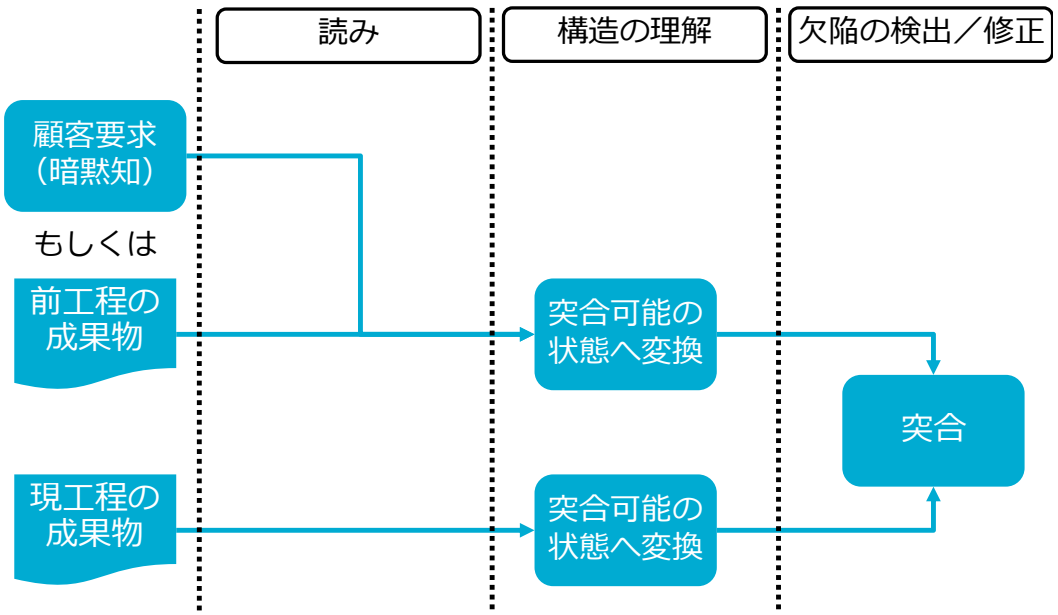


図 1.4　閲読における三つの認知プロセス

1.4　本論文の構成

本論文では，閲読能力と視線の関係を調査するために，三つの実験を行った．

第 2 章では，実際の業務にて使用された要件定義書を用いて，意図的に欠陥

を混入させて，閲読する実験を実施した．本実験では，混入させた欠陥に対する検出感度を導入し，感度と視線の関係を分析した．欠陥検出感度は，欠陥検出率とは異なり，対象の欠陥に対する検出感度を示しており，重要な欠陥を検出できる能力である．特に視線情報のうちの瞬きに着目し，感度との関係を議論する．

第 3 章では，1.2 で定義した閲読の認知プロセスである，顧客要求と要件定義書の突合能力を測る課題を作成する実験を行った．実際の業務にて使用された要件定義書を用いた第 2 章の実験では，顧客要求は被験者に明確に提示されなかった．実際のソフトウェア開発では，顧客要求は複雑かつ多岐に渡っているため，実際の要件定義書の顧客要求を明示することは難しい．そのため，本実験ではソフトウェアを幾何学図形の集合で象徴化し，それを基に要件定義書のペアを作成する実験を行う．作成した課題を用いることで，顧客要求と要件定義書を突合する能力を測ることができる．

第 4 章では，第 3 章で作成した顧客要求と要件定義書のペアを用いて，閲読する実験を行った．顧客要求と要件定義書の突合能力に影響を与える視線情報を分析し，品質の高い閲読者の特徴を明らかにする．

第 5 章では，三つの実験で得られた結果をまとめ，閲読能力に影響を及ぼす視線の特徴を整理し，本論文のまとめと今後の展望を記述する．

第2章　実験１：欠陥の検出感度と視線の関係[1]

2 章では，実際の開発で使用された要件定義書を用いて，意図的な欠陥を含ませて，含ませた欠陥に対する検出感度と視線の関係を議論する．閲読品質を測る既存の指標である欠陥検出率に代わり，欠陥の検出感度という新たな指標を提案する．視線情報のうち，特に瞬目に着目して，欠陥の検出感度との関係を分析した．その結果，欠陥の検出感度との瞬目率の間に有意な相関関係が存在し，瞬目が閲読品質を特徴づける重要な特徴量であることがわかった．

2.1　実験の目的

本実験では，閲読品質を測る新たな指標を提案する．序論でも述べたように，閲読品質の既存手法である欠陥検出率(Thelin et al., 2003; Uwano, 2011)は，正確に品質を評価できない．欠陥の検出数が多いとしても，後工程で重大なバグに繋がらないような欠陥が多数を占めるのであれば，その品質が高いとはいえない．つまり，要件定義書において重要である欠陥を感度よく検出できるかが重要であると考えられる．

そこで，本研究では，欠陥の検出感度を新たに定義した．実際のソフトウェア開発で用いられた要件定義書を基に，意図的な欠陥を混入させて，欠陥の検出感度を算出した．検出感度は，意図的に混入させた後工程でのバグに繋がる欠陥を感度よく検出できているかを示す．上流工程で検出された欠陥は omission（抜け漏れ）と commission（誤り）の二つに代別されるため(Porter & Votta, 1998)，本実験での欠陥は，実験上の操作が容易な抜け漏れを対象とする．

閲読では(1)読みに加えて,(2)構造の理解と(3)欠陥の検出／修正の認知プロセスがあり(Uwano et al., 2007)，それらと関連して，過去研究では三つの瞬目の特徴について以下のように報告している：

(A). 平常時の成人の場合，1 分間あたりの瞬目の回数は約 20 回である．(Bentivoglio et al., 1997)

(B). 読書など外部の情報を取り込む課題では外的な注意が活性化され，単位時間当たりの瞬目が減少する．(Cho et al., 2000; Karson et al., 1981).

(C). 暗算や連想などの課題では内的な注意が活性化され，単位時間当たり瞬目が増加する(Cho et al., 2000; Karson et al., 1981).

上記の結果を基にして，閲読プロセスの三つのサブプロセスは以下のように視線情報と関連すると仮説をたてた：サブプロセス(1)では外的注意を求められ

[1] 本実験は日本認知科学会第 35 回大会抄録集(斉藤・土肥, 2018)，Unisys 技報(斉藤, 2019)，研究業績 2 の Proceedings of The 41th Annual Meeting of the Cognitive Science Society (CogSci2019) (Saito & Hidaka, 2019)，研究業績 1 の認知科学 2020 年 27 巻 3 号(斉藤・日髙, 2020)にて発表・掲載された論文としてまとめられた研究成果に該当する．

るため，結果(B)と関連して，瞬目率が減少すると考えられる．サブプロセス(2)(3)では内的注意が求められるため，(C)と関連して，瞬目率が増加すると考えられる．そして，良い閲読者とそうでない閲読者では(1)のサブプロセスは共通であるが，(2)(3)のサブプロセスで差異が発生し，瞬目の特徴が異なると想定される．良い閲読者はそうでない閲読者と比べて，サブプロセス(2)(3)により認知資源を費やすため，瞬目が平常時より増加すると考えた．

そこで，本研究では要件定義書閲読を模した実験を行い，実験中の視線情報を計測し，瞬目と閲読品質の間の関係を調査した．実験では，閲読対象の要件定義書に意図的な欠陥を含ませ，欠陥の検出有無を基にして閲読品質を定義した．

2.2　実験の概要

本実験では，要件定義書を閲読し，欠陥を検出する課題に実施し，その際の視線情報を基に欠陥検出感度との関係を分析する．本節では，欠陥検出感度を用いた分析の概要を説明する．

分析 1～3 では，意図的に混入した欠陥に対する感度を算出して，視線情報との関連を分析し，合わせて意図しない潜在的な欠陥の有無についても調査する．実験で用いる要件定義書は実際のソフトウェア開発で用いられたものであり，一定の品質は担保されていると想定されるものの，意図しない潜在的な欠陥が含まれている可能性は否定できない．そのため，被験者が意図しない潜在的な欠陥に反応する可能性がある．しかし，意図しない潜在的な欠陥をすべて排除することは難しく，被験者の検出バイアスにも依存する．意図的に混入した欠陥に反応している場合は，欠陥の検出感度は正の値を示し，一方で意図しない欠陥に対して強く反応する場合には負の値を示す．そのため，分析 2 における欠陥検出感度はゼロ付近で最も感度が低く，ゼロから離れるほど意図的な欠陥か否かによらない欠陥に対する感度を示すと想定される．欠陥の検出感度を用いることで，意図しない潜在的な欠陥の有無を調査し，続く分析 4 ではそれらの欠陥を考慮し，視線情報との関連を分析する．

その後，分析 4 にて意図しない潜在的な欠陥を考慮し，意図しない欠陥も含めた欠陥に対する感度を算出し，同様に視線情報との関係を分析する．意図しない潜在的な欠陥は，被験者の検出バイアスにも依存するため，実験設計時点ですべて排除することは難しい．そのため，本実験では被験者が指摘した欠陥を基にして，意図しない潜在的な欠陥を分類し，再度欠陥の検出感度を算出する．そのため，分析 4 にて算出する欠陥検出感度の方が，分析 1～3 と比較して被験者の閲読能力をより反映しているものであると考えられる．

2.3　実験手続き

閲読セッションと閲読後セッションの二つのセッションに分け実験を実施した．閲読セッションのそれぞれの試行で，被験者には 3 種類の要件定義書の 1 頁を閲読し，その後欠陥を含む箇所に下線を引くよう指示した．閲読セッションの 11 試行が終了した後，質問紙調査に回答してもらった．

2.3.1　被験者

被験者は 19 名の正視の成人であり，その内訳は男性 16 名，女性 3 名であった．30 代が 9 名，40 代が 4 名，50 代が 6 名であり，平均年齢は 42.2 歳，標準偏差は 9.1 であった．全員システムエンジニアであり，そのうちの半数は要件定義書の閲読経験がなかった（図 2.1）．

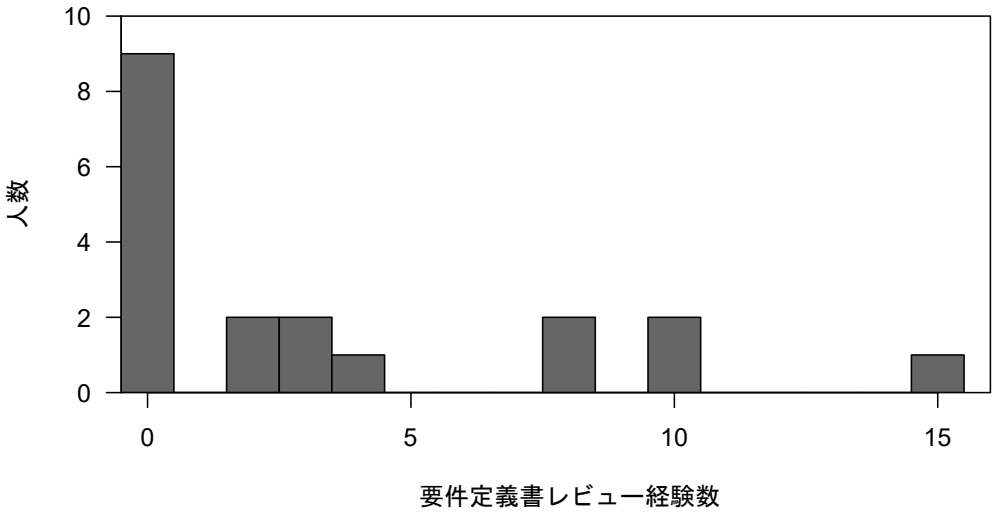


図 2.1　被験者の要件定義書の閲読経験分布

2.3.2　閲読対象文書

実験に使用する閲読対象文書は，日本ユニシス株式会社で実際に使用された 3 種類の要件定義書を基にした．それぞれの要件定義書は概要／機能要件／非機能要件の 3 頁の構成に改編し，サンプル文書を二つ加えた計 11 頁とした．サンプル文書は IPA の情報処理技術者試験の問題より抜粋した．改編した要件定義書のそれぞれの頁に意図的な欠陥を含ませた．各要件定義書において最初の頁で対象システムの背景または目的を閲読する設計とし，全ての被験者がシステムの前提知識を等しく得られるようにした．使用した文書の詳細と欠陥を含む文数を表 2.1 に，実験時の画面サンプルを図 2.2 に示す．本研究では，欠陥の種類は抜け漏れとし，要件定義に必要な情報の一部を削除した．削除する

単位は一つまたは連続する二つの文節とした．一つの文節を削除した欠陥文の例を以下に示す．

欠陥文）…，各店舗より利用実績情報を受信する．

正解文）…，各店舗より利用実績情報を日次（夜間）に受信する．

欠陥が多いと文の不自然さが増し，通常の閲読時の視線の動きが計測されない可能性がある．そこで，1 頁あたり欠陥を含む文は最大 2 個までとし，11 頁全体では欠陥を含む文は 17 個とした．質問紙調査では，年齢／性別／要件定義書の閲読経験（経験したプロジェクトの数）／文書（要件定義書，設計書，報告書などの文書全般）の閲読頻度（ほとんどない／1 月に 1 度程度／2 週間に 1 度程度／1 週間に 1 度程度）／閲読時の集中度（1 できなかった／2 あまりできなかった／3 どちらともいえない／4 まあできた／5 できた）／閲読対象文書への理解度（1 できなかった／2 あまりできなかった／3 どちらともいえない／4 まあできた／5 できた）を収集した．

表 2.1　実験で使用した閲読対象文書の詳細情報

種別	頁番号	内容	文字数	全文数	欠陥を含む文数
サンプル文書	1	仕入れ納品システムの概要 （出典：平成28年度 秋季 システムアーキテクト試験 午後 問1）	116	2	0
	2	手術支援システムの概要 （出典：平成26年度 秋季 システムアーキテクト試験 午後 問4を一部改変）	140	2	1
要件定義書1	3	概要：本ビジネスの背景	519	7	2
	4	機能要件：開発対象機能	367	8	2
	5	非機能要件：制約，可用性要件	458	10	1
要件定義書2	6	概要：システム化の目的	474	6	2
	7	機能要件：開発対象機能	432	14	2
	8	非機能要件：運用要件	393	9	2
要件定義書3	9	概要：システム化の目的	581	7	2
	10	機能要件：開発対象機能	532	11	2
	11	非機能要件：可用性要件	526	12	1
合計				88	17

図 2.2　実験時の画面サンプル（社外秘情報が含まれるため本文を黒塗りとする）

2.3.3　実験手続きと装置

それぞれの試行にて，被験者に対して上述の閲読対象文書をモニタ上に提示し，閲読時の視線をアイトラッカで計測した．アイトラッカは視線情報を計測する専用のデバイスであり，本実験では gazepoint 社の GP3 HD Eye Tracker（サンプリングレート：150Hz，正確度（Accuracy）：0.5-1.0°）を用いた（図 2.3）．閲読時間の制限は設けなかった

被験者には，モニタ上の要件定義書を 1 頁閲読した後，紙の要件定義書に対して改善すべき箇所に下線を引くように指示した．下線を引く範囲は制限せず，文の一部でも全体でも可能とした．被験者には意図的に含めた欠陥の種類は伝えなかった．この試行を 11 頁分繰り返し，頁順は全被験者で同一とした．なお，途中休憩は挟まなかったが，測定の精度を確保するためになるべく頭部を動かさないように指示した．

閲読セッション終了後，被験者は質問紙調査に回答した．

図 2.3 (左)実験の様子,(右)アイトラッカ（モニタ下部にアイトラッカを設置）

2.4 結果

閲読セッションでの被験者 19 名の閲読時間の平均／最小／最大は，それぞれ 21 分／8 分／40 分であった．被験者 19 名に対して 11 頁分の視線情報を取得した計 209 頁のデータには，視線情報が正しく計測できていない頁が存在する可能性がある．その原因としてはアイトラッカが被験者の目の検知に失敗することが考えられ，その場合，後述する fixation（固視）の情報が無効となる．そこで，スミルノフ・グラブス検定(Grubbs, 1969)を行い，有効な fixation の割合が 0.6 未満の 4 頁分（図 2.4）を視線情報の計測不良データとして削除した．以降の分析では，計測不良データの 4 頁分を削除した計 205 頁のデータを用いた．

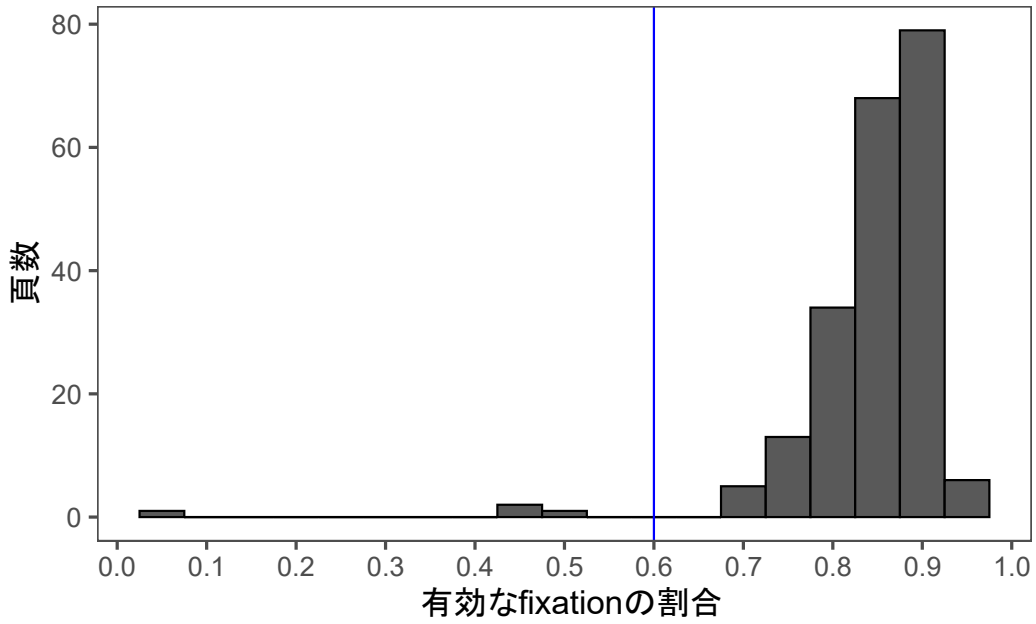


図 2.4　頁ごとの有効な fixation の割合のヒストグラム

2.4.1　閲読品質

本研究では下線を引いた箇所が，意図的に含んだ欠陥部分を一部でも含む場合に正しく欠陥を検出できたとした．そこで，本タスクを信号検出として定式化した．本タスクでのシグナルは意図的な欠陥であり，文をノイズとして信号検出理論を適用する．被験者は欠陥を含む文（シグナル＋ノイズ）またはそうでない文（ノイズ）を与えられた際に，欠陥を検出したかどうかを報告する．閲読品質として，信号検出理論(Green & Swets, 1966)で算出される d-prime を用いた．信号検出理論では，被験者の反応バイアスと検出力（d-prime）は，欠陥を含む文のうち下線を引いた文の割合（正答率）と欠陥を含まない文のうち下線を引いた文の割合（誤答率）から独立した指標として算出できる．d-prime は，シグナル+ノイズの分布の平均値とノイズの平均値との差をノイズの分布の標準偏差で割った値であり，次式

$$\text{d-prime} = \frac{M_{SN} - M_N}{\sigma_N} \tag{2.1}$$

で定義される．ここで，M_{SN}はシグナルとノイズの分布の平均，M_Nはノイズの分布の平均，σ_Nはノイズの分布の標準偏差である．タスクにおけるシグナル＋ノイズとノイズに対する反応を用いて，次式

$$\text{d-prime} = \text{Z}_N - Z_{SN} \tag{2.2}$$

で計算する．ここで，Z_Nは$(1 -$誤答率（false alarm 率）$)$を z 得点に変換した値，Z_{SN}は$(1 -$正答率（hit 率）$)$を z 得点に変換した値である．誤答率は

$\frac{False\ alarm}{False\ alarm+Correct\ rejection}$, 正答率は$\frac{Hit}{Hit+Miss}$で定義され, Hit, Miss, False alarm, Correct rejection は表 2.2 での四つの指標である.

正答率（hit 率）と誤答率（false alarm 率）は被験者の潜在的な応答バイアス（シグナルの有無に関係なく，検出を報告する行動傾向）に影響を受けるため，閲読指標としては適していない．例えば，被験者が全文に下線を引いた場合には，正答率は 100%となるものの，その閲読品質は高いとはいえないため正答率では品質を判別できない．しかし，その場合は誤答率も 100%となり，d-prime は低い値を示す．したがって，d-prime は被験者の潜在的な応答バイアスの影響を回避できるため，閲読品質の指標として適している.

合わせて，d-prime と比較する目的で，F 値を用いて瞬目率との関連を議論する．F 値は次式

$$\text{F 値} = \frac{2Recall \cdot Precision}{Recall + Precision} \tag{2.3}$$

で定義され, recall と precision の調和平均である．Recall, precision は表 2.2 の四つの指標を用いて，それぞれ次式

$$\text{Recall} = \frac{Hit}{Hit + Miss} \tag{2.4}$$

$$\text{Precision} = \frac{Hit}{Hit + False\ alarm} \tag{2.5}$$

で定義され，recall と hit 率は同じ値を示す.

表 2.2　信号検出理論における反応の分類

	欠陥ありと反応	欠陥なしと反応
シグナル（意図的な欠陥）がある文	Hit	Miss
シグナル（意図的な欠陥）がない文	False alarm	Correct rejection

2.5　分析

仮説を検証するため，分析 1 では経験と d-prime により測定された閲読品質の相関関係を，分析 2 では瞬目率と閲読品質の相関関係を分析した．分析 3 では瞬目率を含むそれ以外の視線情報（fixation，saccade など）と閲読品質との関係をモデルベースで分析した．分析 2 にて，本研究で意図しない欠陥が含まれていることが示唆されたため，分析 4 では，意図しない欠陥を考慮した閲読品質と経験／瞬目率の相関関係を分析した．なお，相関関係の算出やモデルの構築に必要な統計解析は，フリーソフトウェアの R 言語（R version 3.4.1）で行った.

2.5.1 分析 1：経験と閲読品質の関係

質問紙調査では，要件定義書の閲読経験数と文書閲読の頻度の 2 種類の閲読経験に関する情報を収集した．文書閲読は要件定義書に限らず，設計書や報告書などの文書全般の閲読を対象とし，その頻度の情報を収集した．分析では要件定義閲読経験数は比率尺度として，文書の閲読頻度は順位尺度として扱う．そのため，文書の閲読頻度を含む場合の相関係数算出にはスピアマンの相関係数を，含まない場合にはピアソンの相関係数を用いた．

最初に，2 種類の閲読経験の関係を確認するため，散布図を図 2.5 に示す．二つの経験の間には正の有意な相関がみられた（スピアマンの相関係数 $r = .535, p = .018$）．一般に閲読者となる前には作成者としての経験を積み，その後閲読者となる傾向がある．そのため，要件定義書の閲読経験がゼロの被験者は，日々の業務において作成者としての役割を果たしているため，文書の閲読頻度が低くなったと考えられる．一方で，要件定義書の閲読経験が一度でもある被験者は閲読者としての役割も果たすことができるため，文書の閲読頻度が高くなる傾向にあった．ただし，文書の閲読頻度は調査時の業務内容にも依存するため，要件定義書の閲読経験があるにもかかわらず頻度がほとんどない場合も存在した．

閲読経験と閲読品質の関係を評価するため，それぞれの経験と d-prime の散布図を図 2.6，図 2.7 に示す．2 種類の閲読経験と d-prime の間には有意な相関はみられなかった（要件定義閲読経験数と d-prime のピアソンの相関係数 $r = -.149, p = .542$，文書の閲読頻度と d-prime のスピアマンの相関係数 $r = -.078, p = .750$）． 2 種類の閲読経験と d-prime の絶対値にも有意な相関は見られなかった（要件定義閲読経験数と d-prime 絶対値のピアソンの相関係数 $r = -.323, p = .178$，文書の閲読頻度と d-prime 絶対値のスピアマンの相関係数 $r = -.380, p = .108$）．これらは仮説を支持する結果であり，閲読経験が豊富な閲読者が良い閲読者とは限らないということを示唆している．

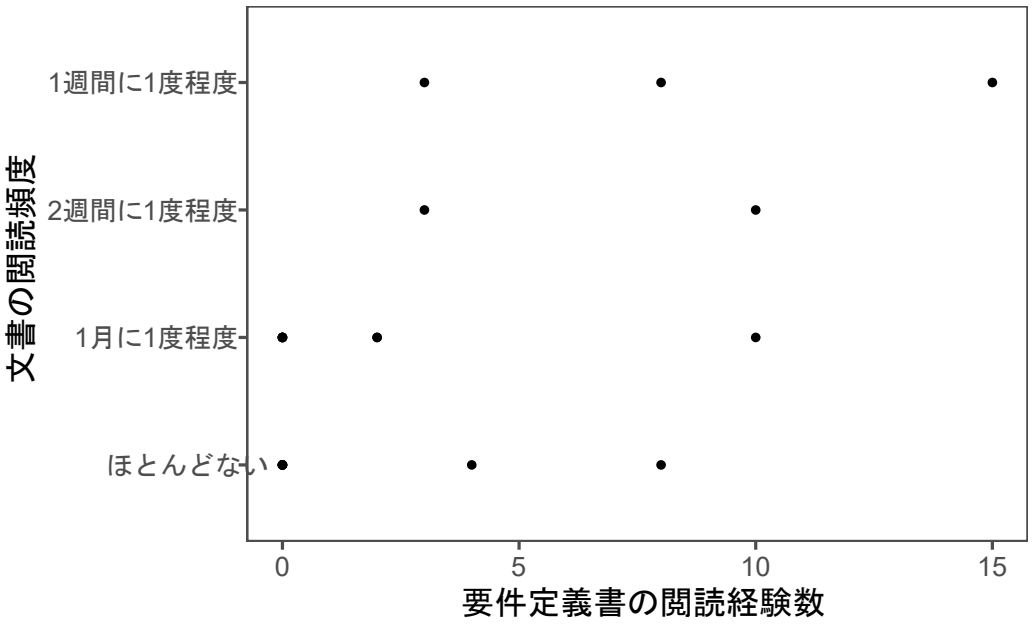


図 2.5　要件定義書の閲読経験数と文書閲読の頻度の散布図

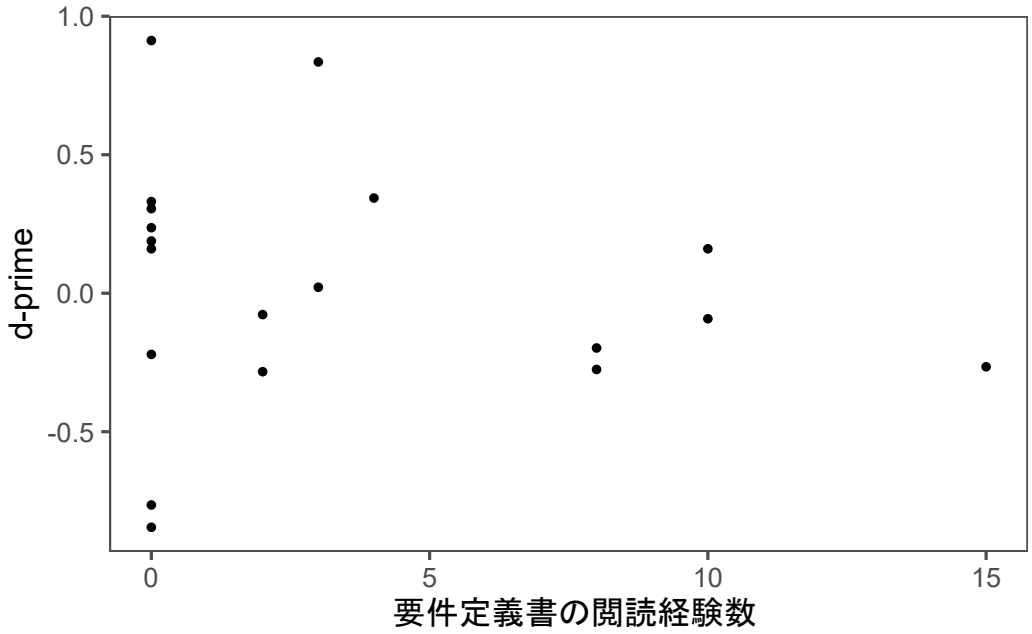


図 2.6　要件定義書の閲読経験数と d-prime の散布図

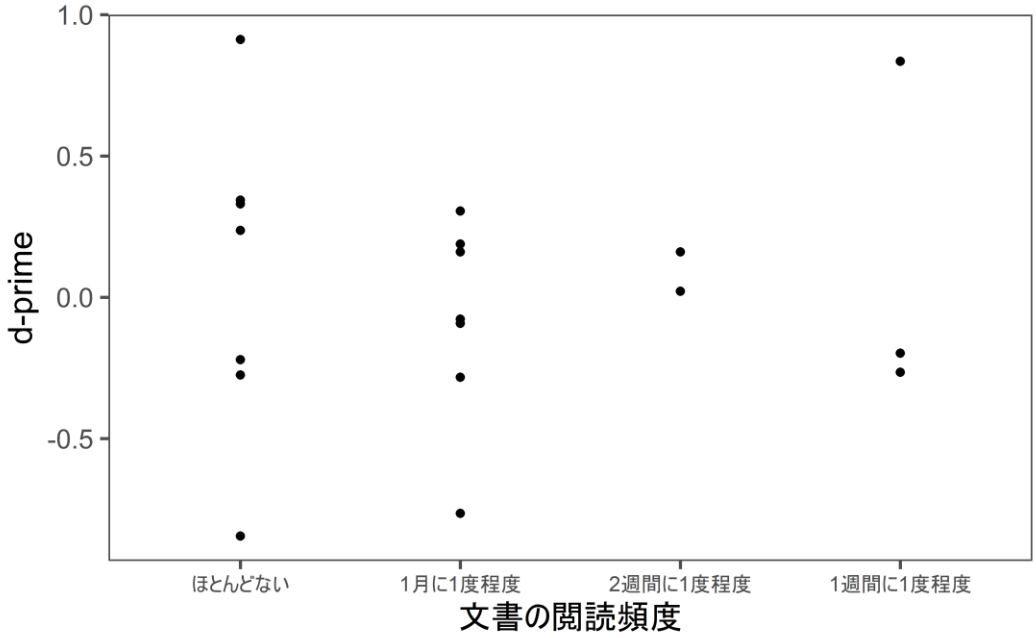


図 2.7　文書の閲読頻度と d-prime の散布図

2.5.2　分析 2：瞬目率と閲読品質の関係

本研究の仮説では，閲読において重要なサブプロセスでは内的注意が要求され，瞬目率が増加すると想定される．瞬目率と閲読品質の関係を確認するために，瞬目率と d-prime の散布図を図 2.8 に，相関係数を算出した結果を表 2.3 に示す．Maximal information coefficient（MIC）(Reshef et al., 2011)は相互情報量を用いて算出される相関係数であり，非線形な関係でも適用可能である，MIC-ρ^2 は非線形性の指標（値が大きければ非線形）を示し，maximal asymmetry score（MAS）は単調性の指標（値が小さければ単調）である．これらの結果から次の事実が確認できる：

- d-prime は-1 から+1 の間に分布した．
- d-prime がゼロ付近では瞬目が平均より減少し，ゼロから離れるにつれて瞬目が増加した．
- ピアソンの相関係数は有意な負の弱い相関を示した．
- MIC と MIC-ρ^2 の値が大きく，非線形な相関を示した．

以上の事実から，d-prime と瞬目率は，d-prime がゼロ付近で瞬目率が最小となるような U 字型もしくは V 字型の非線形な相関を示すことがわかった．

これに加えて，表 2.11 に示す被験者の検出した欠陥の分類を合わせて考えると，正と負の d-prime が存在したことは，被験者内で二つの異なるグループが存在したことを示唆している（被験者の d-prime と属性情報は付録 A.1 を参照）：一つのグループは本実験で混入させた欠陥を検出していた．一方のグループはランダムな文だけではなく本実験で十分には統制できなかった異なる種類

の欠陥を検出していた．実験では意図的に欠陥を混入させたが，元の要件定義書に特定の異なる種類の欠陥が含まれていた可能性がある．その場合，正の d-prime は本実験で含ませた欠陥に対する検出感度を．負の d-prime は元の要件定義書に含まれていた特定の予期しない欠陥に対する検出感度を示している．

全被験者の d-prime とノイズへの反応数の散布図を図 2.9 に，その相関係数を表 2.4 に示す. 負の d-prime の被験者は正の d-prime の被験者と比較してノイズへの反応数が多くなり，シグナルをノイズとして検出する傾向にあることがわかった．したがって，実験の教示にて欠陥の種類を提示しなかったことにより，負の d-prime の被験者は本実験ではノイズに分類した他の種類の欠陥を検出していると考えられる．

正と負（ゼロ以外）の両方の d-prime は，特定の種類の欠陥に関する感度が高いことを示していると解釈できる．さらに，瞬目率は潜在的に混在するタイプの欠陥の閲読品質と有意に相関していた．

続いて，瞬目率と F 値の散布図を図 2.10 に示す．両者の間に正の有意な相関がみられた（ピアソンの相関係数 $r = .166, p = .0171$）．F 値は recall と precision の調和平均であり，共に hit（欠陥に対する反応）を分子とする指標である．そのため，d-prime のようにノイズに対する反応の強さは測れない．しかし，瞬目率と正の有意な相関が得られたことから，d-prime が正の場合における意図的な欠陥に対する感度指標としては F 値でも十分であると考えられる．

これらの結果は本研究の仮説を支持する．しかし，瞬目率以外の視線情報も含めた場合，他の視線情報が潜在要因となり疑似相関となっている可能性がある. そこで, 分析 2 では, 他の視線情報も含めたモデルベースの分析を行った．d-prime を予測する上で fixation や saccade などの他の視線情報と比較して，瞬目率の重要度を評価した．

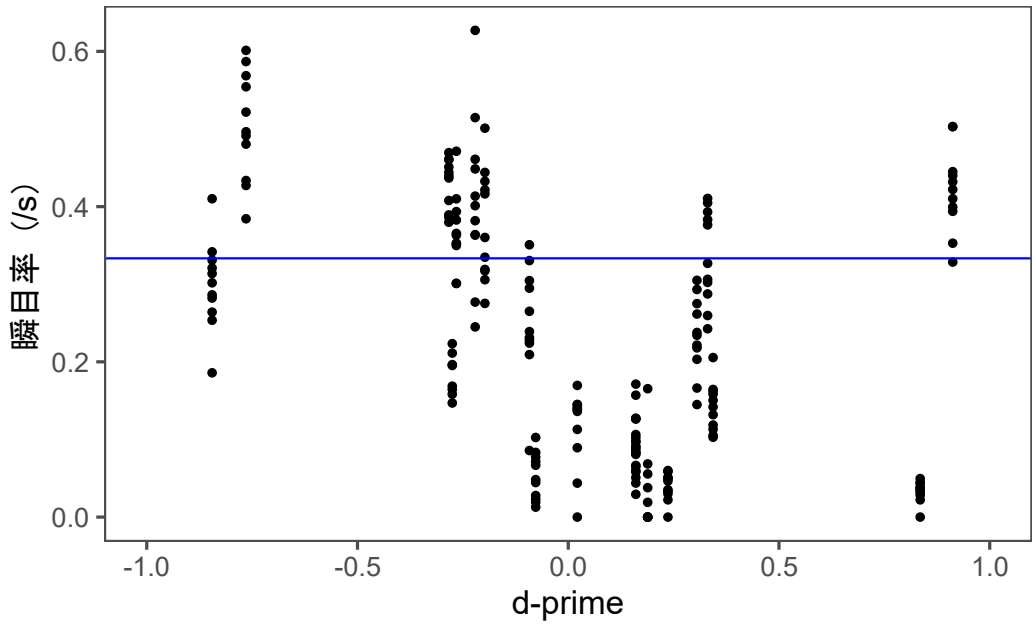


図 2.8　d-prime と瞬目率の散布図（青線は瞬目率の平均値を示す）

表 2.3　d-prime と瞬目率の相関係数

		瞬目率
d-prime	ピアソンの相関係数	-.393
	p値	.000
	MIC	.865
	MIC-ρ^2	.711
	MAS	.450

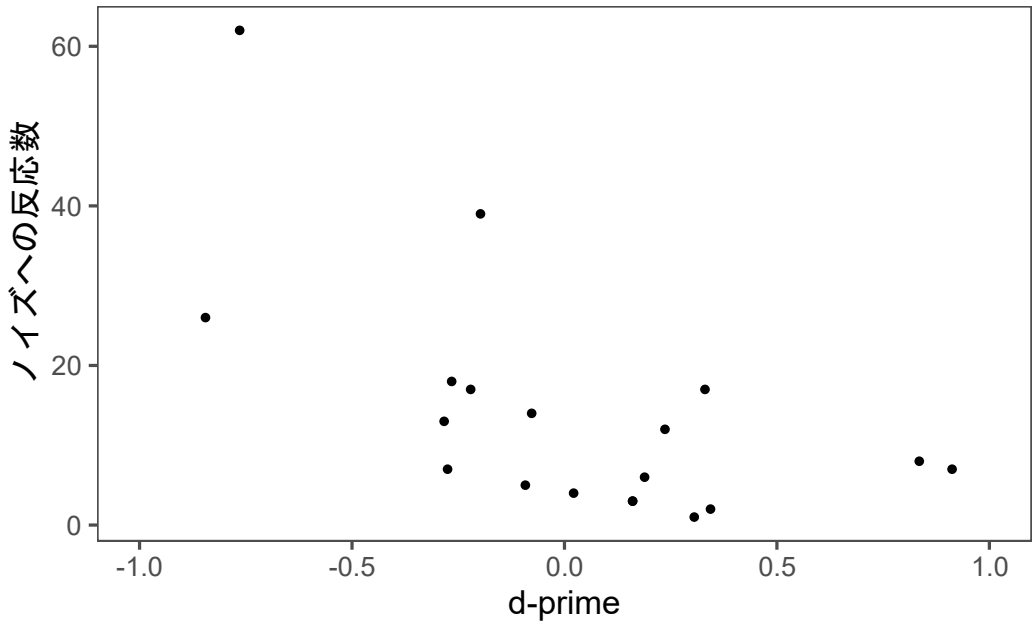


図 2.9　d-prime とノイズへの反応数の散布図

表 2.4　d-prime とノイズへの反応数の相関係数

		ノイズへの反応数
d-prime	ピアソンの相関係数	−.618
	p値	.005

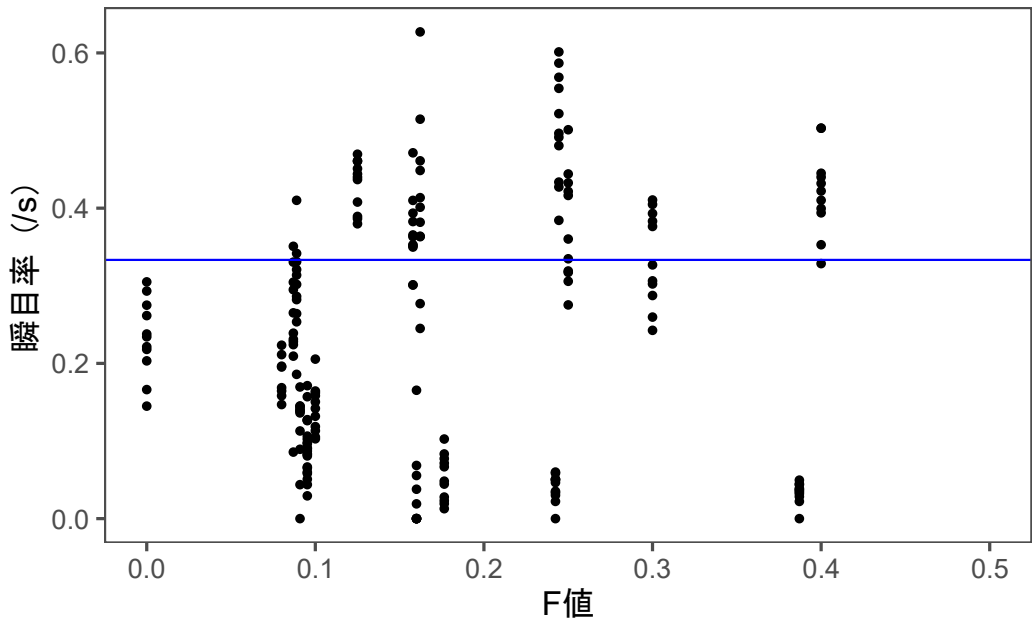


図 2.10　F 値と瞬目率の散布図（青線は瞬目率の平均値を示す）

2.5.3 分析 3：閲読品質のモデルベース分析

瞬目率と他の視線情報を用いて，機械学習アルゴリズムにて閲読品質予測モデルと d-prime 正負分類モデルを構築する．まず，視線情報から特徴量を算出し，閲読品質に影響を及ぼしている特徴量を抽出する．その後，抽出した特徴量を使って機械学習を行い，閲読品質を予測するモデルと d-prime の正負を分類するモデルを構築する．

2.5.3.1 特徴量の抽出

アイトラッカで取得できる以下の四つの基本情報（fixation, saccade, blink, pupil）を基に，47 個の特徴量を抽出した．Fixation は，gazepoint 社の Gazepoint analysis に実装されている独自アルゴリズム（Jacob の研究(Jacob, 1995)を基に作成）を用いて算出された．47 個のうち 46 個は Bixler らの研究(Bixler et al., 2015)で定義された特徴量を用い，本研究の目的のために瞬目率を特徴量に加えた（47 の特徴量の詳細は付録 A.2 を参照）．47 個の特徴量は，被験者ごとに 1 頁ごとの四つの基本情報を用いて算出した．

① fixation（固視）：1 箇所を注視している視線の集まり
② saccade（跳躍）：fixation 間の素早い目の動き
③ 瞬目：瞬目の有無
④ 瞳孔径：瞳孔径の大きさ

47 個の特徴量を基に，Random forest（以降，RF）(Breiman, 2001)による機械学習アルゴリズムを使って，閲読品質に影響を及ぼす特徴量を抽出した．RF は各特徴量の重要度も同時に算出できるため，重要度の低い特徴量は閲読品質への影響は少ないとして削除できる．RF における重要度はジニ係数を基にした．

特徴量の抽出は，RF 重要度を基にして変数増加法を用いた．最初に，全データ（205 頁分のデータ）を用いて RF を実行して特徴量ごとの重要度を算出した，最も重要度が高い特徴量を用いて RF を実行して実測値との誤差である RMSE を算出する．RMSE は次式

$$\mathbf{RMSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}}(\boldsymbol{y_k} - \boldsymbol{\hat{y}_k})^{\mathbf{2}}} \qquad (\mathbf{2.6})$$

で定義される．ここで，n は頁数，y_kは k 頁目の d-prime の実測値，$\hat{y}_k$はモデルによる予測値である．その後，次に重要度が高い特徴量を追加して RF を実行し，同様に RMSE を算出する．RMSE が最も小さくなるまで繰り返し，閲読品質に有効な特徴量を抽出する．

特徴量の抽出によって得られた 15 個の特徴量と重要度を表 2.5 に示す．瞬目率が最も重要度が高く，瞬目率が d-prime に最も影響を与える特徴量であることがわかった．

表 2.5　RF により抽出された特徴量と重要度

順位	特徴量	重要度
1	瞬目率	2.23
2	saccadeの持続時間の尖度	2.09
3	fixationの持続時間/saccadeの持続時間の比率	1.93
4	瞬目回数	1.93
5	saccadeの持続時間の最大	1.88
6	saccadeの持続時間のレンジ	1.80
7	瞳孔の直径の歪度	1.77
8	fixationの持続時間の最小	1.56
9	瞬目時間の割合	1.50
10	saccadeの持続時間の平均	1.31
11	saccadeの持続時間の標準偏差	1.30
12	saccadeの持続時間の歪度	1.27
13	水平なsaccadeの割合(水平軸から30度未満の角度)	1.14
14	saccadeの距離の中央値	1.08
15	fixationの持続時間の中央値	1.01

2.5.3.2　閲読品質予測モデル

抽出した 15 個の特徴量を用いて，閲読品質予測モデルを複数のアルゴリズムで構築した．本研究で採用したアルゴリズムは，RF，Decision tree（以降，DT），Support vector regression（以降，SVR）(Chang & Lin, 2011)，および Multiple linear regression（以降，MLR）である．各アルゴリズムは R のパッケージを用いて，パラメータ等のチューニングは行わずライブラリのデフォルト値を採用した（表 2.6）.

各アルゴリズムによる閲読品質予測モデルの有効性の確認は全データ（205 頁分のデータ）で 10-fold cross validation を行い，mean absolute errors(MAE), RMSE を算出して比較する．MAE は次式

$$\mathbf{MAE} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}}|\boldsymbol{y_k} - \boldsymbol{\hat{y}_k}| \qquad (\mathbf{2.7})$$

で定義される．

閲読品質予測モデルを構築し，アルゴリズムごとに得られた結果を表 2.7 に，d-prime 実測値と SVR で予測した d-prime を図 2.11 に示す．SVR による閲読品質予測モデルの MAE，RMSE が最も低く，次いで RF である．モデルの有効性を確認するため，d-prime の実測値と予測値の散布図を図 2.11 に，相関

係数を表 2.8 に示す.

以上の結果をまとめると，47 個の特徴量のうち 15 個が閲読品質に影響を与えており，そのうち「瞬目率」が最も高い重要度を示した．モデルベースの分析の結果は分析 2 の結果を支持し，分析 2 で示された瞬目率と閲読品質の相関は疑似相関ではないと示唆される.

表 2.6　アルゴリズムごとのライブラリとパラメータ

アルゴリズム	ライブラリ	パラメータ
RF	randomForest	全てデフォルト値を使用
DT	rpart	全てデフォルト値を使用
SVR	e1071	全てデフォルト値を使用
MLR	基本パッケージ（lm()）	全てデフォルト値を使用

表 2.7　アルゴリズムごとの閲読品質予測モデルの性能

		RF	DT	SVR	MLR
d-prime	MAE	0.224	0.323	**0.214**	0.283
	RMSE	0.304	0.451	**0.289**	0.361

表 2.8　d-prime の実測値と予測値の相関係数

			d-prime （実測値）
d-prime （予測値）		ピアソンの相関係数	.750
		p値	.000

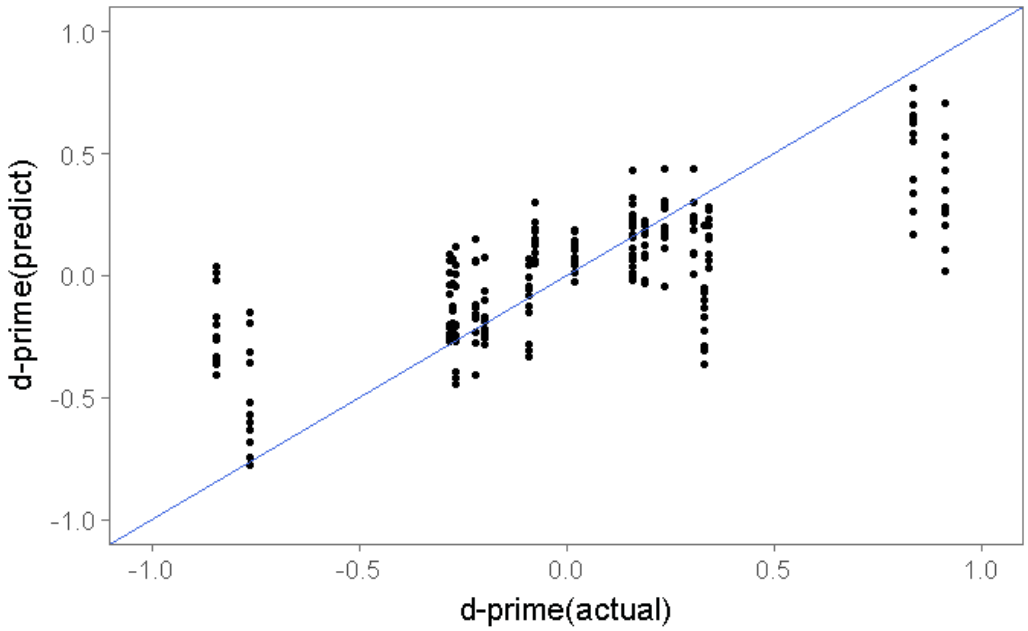


図 2.11　d-prime の実測値と予測値（SVR の結果）の散布図

2.5.3.3　d-prime 正負分類モデル

視線情報を用いて欠陥の種類を分類できるかどうかも重要である．そして，欠陥の種類は d-prime の正負の違いであると想定される．そこで，抽出した 15 個の特徴量を用いて，d-prime 正負分類モデルを複数のアルゴリズムで構築する．本研究で採用したアルゴリズムは，前節で d-prime の予測精度が高かった RF と Support vector machine（以降，SVM）である．前節のモデルベース分析とは異なり，d-prime の値ではなくその正負をクラスラベルとして用いた

各アルゴリズムによる欠陥分類モデルの有効性の確認は全データでランダムに 10-fold cross validation を行い，Accuracy を算出して比較する．Accuracy は予測値と実測値との一致率であり，次式

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.8}$$

で定義される．ここで，TP，TN，FP，FN は表 2.9 に示す混合行列の各要素である．

表 2.9　混合行列

		真の結果	
		正	負
予測結果	正	*TP*	*FP*
	負	*FN*	*TN*

d-prime 正負分類モデルを構築し，アルゴリズムごとに得られた結果を表 2.10 に示す．この結果から，瞬目率と他の視線情報を用いることで，約 84%の accuracy で d-prime の正負を分類できることがわかった．

表 2.10　d-prime の正負分類モデルの accuracy

	RF	SVM
accuracy	83.83%	**84.38%**

2.5.4　分析 4: 意図しない欠陥を考慮した閲読品質と瞬目率の関係

分析 2 にて，元の要件定義書に本研究で混入させた欠陥以外に意図しない潜在的な欠陥が含まれていたことが示唆された．そこで，被験者が任意で下線箇所に指摘事項などの情報を記載していたため，それらの情報を基にノイズに含まれる意図しない欠陥を分類した．分類した欠陥を基に，意図しない欠陥を考慮した閲読品質（d-prime）を算出し，分析 1，2 と同様の分析を実施した．

2.5.4.1　意図しない欠陥の分類

被験者が下線箇所に記載した指摘事項などの情報を分析することで（被験者がノイズに対して指摘した欠陥分類は付録 A.3 を参照），ノイズに含まれる欠陥を分類した（表 2.11）．ただし，一つのノイズ（文）に対して複数種類の欠陥を指摘していた被験者がいたため，それぞれの欠陥ごとに 1 件としてカウントした．そのため，付録 A.3 と表 2.11 の件数は，図 2.9 で示したノイズへの反応数以上となる．

その結果，一部は標準化や文の校正といった軽微な欠陥の指摘もあったが，半数弱ははじめにで述べた抜け漏れまたは誤りに分類される重要な欠陥の指摘であった．さらに，d-prime が近い値の被験者だとしても，ノイズに対して指摘した欠陥分類ごとの件数が異なることがわかった．そのため，抜け漏れと誤りに分類される欠陥を検出している被験者の閲読は，その品質が高い可能性がある．被験者がノイズに対して抜け漏れや誤りを検出している場合，被験者にとってその文はノイズではなくシグナルである可能性がある．

そこで，被験者がノイズに対して抜け漏れまたは誤りを一つ以上検出してい

る場合は,ノイズではなくシグナルに反応したと仮定して d-prime を算出した.

表 2.11　ノイズに含まれる欠陥分類と件数（抜け漏れ：定義すべき要件の抜け，曖昧：二つ以上の解釈が可能，誤ったセクション：記載すべきセクションの誤り，標準化：用語が未定義または用語が未統一，校正：文章表現の修正や誤字脱字，確認／提案：要件に対する確認または要件の修正提案，冗長：複数の箇所にて同じ要件の繰返し）

欠陥分類		件数
抜け漏れ	−	65
誤り	曖昧	60
	誤ったセクション	11
その他	標準化（用語の統一など）	24
	校正（表現の修正，誤字脱字）	55
	確認／提案	34
	冗長	7
	不明	46

2.5.4.2　経験と意図しない欠陥を考慮した閲読品質の関係

分析 1 と同様に d-prime と閲読経験の関係を分析した（図 2.12）．2 種類の閲読経験と d-prime の間には有意な相関はみられなかった（要件定義閲読経験数と d-prime のピアソンの相関係数 $r = -.168, p = .493$，文書の閲読頻度と d-prime のスピアマンの相関係数 $r = -.142, p = .563$）．したがって，閲読経験は抜け漏れと誤りを含めた閲読品質には影響を及ぼさないことがわかった．これははじめにで述べた閲読経験の有無が品質に大きな影響を与えなかったという既存研究と同じ結果であり，閲読経験は品質を評価する上では不十分であることが示唆された.

これに加えて，意図しなかった欠陥を欠陥とした場合，d-prime がすべてゼロ以上になることが示された(図 2.12)．この結果は，分析 2 の図 8 に示す正負の d-prime が混在する結果に対して，意図しなかった欠陥を検出した結果として負の d-prime を示す閲読者がいたという解釈と整合的である．つまり，新たに意図しなかった欠陥を含めて欠陥とする基準及びそれに基づく d-prime は，より適切に閲読者の欠陥検出の能力を捉えると考えられる．したがって，次に改めてこの d-prime と閲読時の瞬目率の関係を分析する.

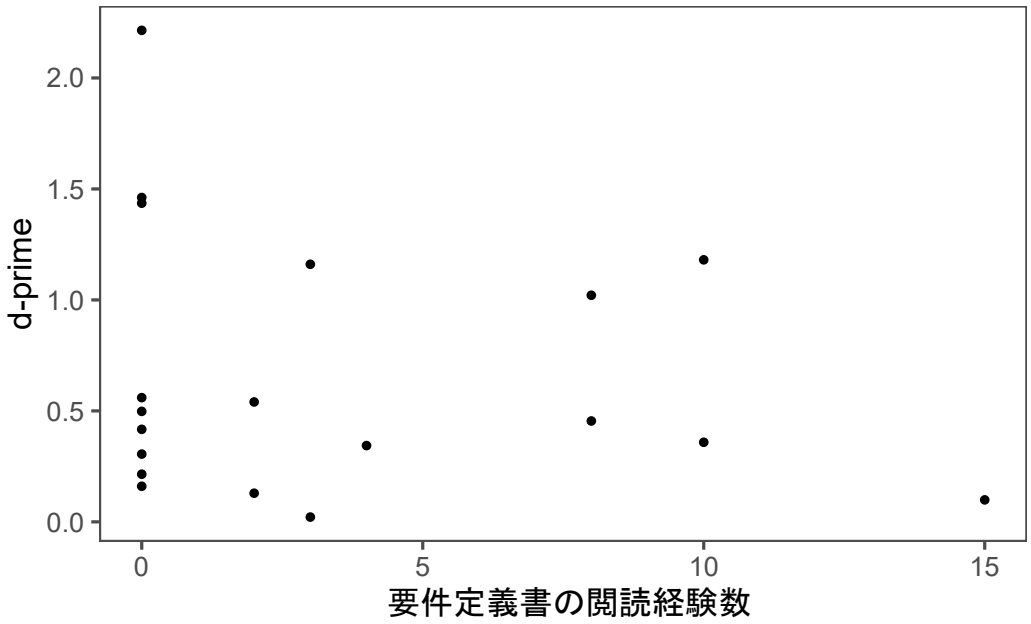


図 2.12　要件定義書の閲読経験数と意図しない欠陥を考慮した d-prime の散布図

2.5.4.3　瞬目率と意図しない欠陥を考慮した閲読品質の関係

分析 2 と同様に d-prime と瞬目率の関係を分析した（図 2.13, 表 2.12）．その結果，d-prime と瞬目率は有意な正の相関がみられた．分析 2 と同様に，この結果は本研究の仮説を支持する結果であり，瞬目率は抜け漏れと誤りも含めた閲読品質を予測する上で重要な特徴であることが示唆された．

合わせて，瞬目率と F 値の散布図を図 2.14 に示す．d-prime 同様に有意な正の相関がみられた（ピアソンの相関係数 $r = .571, p < .00001$）．さらに，F 値と d-prime の間にも有意な正の強い相関がみられた（ピアソンの相関係数 $r = .802, p < .00001$）．したがって，d-prime が正の場合においては，F 値は d-prime をほぼ同じ意味を持つ指標であると考えられる．

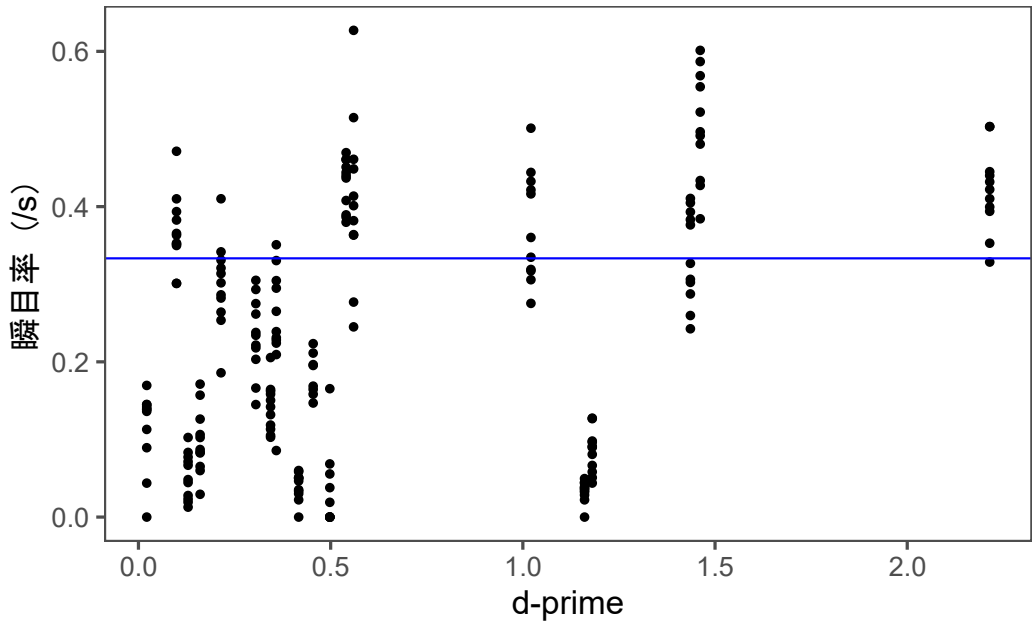


図 2.13　意図しない欠陥を考慮した d-prime と瞬目率の散布図（青線は瞬目率の平均値を示す）

表 2.12　意図しない欠陥を考慮した d-prime と瞬目率の相関係数

		瞬目率
d-prime	ピアソンの相関係数	.380
	p値	.000

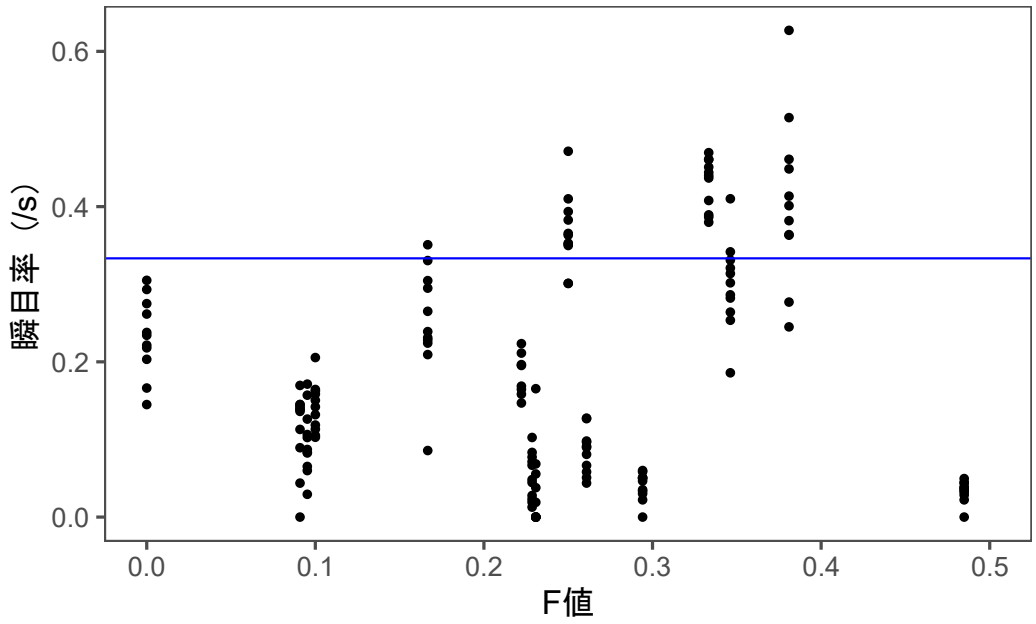


図 2.14　意図しない欠陥を考慮した F 値と瞬目率の散布図（青線は瞬目率の平均値を示す）

2.6　考察

2.6.1　瞬目率と閲読品質

SE 分野での過去の研究では，視線情報は認知プロセスの解明に用いられていたが，多くの研究は fixation と saccade に着目しており瞬目率はほとんど考慮されていなかった(Sharafi et al., 2015)．そして，その傾向は物語の読みや理解に関する研究において同様であった(Augereau et al., 2016; Campbell & Maglio, 2001; Okoso et al., 2015)．本研究では，閲読のサブプロセスに関連する瞬目率に着目し，閲読品質との関係を分析した．

分析 2 では，d-prime と瞬目率の間に，U 字型の非線形な相関を示した．この結果は d-prime で定義された閲読品質と瞬目率で計測された内的注意が関連するという仮説と一致した．瞬目率はストレスや緊張といった状態によって影響を受けることは報告されているものの，課題によって増加する場合もあれば減少する場合もある(田多他, 1991)．本課題は要件定義書を読むプロセスでは視覚課題に近いと考えられるものの，構造の理解と欠陥の検出／修正のプロセスでは視覚課題とは異なる可能性がある．そのため，どのプロセスが優位なのかによって取り組む課題が異なっており，瞬目率の増加がストレスや緊張によって引き起こされているかを判別することは難しい．合わせて，課題に対する取組姿勢（集中していたか否かなど）によっても瞬目率が影響を受ける可能性はあるものの，どのプロセスに集中したのかによって瞬目率に与える影響は異なる．例えば，読みに集中している人であれば瞬目率は減少傾向にある．した

がって，瞬目率の増減は閲読の三つのサブプロセスにおいて読み（外的注意が活性化）のプロセスにとどまるのか，構造の理解と欠陥の検出／修正（内的注意が活性化）のプロセスに至っているのかの違いによるものであると考えられる．分析 3 では，他の視線情報と比較した場合に，その相関が疑似相関である可能性を調査した．fixation，saccade，瞬目，瞳孔径から抽出した 46 個の他の視線情報を用いて回帰分析を行った．この結果，d-prime への影響度は瞬目率が最も高いことがわかった．さらに，瞬目率が閲読品質を予測する上で最も主要な特徴であることを示唆している．

分析 4 では，本研究で意図していなかった欠陥を考慮した d-prime を算出したところ，瞬目率との間に有意な正の相関を示した．分析 2，3 で算出した d-prime は本研究で定義した欠陥（抜け漏れ）に対する閲読品質を，分析 4 で算出した d-prime は潜在的に含まれていた 2 種類の欠陥（抜け漏れと誤り）に対する閲読品質を示している．したがって，瞬目率は，抜け漏れだけではなく誤りも含めた閲読品質を予測できる可能性が示唆された．

d-prime と比較する目的で分析 2 と 4 では F 値と瞬目率の関係を分析し，d-prime が正の場合にて F 値は d-prime と同じように欠陥の検出感度を示していることが示唆された．しかし， F 値は欠陥を検出したか否かを中心とした指標であるため，意図しない潜在的な欠陥が存在した場合それらに対する感度は測れない．

2.6.2 本実験の限界

分析 2 の結果で着目すべきは，閲読セッションでの教示の曖昧性に起因し，異なる種類の欠陥を検出している二つの潜在的なグループ（正の d-prime と負の d-prime）の存在が示唆されたことである．この実験の限界を考慮すると，両方の種類の欠陥（本研究で定義した欠陥とそれ以外の欠陥）を検出している優れた少数の閲読者が存在した可能性がある．なぜなら，そのような閲読者は本研究で定義したシグナルとノイズの両方を検出しており，d-prime がゼロに近い値となると想定されるからである．そこで，分析 4 ではそのような優れた閲読者の存在を考慮し，それ以外の欠陥のうち重要な欠陥を検出できている場合にシグナルを検出したとして d-prime を算出した．その結果，閲読経験と d-prime の間に有意な相関はみられず，経験数が多い閲読者にて 1 以上の d-prime を示す人が存在した．分析 2 ではゼロ付近の d-prime かつ，分析 4 では 1 以上の d-prime を示す閲読者が，優れた閲読者である可能性が高い．今後の実験デザインでは，複数の種類の欠陥が含まれることを防ぐため，要件定義書に含まれる多くの欠陥を網羅するリストが必要である．

瞬目率だけでは欠陥の種類は判別できないという課題があるが，分析 3 より他の視線情報も用いることで事前に定義した欠陥かどうかは分類できることがわかった．15 個の特徴量にて saccade の持続時間に関する特徴量が多かった．

一般的に saccade は視線の移動した軌跡を示し，その持続時間は軌跡の時間を示す．したがって，検出する欠陥の種類によって読み方や読む速度の強弱などの読みのスタイルが異なる可能性を示唆している．

本研究で使用した要件定義書は，実際に顧客のシステム開発に使用された文書であり，一定レベルの品質は担保されていたと考えられる．しかし，不完全な要件定義書には意図的に加えた欠陥の種類とは異なる潜在的な欠陥が含まれていた．複数の種類の欠陥が含まれていたことが閲読時の視線に影響がある可能性は否定できない．そのため，欠陥の種類を統制し，対象となる欠陥のみの検出力と視線情報の関係を調査するような実験計画の再検討が必要である．

実験終了後の質問紙調査では,「頭を固定することで集中しづらかった」,「紙で閲読をしたい」,「他の項目と行き来して閲読をしたい」などの意見があり，実際の要件定義書閲読と一部乖離している要素がある．モニタ上の 1 頁の要件定義書閲読においては汎用性があると考えられるものの，紙の閲読にて他の頁との行き来がある場合には視線の傾向が異なる可能性がある．この点については今後の課題とし，メガネのように装着するアイトラッカを利用し，実際の要件定義書閲読の場面に近づけた実験を検討する．

第3章　実験 2：要件の必要十分性判別能力課題作成[2]

3 章では，要件定義書の閲読能力を測る課題を作成する実験を行う．要件定義書の閲読では，顧客要求を必要十分に満たす定義書になっているかの確認を突合により行っている．しかし，実際の開発における顧客要求は複雑かつ多岐に及んでいるため，顧客要求と要件定義書の単純な比較はできない．そこで，ソフトウェアを幾何学図形で象徴化することにより，顧客要求と要件定義書のペアを作成するゲームを考案した．その結果，幾何学図形の集合（顧客要求）と要件定義書のペアを得た．ペアにおいて，顧客要求を必要十分に満たす要件定義書だけではなく，満たさない定義書も得られた．さらに，考案したゲームは実際の要件定義書作成における特徴を持っており，要件定義工程を十分にもした実験であることが示唆された．

3.1　実験の目的

実験 1 では，実際のソフトウェア開発に使用された文書であったものの，実験者が事前に想定してなかった潜在的な欠陥が含まれており，閲読品質と視線に影響を与えた可能性があった．さらに，実際のソフトウェアに関する暗黙的な顧客要求は明確には知ることはできず，要件定義書の概要にて顧客要求の概略のみ知ることができた.そのため,正確な顧客要求と突合することは難しく，要件定義書そのものの整合性や必要な要求が抜けていないかという閲読が主であったと想定される．

序論でも述べたように，閲読とは顧客要求と要件定義書の突合であり，実際のソフトウェア開発においては顧客要求を基に要件定義書を閲読している．その際，顧客要求は明文化されておらず，断片的な情報（例えば，顧客の発言をまとめた議事録や業務フローなど）のみが存在し，閲読者はそれらを基に顧客要求を整理していると考えられる．つまり，閲読者は断片的な情報を基に顧客が求めるソフトウェアをイメージし，要件定義書がそれを必要十分に満たすか突合していると想定される（図 3.1）．ここで，顧客要求は顧客が求めるソフトウェアを示すため非言語的な対象であり，一方で要件定義書は言語的な対象である．したがって，非言語的な対象と言語的な対象を突合していると考えられる．

要件定義書の閲読には，顧客要求を正確に把握する能力と要件定義書が顧客要求を必要十分に満たすかを突合する能力の二つが求められる．本実験では，閲読能力として後者の突合能力を対象とする.正確な閲読能力を測るためには，顧客要求と要件定義書のペアが求められる．しかし，そのようなペアの課題は

[2] 本実験は研究業績 4 の日本認知科学会第 36 回大会抄録集(斉藤 & 日髙, 2019)，研究業績 3 の Proceedings of The 42nd Annual Meeting of the Cognitive Science Society (CogSci2020)(Saito & Hidaka, 2020)にて発表・掲載された論文としてまとめられた研究成果に該当する．

開発されておらず,実際のソフトウェア開発は複雑かつ多岐に及んでいるため,実験 1 のように実際の要件定義書は突合の認知過程を調べる課題としては適していない．そこで，本実験ではソフトウェアと要件定義書の関係に着目し，そのペアを作成する実験を考案する．

ソフトウェア開発の各工程における成果物は，顧客要求を満たすソフトウェアを異なる状態で記述したものである．つまり，それらの成果物は一つのソフトウェアを記述しているものの，記述の手段（自然言語やプログラムなど）やその粒度が異なっている．要件定義工程において，要件はソフトウェアの機能または特性であることから，ソフトウェアは多くの特性の集まりであり，要件定義書ではその特性を自然言語で記述したものである．

したがって，要件定義書はそれぞれの特性をもつ集合の「内包的定義」であり，各ソフトウェアはそれらの「外延」の一例であるといえる．内包／外延は集合論で用いる用語であり， ある集合の内包的定義は，その集合の要素が持つ性質を記述することによる定義で，集合の外延的定義は，その集合のすべての要素を列挙することで具体的に示す定義である．例えば，通販サイトのログイン機能であれば，その機能の満たすべき条件を言語的に記述した文章が内包的定義であり，ある機能（画面などを通してログインできること）を持つ特定のプログラム自体が外延であり，あるいはそうしたプログラムを提示することによる定義が外延的定義である．つまり，ソフトウェア開発とは，顧客の要求する性質を記述した要件定義書として与えられるソフトウェアの内包的定義を読み解き，外延であるソフトウェアを構築する工程である．

品質の良い要件定義書（内包的定義）は，それにより定義された外延的定義が顧客の求めるソフトウェア（外延的定義）と必要十分に一致する．なぜなら，要件定義書はソフトウェアが満たす特性の要件のリストを記述しており，必要かつ十分である必要があるためである(Sawyer & Kotonya, 2001)．しかし，ほとんどの場合，一度で品質の良い要件定義書を作成することは難しく，1.2 で述べた閲読によりその品質を高めていく．閲読の目的は欠陥を検出することであり，閲読者は要件定義書が顧客の求めるソフトウェアを必要十分に記述しているかを確認していると想定される．

閲読の二つ目のプロセスである構造の理解（図 1.4）においては，内包的／外延的定義を基に比較可能な状態に変換して，突合により必要十分か否か確認していると想定される．具体的な例として，内包的定義を基に具体的な特性を定義し，それらが外延的定義と一致するかどうかの突合をしている（図 3.2）．もちろん，外延的定義を基に内包的定義に変換して，突合していることも考えられる．

要件定義書と顧客が求めるソフトウェアの突合において，その内包的定義と外延的定義の間には 4 種類の関係がある（図 3.3）．ここで，**R（Requirement）**は要件定義書によって定義された特性の集合を，**P（Product）**は顧客の求める

特性の集合（ソフトウェア製品）を示す．4 種類のタイプにおいて，タイプ NS (Necessary and Sufficient)の要件定義書が品質高い，目指すべきゴールである．ここで, R と P は順序関係を満たす集合ではなく, 概念的な特性の集合とする．

タイプ N（機能不足）：集合 R は集合 P の必要条件（necessary）であるが，十分条件（sufficient）でない.顧客が要求した特性（R）を一部もつが，全ての要求された特性はもたないソフトウェア（P）

タイプ S（余分な機能）：集合 R は集合 P の十分条件であるが，必要条件でない.顧客が要求した特性（R）を全てもつが，要求していない特性までももつソフトウェア（P）

タイプ U（機能不足と余分な機能）：集合 R は集合 P の必要条件でも十分条件でもない.顧客が要求した特性（R）を一部しかもたない，かつ要求していない特性までもつソフトウェア（P）

タイプ NS（完全な機能）：集合 R は集合 P の必要十分条件である.顧客が要求した特性（R）のみをもち，要求していない特性はもたないソフトウェア（P）

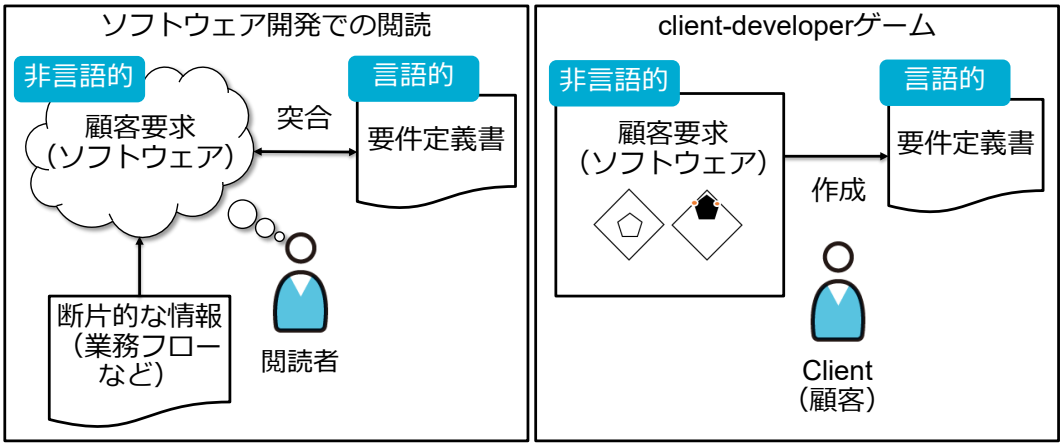


図 3.1　ソフトウェア開発での閲読と CD ゲームの関係

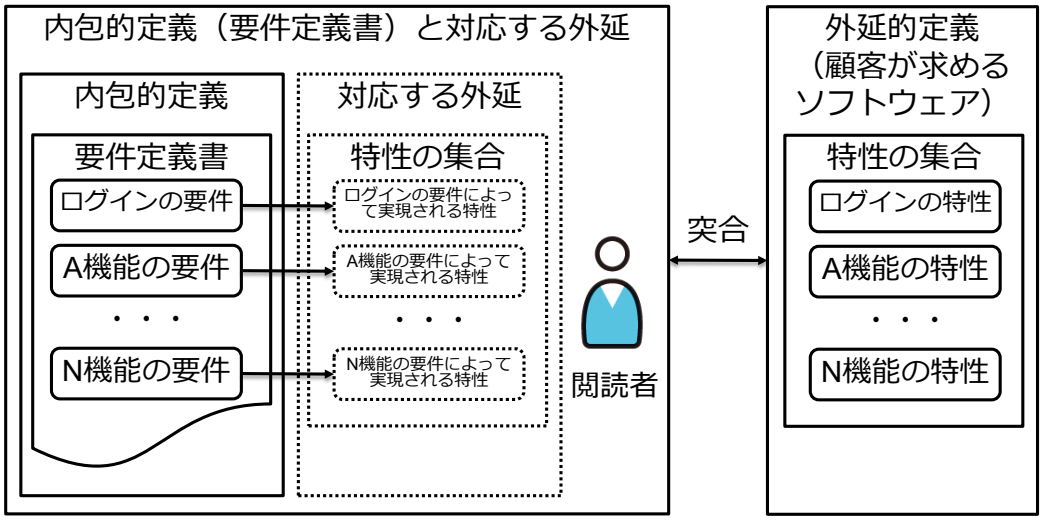


図 3.2　閲読者によるソフトウェアの内包的定義と外延的定義の突合の例

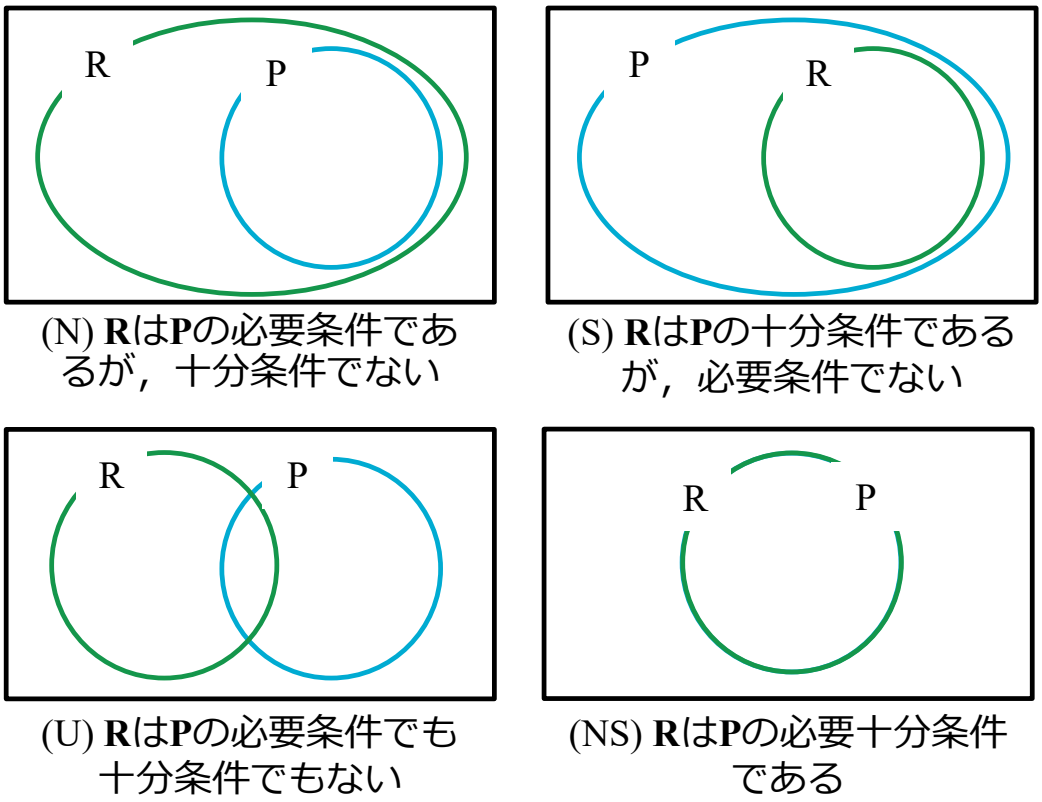


図 3.3　集合 R（要件定義書）と集合 P（ソフトウェアプロダクト）の間の 4 種類の関係

実際のソフトウェア開発では，要件定義書は自然言語で書かれており論理構造が非常に複雑でありかつそのソフトウェアの構造も複雑であるため，要件定義書とソフトウェアをこれらの四つの理想的なタイプに分類することは難しい．

そこで，仮説を検証するために，顧客と開発者のインタラクティブなコミュニケーションをシミュレーションするゲームを開発した（図 3.4）．これを本研究では client-developer ゲーム（CD ゲーム）と呼ぶ．CD ゲームと実際のソフトウェア開発の閲読の関係性について図 3.1 に整理した．非言語的な顧客要求を幾何学図形で象徴化することで，実際のソフトウェア開発における閲読の状況に近づけた設計とした．ただし，実験者が顧客要求を定義しているため顧客要求を正確に把握する能力は測れない．CD ゲームのスキームは，実験記号学における繰り返し学習パラダイム(Smith et al., 2003)に着想を得て設計した．このゲームでは，個々の幾何学的図形がソフトウェア特性の要素を表し，それらの要素からなる集合をソフトウェアとみなす．顧客は求めるソフトウェア（集合）を内包的定義によって表現する（要件定義書）．一方で，開発者は顧客の要件定義書を基に顧客が求める幾何学図形の集合（ソフトウェア製品）を外延的に構築する．顧客と開発者は，要件定義書の作成とソフトウェア製品の構築を繰り返し行う．このようなインタラクションを経ることで，最初は不必要あるいは不十分な要件定義書（タイプ U, N, S）が，顧客が求める製品を必要十分に満たす要件定義書（タイプ NS）へと近づいていくことが期待される．CD ゲームの結果として，不一致（タイプ N，S，U）から一致（タイプ NS）まで様々なタイプの要件定義書と製品のペアの系列を得るられる．その後，得られた要件定義書と製品のペアを実験刺激として用いて，その閲読プロセスの更なる実験を行う（4 章の実験 3）．

まとめると，本実験（CD ゲーム）の目的は，上記に示した方法にて象徴化された要件定義書とソフトウェア製品の対の集合を構築することである．この実験は，要件定義書が自然と洗練されて作成されるプロセスを模したもので，製品の要件を不正確に記述されたものから必要十分に記述されたものへの遷移を捉えている．実験 3 では作成された要件定義を用いて，閲読者の視線の振る舞いを調査する．

合わせて，CD ゲームにおけるラウンドの経過による要件定義書への影響を調査することで，CD ゲームにおける閲読の効果を検証できると考える．なぜなら，developer は要件定義書を基にソフトウェア製品を開発しており，client にとっては定義書の閲読のような役割を果たしている．そのため，CD ゲームにおいて 1 ラウンドごとに閲読が行われている状態であり，閲読によって要件定義書の品質の変化（タイプの変化）を分析することで閲読の効果が検証できる．閲読によって要件定義書の品質が高められるならば，ラウンドを経るごとにタイプ NS に近づいていくと想定される．

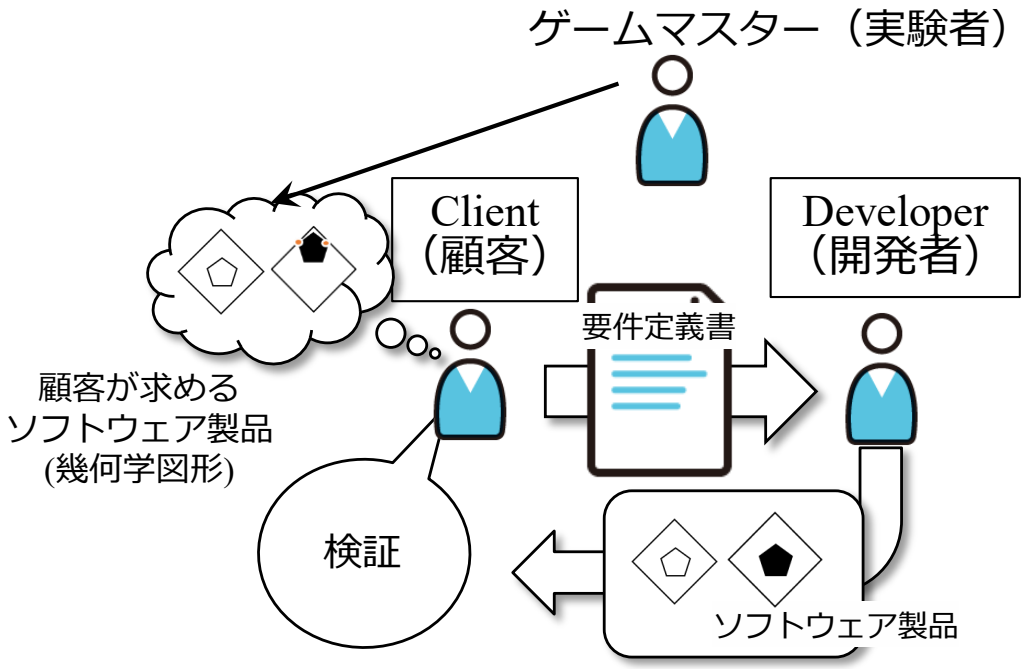


図 3.4　client-developer ゲームにおける全体概要図

3.2　実験手続き

先に述べたように，顧客要求文は求める製品の特性（内包的定義：求める図形の集合を説明する命題の集合）を記述する言語形式で与えられ，開発者が作成する製品は正しい幾何学図形（正例）の集合（外包的定義：図形の集合）である．以下では，正しいソフトウェア（集合）に含まれる幾何学図形を正例と呼び，それに含まれないものを負例と呼ぶ．

本実験では，顧客が求める製品を象徴化したものとして，図 3.5 で示すような幾何学的パターンの集合を採用した．求める製品は，幾何学図形の集合として表現し，それぞれの製品を 6 つの異なる図形で構成した．その製品を正確に記述する要件定義書(命題の集合)は 8 つまたは 9 つの文から成るように設計．五つの製品と要件定義書のペアを用いて，client-developer ゲームを実施した．

3.2.1　被験者

被験者は 10 人の日本人の成人（男性：8 名，女性：2 名）で，30 代が 4 名，40 代が 1 名，50 代が 4 名，60 代が 1 名であり，平均年齢は 45.3（SD = 9.4）であった．全て日本ユニシス株式会社のシステムエンジニアであった．

3.2.2　手続き

10 人の被験者のそれぞれを顧客または開発者にランダムに割り当て，5 人の顧客と 5 人の開発者に分けた．それぞれの顧客は 2 人の開発者とペアになり，

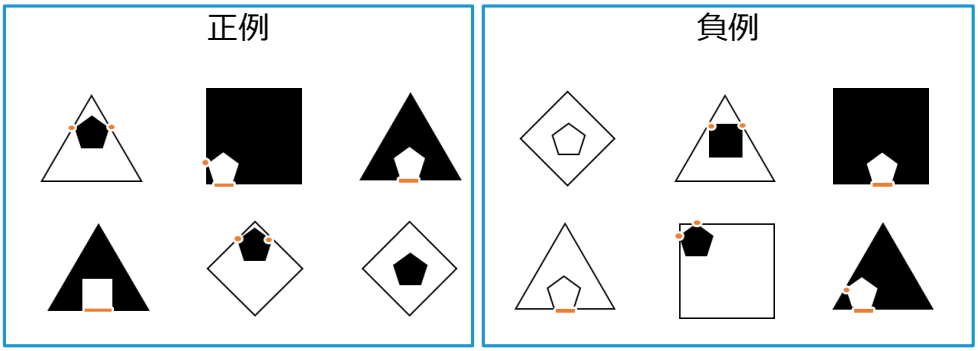
特定の図形の集合を基にゲームを実施した．それぞれの開発者は 2 人の異なる顧客とペアとなり，二つの異なる製品（開発者にはその製品については知らされていない）を使用したゲームに取り組み，顧客から与えられた要件を基に製品を作成した．したがって，一つのゲームは 1 人の顧客と 2 人の開発者で構成され，2 人の開発者が作成した製品を結合したものが，各ゲームにて顧客にフィードバックされた．開発者がかならずしも求められたすべての幾何学図形を製品として含めることができるわけではないため，2 人の開発者を割りあてた．五つのゲームにおけるプレイヤーの組合せは，表 3.4 の左端の 2 列に示す．

ゲームの最初に，5 人の顧客ぞれぞれに正例と負例が提示された．正例は，顧客が求める図形を，負例は求めていない図形を示す．五つの異なる図形のセット（正例と負例）を用いて，それぞれの顧客は異なる図形のセットを提示された．正例と負例の一例を図 3.5 に示す．その後，顧客は正例のみを表現する文のリストを作成するように指示された．文のリストの詳細な作成ルールは，次節にて説明する．顧客がリストを作成した後，そのリストは対応する開発者に渡される．開発者はルールに沿って，幾何学図形作成するように指示された．その後，2 人の開発者によって作成された図形群を結合したものが，対応する顧客へと渡された．ここまでを一つのラウンドとする．開発者によって作成された図形群が顧客に提示された正例が一致した場合（図 3.5 (a)における正例の正しい要求文のリストを図 3.5(b)に示す），ゲームはこのラウンドで終了する．一致しない場合は，最大 4 ラウンドまで繰り返す．顧客と開発者の両方は，ゲーム開始前に本ゲームの目的とルールを事前に説明された．

最初のラウンドでは，実験になれるために二つの例題が顧客と開発者にそれぞれ提示された．二つの例題は，本ゲームのルール確認を目的として実施された．そのため，顧客と開発者のペアによるラウンドは行わなかった．顧客は提示された正例と負例を基に文のリストを作成し，その後正答例が提示された．開発者は文のリストを基に幾何学図形を作成し，その後正答例が提示された．

実験終了後，質問紙調査を実施し，またインタビューを行い，個人特性を収集した．

(a)



(b)

顧客要求文

1. 大は正三角形または四角形である
2. 小は正四角形または正五角形である
3. 大の辺と小の辺が1辺のみ接する場合、大は黒塗りである
4. 大が正四角形の場合、大の辺と小の点は3点のみ接する
5. 大が正三角形の場合、大の辺と小の点は2点のみ接する
6. 大の辺と小の点は接しないまたは2点以上接する
7. 大の辺と小の辺が接しない場合、小は黒塗りである
8. 小が黒塗りの場合、小は正五角形である

図 3.5 （a，左）正例，（a，右）負例，（b）顧客要求文

3.2.3 課題難易度の調整

異なる製品を用いたゲームの難易度を揃えるために，正例は 6 つとし，それを表現するのに必要な文数は 8 つまたは 9 つとなるように設定した（図 3.5）．

3.2.3.1 正例の作成規定

自由に図形を作図して事例を作成すると，要件定義書の曖昧を避けるために文が過剰に長くなるか，文の例外が生じる．従って，実験で用いる図形および文の種類に制限をいれ，同値類（図 3.7）を定義することで，こうした問題を解消した．具体的には，単位図形を 20 種類定義し，製品（幾何学図形）の作成に使用した（図 3.6）．各 1 回まで一つの組み合わせに使えるとし，二つを組み合わせて図形を作成した（6 組の単位図形の組合せ集合で一つの製品を作成した）．正例，負例の両方において，図形を生成できる組合せ数を減らすために，単位図形の操作規定を以下のように設定した．

① 全てに図形は，二つの単位図形（大と小）で構成され，小の単位図形（以降，小）は大の単位図形（以降，大）に包含される

② 単位図形の可能な操作は，平行移動のみである（回転，拡大縮小，反転は不可）

③ 大と小の色は，ともに黒でない限り，黒または白のどちらでもよい

④ 大と小を関連づける方法は以下の 9 通りのみである：表 1 の(a)–(i)，および図 3.7 に例を示す．大と小が接し方の違いのみが関連づけのパターンとして考慮され，接する辺や点などの位置は考慮しない．

上記制限規定に基づき，作図可能な組み合わせ図形の個数を表 3.1 に示す．

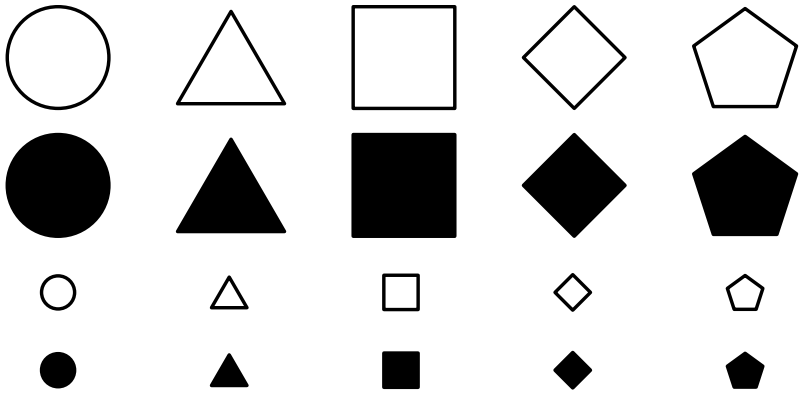


図 3.6　20 種類の単位図形

表 3.1　作成可能な図形の組み合わせ数

点と点	辺と点	辺と辺	組合せ	個数
0点で接する（接しない）	0点で接する（接しない）	0辺で接する（接しない）	(a)	75
	1点のみで接する	0辺で接する（接しない）	(b)	48
	2点のみで接する	0辺で接する（接しない）	(c)	48
		1辺のみで接する	(d)	30
	3点のみで接する	1辺のみで接する	(e)	9
1点のみで接する	0辺で接する（接しない）	0辺で接する（接しない）	(f)	24
	1点のみで接する	1辺のみで接する	(g)	9
	2点のみで接する	2辺のみで接する	(h)	12
2点のみで接する	0点で接する（接しない）	0辺で接する（接しない）	(i)	12
			合計	267

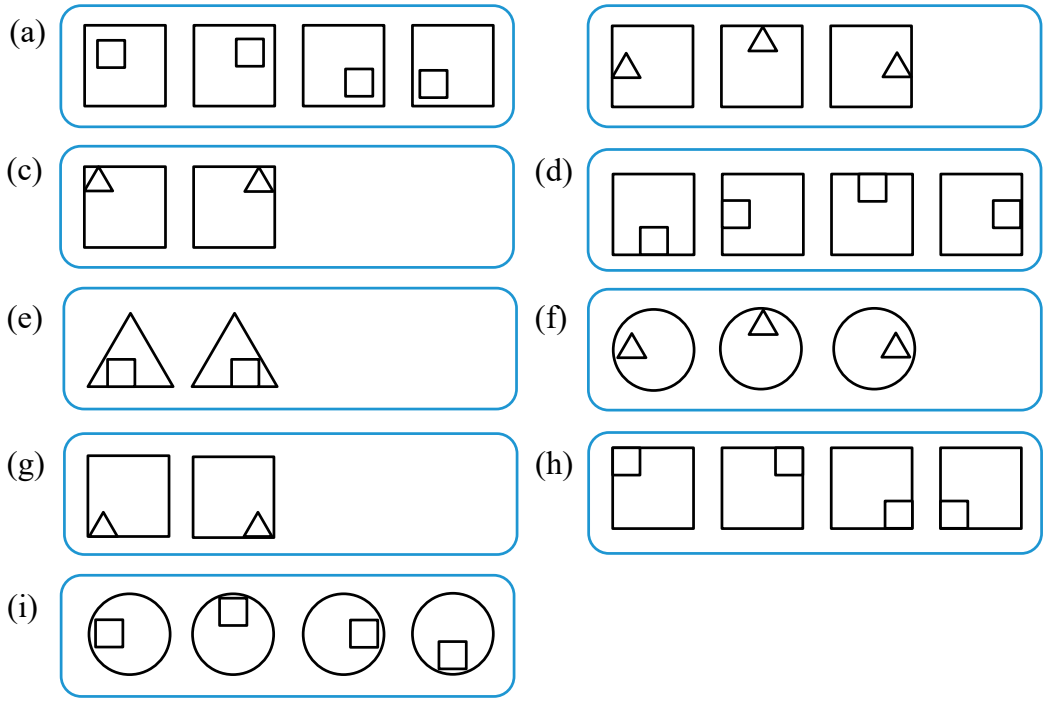


図 3.7　同値な図形の例（青枠内の図形が同値グループ）

3.2.3.2　顧客要求文の言語的記述規定

顧客要求文の言語的な記述にて，被験者間で表現のばらつきを抑えるため，顧客要求で用いることができる用語と記述規則を以下のように規定した．

顧客要求を記述する上で使用可能な用語

- 形：円，正三角形，正四角形，ひし形，正五形
- 四角形：正四角形とひし形
- 正多角形：正三角形，正四角形，正五角形
- 多角形：正三角形，正四角形，ひし形，正五角形
- 大きさ：大，小
- 塗り：白塗り，黒塗り
- 辺／頂点の数：辺，頂点共に円はゼロ，正三角形は 3，正四角形とひし形は 4，正五角形は 5 とする
- 接する：単位図形同士の枠線が重なっている場合は，接するとし，前節の正例の作成規定に基づき 9 通りで表現される

顧客要求を記述する上で使用可能な構文

- A は B である／でない
- A が B である／でない場合，C は D である／でない

A，B，C，D は 1 種類の要求のみ記述でき，同種類の要求であれば「または」で複数記述できるとした．上記制限規定に基づき，A は B である構文の組合せ数を表 3.2 に示す．A は B でない構文においても同数の文が作成でき，かつ C は D である／でない構文においても同じ文を用いることができる．したがって，A は B である／でない構文で 160 個の文が，160 個から二つを選ぶ組合せでは 12720 個の文が作成できるため，計 12880 個の文が作成できる．ただし，文の組合せ次第では A が B である／でないと C は D である／でないが矛盾する場合も含まれているため，現実的に作文可能な文数は 12880 個より少なくなる．

表 3.2　A は B である構文において作文可能な組み合わせ数

	A	B	個数
形に関する条件	大	5つの形から1つ選択	5
		5つの形から2つをまたはでつなげる	10
		5つの形から3つをまたはでつなげる	10
		5つの形から4つをまたはでつなげる	5
		5つの形から5つをまたはでつなげる	1
	小	5つの形から1つ選択	5
		5つの形から2つをまたはでつなげる	10
		5つの形から3つをまたはでつなげる	10
		5つの形から4つをまたはでつなげる	5
		5つの形から5つをまたはでつなげる	1
	大と小の形	同じ	1
		異なる	1
塗りに関する条件	大	白塗り	1
		黒塗り	1
	小	白塗り	1
		黒塗り	1
	大と小の塗り	同じ	1
		異なる	1
接する条件	大の点と小の点	0点で接する（接しない）	1
		1点のみで接する	1
		2点のみで接する	1
	大の辺と小の点	0点で接する（接しない）	1
		1点のみで接する	1
		2点のみで接する	1
		3点のみで接する	1
	大の辺と小の辺	0辺で接する（接しない）	1
		1辺のみで接する	1
		2辺のみで接する	1
		合計	80

3.3　結果

3.3.1　Client グループ

Client グループの 5 人のうち，2 人は 2 ラウンド目で，1 人は 3 ラウンド目で，1 人は 4 ラウンド目で終了した．1 人は 4 ラウンド目でも終了しなかった

ため，4 ラウンドで終了とした．

被験者が定義した顧客要求を満たす事例の集合 R と正例 P との関係のラウンドごとの推移を表 3.3 に示す．ラウンド 1 では，client グループ 3 と 5 にて R が必要性を満たさないタイプ（タイプ S と U）があったものの，まずは集合 R を正例 P の必要条件を満たすよう（タイプ N）に記述する傾向があった．その後，developer グループの作成事例を基に，十分性を満たすよう（タイプ NS）に記述を修正していた．タイプ NS の前は必ずタイプ N になっており，developer グループの検証なしにタイプ NS の顧客要求文を作れた client は存在しなかった．合わせて，タイプ N からタイプ S や U に変化するパターンも存在しなかった．

ラウンドごとの顧客要求文の数推移を図 3.8 に示す．ラウンドの進行とともに顧客要求の文数は，一部減少することがあったものの基本的には変化しないまたは増加する傾向にあった．N から NS に推移する場合においては，顧客要求文の数は減少することなく，変化しないまたは増加していた．Client グループ 2 における，文章の変化の例を図 3.9 に示す．

したがって，client グループでは正例を満たすために必要な顧客要求を列挙し，その後十分性を満たすように要求を追加または修正することで必要十分性を満たす要求を作成する傾向にあることがわかった．

表 3.3　集合 R と P の関係の推移（タイプ N: 集合 R は集合 P の必要条件（necessary）であるが十分条件（sufficient）でない，タイプ S: 集合 R は集合 P の十分条件であるが必要条件でない，タイプ U: 集合 R は集合 P の必要条件でも十分条件でもない，タイプ NS: 集合 R は集合 P の必要十分条件である）

Client グループ	集合RとPの関係			
	ラウンド1	ラウンド2	ラウンド3	ラウンド4
1	N	NS	-	-
2	N	NS	-	-
3	U	N	N	NS
4	N	N	N	N
5	S	N	NS	-

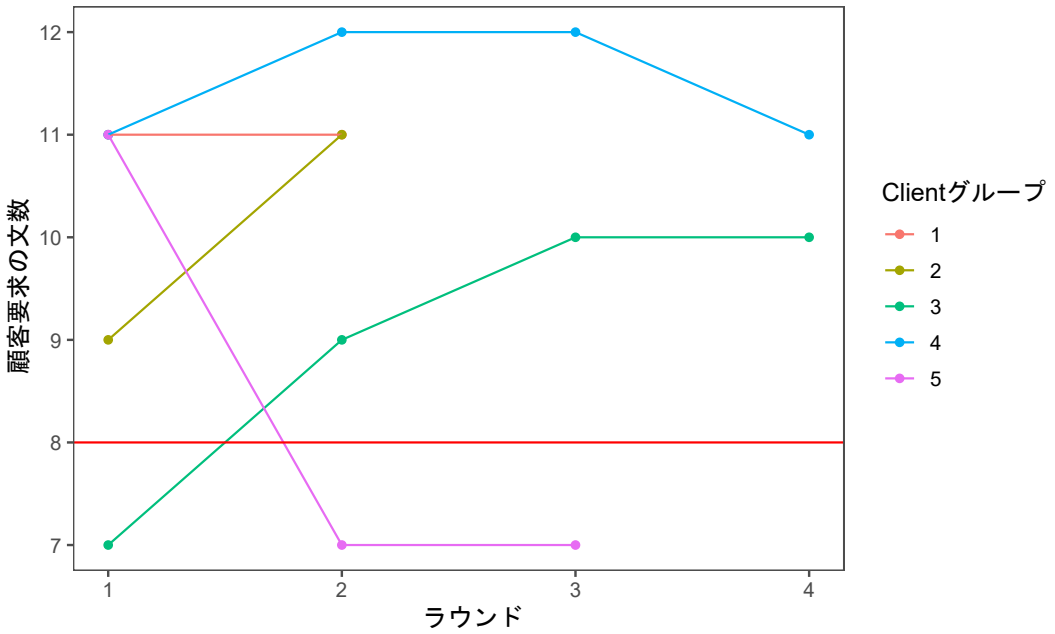


図 3.8　各被験者のラウンドごとの顧客要求文の数推移（赤線は実験者が事前に設定した必要な顧客要求の文数）

(a)ラウンド1

顧客要求文
1. 大は円または正五角形またはひし形である
2. 大が正五角形またはひし形の場合、白塗りである
3. 小は正三角形またはひし形である
4. 大が円または正五角形の場合、小は正三角形である
5. 大の点と小の点は１点以上接する
6. 大と小がひし形の場合、小は黒塗りである
7. 大が円の場合、小は白塗りである
8. 大が黒塗りの場合、大の点と小の点は2点のみ接する
9. 小が白塗りの場合、大の辺と小の辺は接しない

(b)ラウンド2

顧客要求文
1. 大は円または正五角形またはひし形である
2. 大が正五角形またはひし形の場合、大は白塗りである
3. 小は正三角形またはひし形である
4. 大が円または正五角形の場合、小は正三角形である
5. 大が円の場合、小は白塗りである
6. 大と小の形が同じ場合、小は黒塗りである
7. 大の点と小の点は１点以上接する
8. 大が黒塗りの場合、大の点と小の点は2点のみ接する
9. 大と小の塗りが同じ場合、大の点と小の点は１点のみ接する
10. 小が白塗りの場合、大の辺と小の辺は接しない
11. 小が黒塗りの場合、大の辺と小の辺は１辺以上接する。

図 3.9　client グループ 2 における顧客要求文のラウンドによる変化例

3.3.2　Developer グループ

Developer グループの作図した図形事例にて，要求定義グループの正例との類似度を示す Jaccard 係数(Jaccard, 1912)を表 3.4 に示す．Jaccard 係数は二つの集合間の類似度を測る指標であり，次式，

$$\mathrm{J}(R,P) = \frac{|R \cap P|}{|R \cup P|} \tag{3.1}$$

で定義される．ここで，Rは client グループが作成した顧客要求文を基に作成できる図形の集合であり，Pは developer グループが作図した図形の集合であ

る．

Rはタイプ NS の場合であれば 6 個の正例と一致し，タイプ N であれば 6 個より多くなる．本実験において Jaccard 係数は，developer グループが顧客要求文を基に図形を作成する能力を示しており，1 に近いほど要求文から作成できる図形を必要十分に作成できているといえる．なぜならば，各ラウンドにおける集合Rは顧客要求文ごとに一定であり，developer グループの作図した図形の集合によってその値が増減するからである．したがって，Jaccard 係数が 1 の時は，$|R \cap P| = |R \cup P|$となり，$R = P$が成り立つため，Jaccard 係数は必要十分性を表す指標である．Jaccard 係数は，多くの被験者でラウンドが進むにつれて増加傾向にあり，ラウンドが進むほど client グループが作成した要求を必要十分に満たす事例を作成できるようになる傾向がある．しかし，Jaccard 係数が次ラウンドで著しく減少する場合もある．そのため，最初のラウンドと最後のラウンドの Jaccard 係数を基に t 検定を実施した結果，有意差は見られなかった．本実験では，開発者グループは誤って負例を正例として作図しても減点を受けることなく，かつ顧客グループにはそのような事例は提示しなかった．そのため，ラウンドが終了しないことを危惧して，網羅性を満たすために多くの事例を作図したことによって Jaccard 係数が減少したと考えられる．

正例を網羅しているかどうかを示すカバー率を表 3.5 に示す．カバー率は，developer グループが作図した図形集合において，client グループが作成した顧客要求を基に作成できる図形集合の網羅率を図る指標であり，次式，

$$\mathrm{C}(R,P) = \frac{|R \cap P|}{|R|} \tag{3.2}$$

で定義される．Jaccard 係数は，顧客要求を基に作成できる図形集合を正確に作成する能力であったものの，誤った図形を作成したとしても減点にならなかったため，網羅性を高めるために誤った余分な図形を多く作成する被験者も存在した．その場合は，Jaccard 係数は小さくなる．Jaccard 係数が小さい場合に，誤った図形を含むのかそれとも必要な図形を作成できていないのかが判別できない．そこで，網羅率を示すカバー率を導入し，必要な図形を作成できているのかを判別する．カバー率が 1 に近いほど，顧客要求を基に作成できる図形集合を漏れなく作成できている．そのため，カバー率が 1 の場合は，$|R \cap P| = |R|$となり，$R \subseteq P$が成り立つため，カバー率は十分性を表す指標である．良い developer は，Jaccard 係数とカバー率がともに高い値を示す．developer グループの個人のカバー率にはばらつきがあり，正例を網羅できていない場合も多く，顧客要求の検証には一人では不十分である．しかし，2 人を合わせた合算でのカバー率は，すべてのラウンドで 86%以上であり，顧客要求の検証には 2 人で十分であることがわかった．合わせて，最初のラウンドと最後のラウンドのカバー率を基に t 検定を行ったところ，有意差が見られた（$\boldsymbol{p < .05}$）．このことから，閲読を繰返すことで，developer 自体の閲読品質と顧客要求文の品質

が高められていると考えられる.

表 3.4　Developer グループ作図事例の正例類似度

Cliend グループ	Developer グループ	Jaccard係数			
		ラウンド1	ラウンド2	ラウンド3	ラウンド4
1	10	1.00	1.00	−	−
	9	0.88	0.86	−	−
2	9	0.60	1.00	−	−
	8	0.39	0.55	−	−
3	8	0.61	1.00	0.43	1.00
	7	0.75	0.36	0.75	0.63
4	7	0.18	0.22	0.44	0.71
	6	0.78	0.50	1.00	1.00
5	6	0.75	0.94	0.35	−
	10	0.67	0.94	0.83	−

表 3.5　Developer グループの作図事例の正例カバー率

Client グループ	Developer グループ	ラウンド1 カバー率		ラウンド2 カバー率		ラウンド3 カバー率		ラウンド4 カバー率	
		個人	合算	個人	合算	個人	合算	個人	合算
1	10	100%	100%	100%	100%	−	−	−	−
	9	88%		100%		−		−	
2	9	67%	89%	100%	100%	−	−	−	−
	8	78%		100%		−		−	
3	8	79%	86%	100%	100%	100%	100%	100%	100%
	7	86%		71%		100%		83%	
4	7	38%	88%	29%	86%	57%	100%	71%	100%
	6	88%		86%		100%		100%	
5	6	100%	100%	100%	100%	100%	100%	−	−
	10	67%		94%		83%		−	

3.4　考察

Client グループにおいて，developer グループの検証なしにラウンド 1 にてタイプ NS の顧客要求文を作成できた被験者はいなかった．このことから，図形事例を基に要件定義書を作成するという比較的難易度の易しい課題だとしても，閲読を全く行わずに顧客要求を満たす定義書を作成することは困難であることが示唆された．合わせて，developer グループの検証によって，ほとんど client グループの被験者がタイプ NS の要件定義書が作成できた．カバー率は

最終ラウンドにて最初のラウンドと比較して，有意に高くなっており，閲読を繰返すことで要件定義書の品質が高められていることが示唆された．この実験結果は,少なくとも象徴化された製品の品質に及ぼす閲読の効果を示しており，要件定義書においては閲読が効果的であると示された．合わせて，Jaccard 係数は必要十分性の指標を，カバー率は十分性の指標を示していることが示唆された．

さらに，developer グループにおいて個人のカバー率が低い被験者が見られたため，単独ではなく複数人の閲読が効果的である．2 人合算のカバー率は最小でも 86%であるものの，個人の正例カバー率の最小は 29%であり，一人での閲読には限界がある．

したがって，要求の必要十分性を満たすためには，複数回のラウンドが必要であり，かつ複数人の観点での検証が必要であることが示された．これは，実際の業務での要求定義書作成においても同様であり，複数人による複数回の閲読が必要とされる．したがって，本実験は要求定義工程における要求定義書作成／閲読のプロセスを十分に模した実験であると考えられる．そのため，本実験での client／developer グループは，それぞれ要求定義工程での，要求定義の作成者／閲読者のような役割を果たしている．

第4章　実験 3：要件の必要十分性判別能力と視線の関係[3]

4 章では，3 章で作成した顧客要求と要件定義書のペアを用いて閲読能力を測る実験を行い，閲読能力と視線の関係を議論する．被験者は，与えられた要件定義書が，幾何学図形であらわされた顧客要求を必要十分に満たすかを判別する課題を実施する．要件定義書が顧客要求を必要十分に満たすかを判別する能力を閲読能力として捉え，視線情報との関係を分析した．その結果，閲読能力の違いによって，瞳孔径と文ごとの fixation にて有意な特徴がみられた．

4.1　実験の目的

実験 2 では，幾何学図形（ソフトウェア製品）に対して，不完全な顧客要求文（タイプ N，S,または U の要件定義書）と完全な顧客要求文（タイプ NS の要件定義書）のペアを得られた．得られたペアを用いることで，要件定義書の閲読において特徴的な顧客要求と要件定義書の突合能力(必要十分性判別能力)を測ることができる. そこで, 本実験では, 得られたペアを課題として用いて, 課題のスコアと視線情報の関係を分析する.

閲読のプロセスを考慮すると，閲読能力の差が現れるのは，図 1.4 における（2）構造の理解と（3）欠陥の検出／修正である．閲読に求められるのは，内包的定義や外延的定義を基に突合可能な状態に変換することと，突合して必要十分か否か判別し欠陥を修正できることである．前者は，(2）構造の理解のプロセスに該当し，内包的定義である文と外延的定義である図形集合を基に突合可能な状態に変換する能力である．後者は，(3）欠陥の検出／修正のプロセスに該当し,本実験では内包的定義が外延的定義を必要十分に満たしているか(タイプ NS か否か）を判別する能力である.

そこで，本実験では二つの定義を基に閲読可能な状態に変換して，突合により必要十分性を判別する能力を閲読能力と捉え，実験を行う．被験者には，顧客要求文（集合＝製品に含まれる図形の満たす性質を表現した文の集合）と対応する製品（図形の集合）のペアが提示され，要求文が製品を必要十分に表現している否かを判断する課題に取り組んだ.

この必要十分性判別能力は，既存の欠陥検出率ではわからない閲読者の特徴を適切に表す指標であると考える.実際のソフトウェア開発における閲読では，欠陥検出率が評価基準となっていることから，必要最低限の欠陥検出ではその検出率が低くなる．その場合に，品質の良い閲読であるにもかかわらず，良い評価に繋がらない．最初から品質の良い要件定義書は欠陥検出率が低くなると想定されるが，その場合に欠陥を十分に検出されていないと評価されることが

[3] 本実験は研究業績 5 の日本認知科学会第 37 回大会抄録集(斉藤・日髙, 2020)，研究業績 3 の Proceedings of The 42nd Annual Meeting of the Cognitive Science Society (CogSci2020)(Saito & Hidaka, 2020)にて発表・掲載された論文としてまとめられた研究成果を一部含む.

ある．しかし，品質において重要なことはタイプ NS の要件定義書を作成することであり，検出率が高いことではないため，多くの欠陥を検出できる閲読者の能力が高いとは限らない．したがって，タイプ NS か否かを判別できる能力が重要である．この能力を本課題のスコアと視線情報を用いて定量的に測ることで，欠陥検出率によらない閲読品質を定量化できる．

ソフトウェアと要件定義書の 4 種類の関係（図 3.3）において，理想的な要件定義書はタイプ NS であり，閲読者は必要条件と十分条件の二つを検証する必要がある．要件定義書が必要条件を満たすかの検証には，提示された正例以外の負例が作成できるかを判別することが求められ，全ての要求文を基に判断する必要がある．一方で，十分条件を満たすかの検証には，特定の要求文の組合せにて作成できない正例があるかを判別することが求められ，全ての要求文を基にすることなく判別できる．したがって，二つの異なる論理的関係によって，視線の動きが異なると想定される．そのため，本実験の分析ではタイプごとに分析した．

閲読能力と視線情報の関係を分析するため，大別して二つの観点で分析する．一つ目は個人の閲読能力であり，本実験では正答率を個人の閲読能力を反映した指標とする．全試行での平均正答率が高い閲読者が，高い閲読能力を持っているとし，視線情報との関連を分析した．全試行での平均正答率が高い閲読者は，正答誤答によらず視線に特徴が現れるとの仮定の基で，正答率を各課題と紐づけ視線との関連を分析した．二つ目は，各試行の正誤答であり，正答時と誤答時では視線情報が異なるとの仮定の基で，正答／誤答における視線情報の特徴を分析した．

4.2 実験手続き

4.2.1 被験者

被験者は 20 人の日本人の成人（男性：16 名，女性：4 名）で，20 代が 1 名，30 代が 6 名，40 代が 4 名，50 代が 9 名であり，平均年齢は 44.9（SD = 9.7）であった．全て日本ユニシス株式会社のシステムエンジニアで，正常視（メガネやコンタクトレンズによる矯正視力を含む）であった．

4.2.2 課題

実験 2 にて 5 種類の幾何学図形の集合（ソフトウェア製品）に対して，それぞれ得られた複数のバージョンの顧客要求文を課題として用いた．具体的には，一つの画面イメージ（図 4.1）は顧客要求文（図 4.1 の左側），対応するソフトウェア（正例と負例，図 4.1 の中央），および質問と回答フォームからなり，顧客要求が製品を必要十分に表現しているかを回答する．1，2，4 ラウンドで作成された 9 つの不完全な顧客要求（必要でないまたは十分でない，または必要かつ十分でない）と 2 から 4 ラウンドで作成された四つの完全な顧客要求（必要かつ十分）を用い

て，計 13 題の課題を作成した（表 4.1）．タイプ N の課題は計 7 問，タイプ S の問題は 1 問，タイプ U の問題は 1 問，タイプ NS の問題は計 4 問とした．

13 問の問題は，client グループとラウンド番号を基にして，問題 ID を設定した．Client グループが 1 で，ラウンド 2 であれば，問題 ID は 1-2 とした．

表 4.1　13 問の課題（太字の顧客要求を用いた）

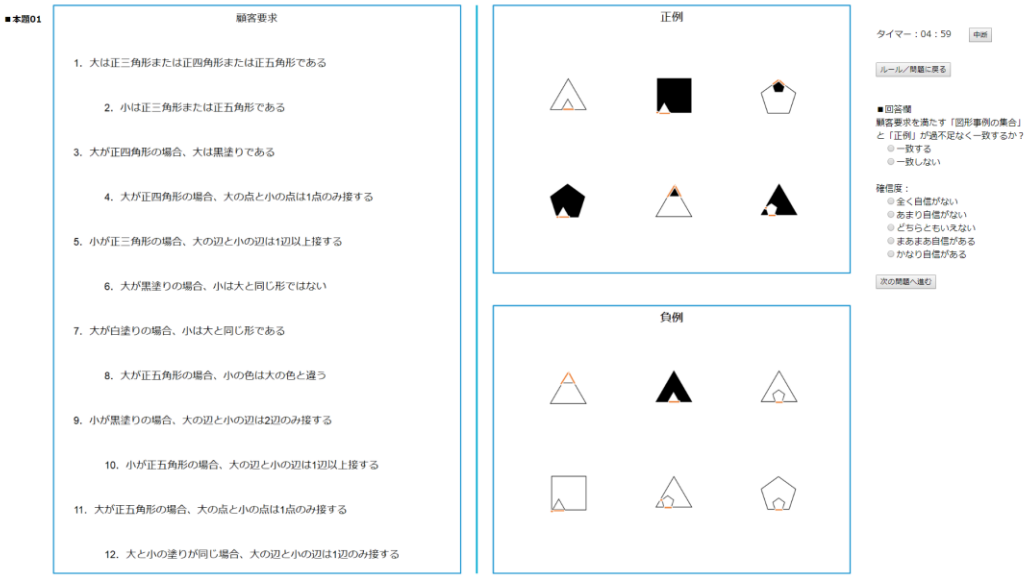
Client	集合PとRの関係			
グループ	ラウンド1	ラウンド2	ラウンド3	ラウンド4
1	N	NS	-	-
2	N	NS	-	-
3	U	N	N	NS
4	N	N	N	N
5	S	N	NS	-

4.2.3　手続き

それぞれの課題において，被験者は顧客要求文とソフトウェアのペアが与えられ，要求が製品を必要十分に満たすかどうかを回答する課題に取り組んだ．課題では顧客要求を満たす「図形事例の集合」と「正例」が過不足なく一致するかどうかを質問した．合わせて，回答の確信度を 5 件法（全く自信がない／あまり自信がない／どちらともいえない／まあまあ自信がある／かなり自信がある）にて回答した．取り組む課題は計 13 問とし，課題の形式に慣れるために二つの例題が提示された．1 題あたりの制限時間は 5 分とし，途中休憩は挟まなかった．実験時の画面イメージを図 4.1 に示す．

被験者が課題実施時の視線をアイトラッカで計測した．アイトラッカは視線情報を計測する専用のデバイスであり，本実験では gazepoint 社の GP3 HD Eye Tracker（サンプリングレート：150Hz，正確度（Accuracy）：0.5-1.0°，非接触型）を用いた．測定の精度を確保するため，実験中はなるべく頭部を動かさないように指示した．

課題終了後に，被験者は質問紙調査に回答した．



顧客要求

1. 大は正三角形または正四角形または正五角形である

2. 小は正三角形または正五角形である

3. 大が正四角形の場合、大は黒塗りである

4. 大が正四角形の場合、大の点と小の点は1点のみ接する

5. 小が正三角形の場合、大の辺と小の辺は1辺以上接する

6. 大が黒塗りの場合、小は大と同じ形ではない

7. 大が白塗りの場合、小は大と同じ形である

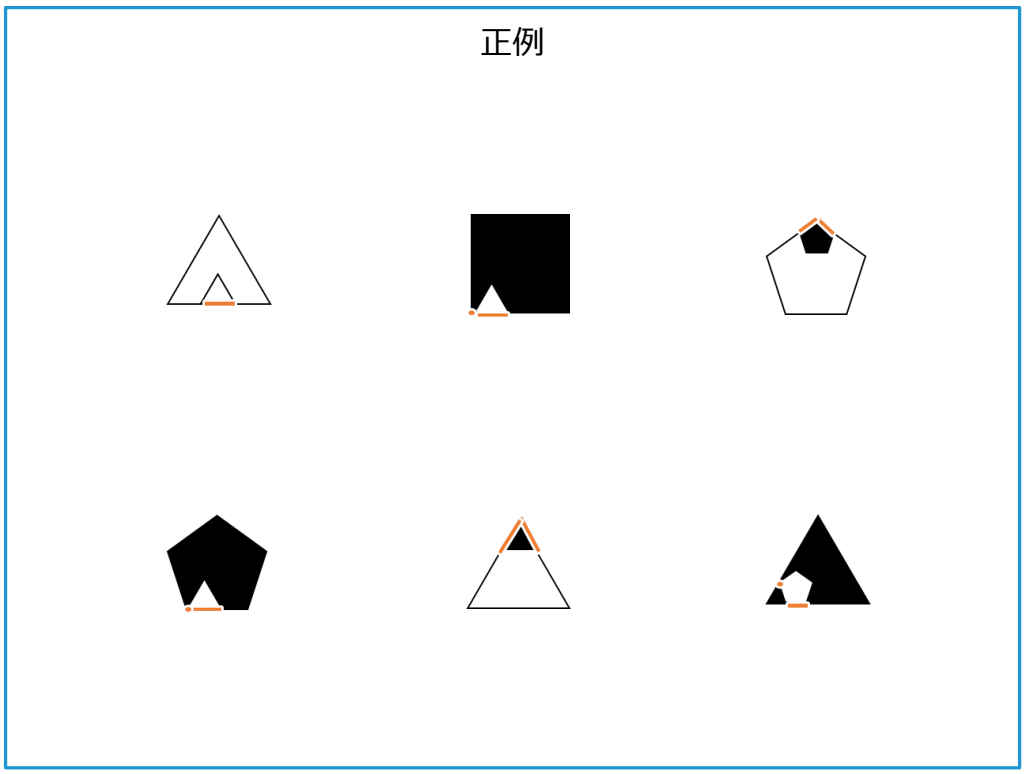
8. 大が正五角形の場合、小の色は大の色と違う

9. 小が黒塗りの場合、大の辺と小の辺は2辺のみ接する

10. 小が正五角形の場合、大の辺と小の辺は1辺以上接する

11. 大が正五角形の場合、大の点と小の点は1点のみ接する

12. 大と小の塗りが同じ場合、大の辺と小の辺は1辺のみ接する



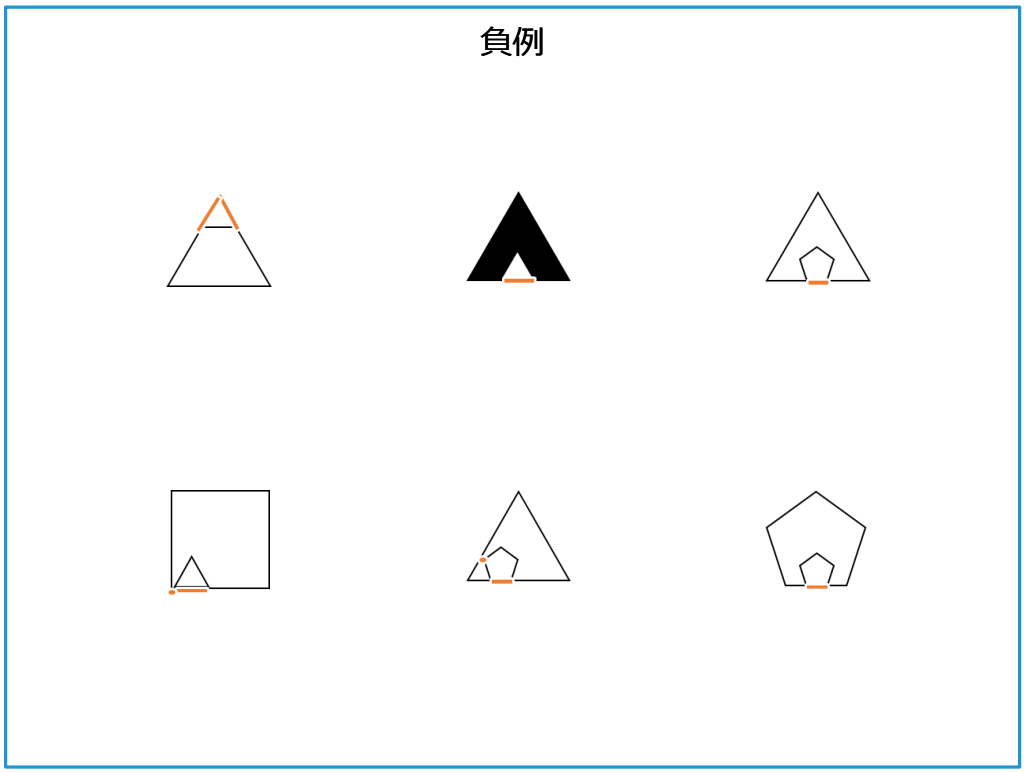


図 4.1　実験時の画面イメージと顧客要求，正例，負例それぞれの拡大図

4.3　結果と分析

要求が製品を必要十分に満たしているかという質問に対して，タイプ NS の問題に対しては「一致する」，それ以外に対しては「一致しない」と回答した場合に正答と定義した．13 問における正答率の平均／最小／最大は，68.6%／46.2%／84.6%であった．

視線情報が正しく取得できていない問題は，スミノルフ・グラブス検定(Grubbs, 1969)を用いて削除した．問題ごとに有効な fixation の割合が 0.4 以下の 4 問分を計測不良データとして削除し（図 4.2），以降の視線を用いた分析では，十分に視線が計測されているデータとして 256 問分のデータを用いた（20 人の被験者で 13 問，計 260 問）．Fixation は，gazepoint 社の Gazepoint analysis に実装されている独自アルゴリズム（Jacob の研究(Jacob, 1995)を基に作成）を用いて算出された．

まず，最初に被験者の正答率を対象として，経験との関連を述べ，タイプごとによる正答率の違いを分析する．問題の正誤によっても視線の特徴は変化すると考えられるため，問題単位で正答／誤答による違いを議論し，視線の違いを分析する．

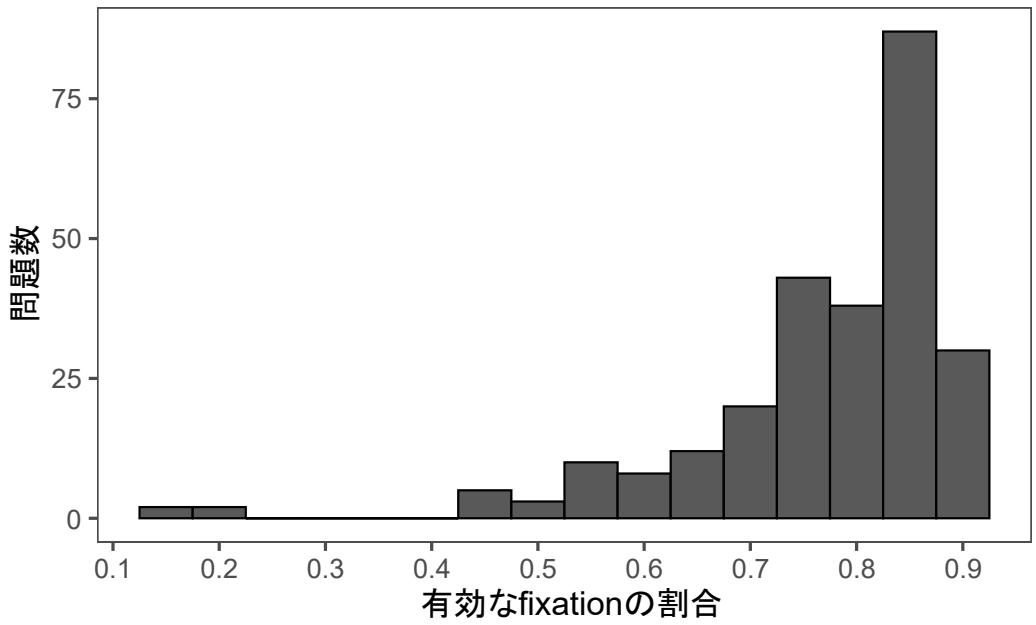


図 4.2　問題ごとの有効な fixation の割合のヒストグラム

4.3.1　分析 1：正答率の分析

4.3.1.1　経験と正答率の関係

質問紙調査では，要件定義書の閲読経験数を閲読経験に関する情報として収集した．分析では要件定義書の閲読経験数は比率尺度として扱うため，相関係

数算出にはピアソンの相関係数を用いた

要件定義書の閲読経験数と正答率の散布図を図 4.3 に示す．両者の間には，有意な相関は見られなかった（ピアソンの相関係数 $r = .0382, p = .873$）．この結果は，2.4.1 の結果と整合的であり，本課題において閲読経験は，閲読品質へ影響を与えないことが示された．

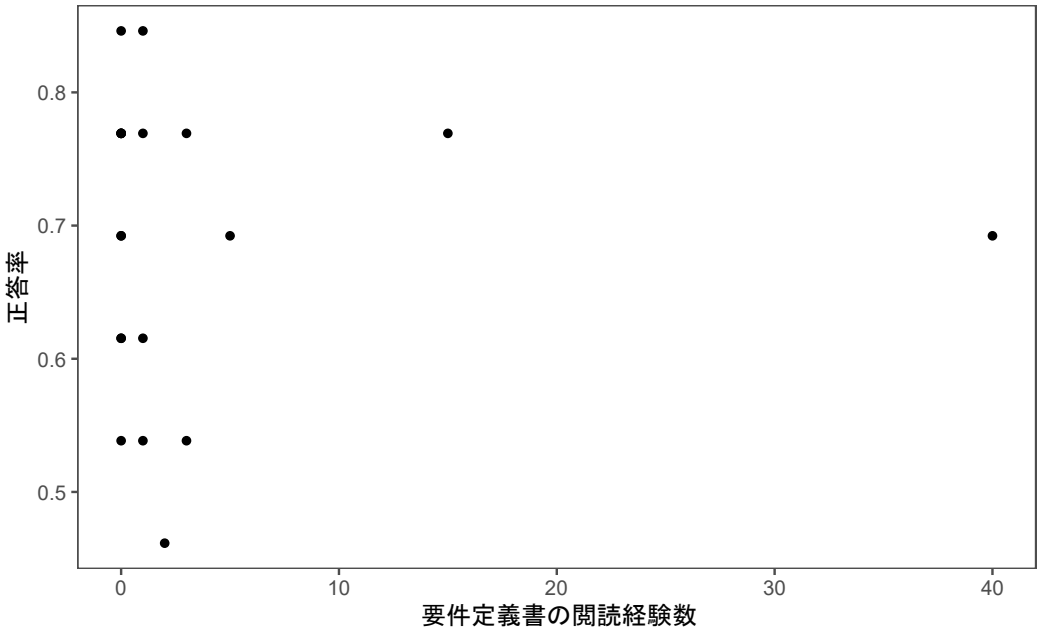


図 4.3　要件定義書の閲読経験数と正答率の散布図

4.3.1.2　タイプごとの正答率と回答時間

顧客要求を満たす事例の集合 R と正例 P のタイプごとの正答率は，タイプ U（必要条件も十分条件も満たさない）と S が最も高かった（表 4.2）．一方で，タイプ N と NS の正答率は，上記二つのタイプと比較すると低く，両者に大きな違いはなかった．必要性と十分性の二つの要因を考慮し，それぞれの因子の正誤に与える効果を二元分散分析にて分析した結果，必要性の因子による主効果が見られた（$p < .0005$）．十分性の因子による主効果や，両因子による交互作用は見られなかった．この結果から，要求と製品の間の二つの論理関係（必要性と十分性）によって閲読プロセスが異なるという仮説は統計的に棄却できないことが示された．

タイプごとに正答率が異なる一つの理由として，必要性を満たさないタイプ U と S の課題では顧客要求からは作成できない正例が製品に含まれており，要求が製品の必要条件を満たしてないことが判別しやすかったと考えられる．さらに，タイプ N と NS の課題では，要求が製品の必要十分条件を満たしているかを判別するためには，タイプ U や S のように例外ケース（作成できない正

例)を探すのではなく,文の意味する図形の特徴をみたす全ての正例を列挙し,余分な負例の有無を判別することが求められた．したがって，要求が製品の十分条件を満たすかを評価するタイプ N と NS より，必要条件を満たすかを評価するタイプ U と S の方が比較的簡単であると考えられる．

問題の難易度は回答時間にも反映されており，最も簡単と考えられるタイプ S と U では回答時間も有意に短くなっている（図 4.4)．タイプ S と U の特徴として，ある要求文にて作成できない正例が発生するため，文章を最後まで読むことなく顧客要求と正例が一致しないと判別できる．このような文を見つけられると，文章の途中で回答できるため，回答時間が短くなっていると想定される．

タイプ N と NS を比較した場合，タイプ NS にて回答時間が有意に長いため，その難易度が高い可能性がある．タイプ N と NS の場合は，タイプ S と U とは異なり,最後まで要求文を読む必要がある.その際,タイプ N においては,顧客要求から作成できる図形事例が正例より多いため，正例には存在しない図形が作成できることが一致しないと判別する基準となる．

さらにタイプ NS では，タイプ N と比較して回答時間が有意に長いため，タイプ N より難易度が高いと考えられる．タイプ NS の場合は顧客要求から作成できる図形事例は正例のみであり,タイプ N のように余分な図形の存在による判別ができないため，時間がかかっていると想定される．つまり，タイプ NS では正例以外に作成できる図形事例が存在するのかについて抜け漏れがないか確認する時間が必要となり,その分タイプ N よりも回答時間が長くなっていると考えられる．

表 4.2　顧客要求のタイプごとの正答率（タイプ N: 集合 R は集合 P の必要条件（necessary）であるが十分条件（sufficient）でない，タイプ S: 集合 R は集合 P の十分条件であるが必要条件でない，タイプ U: 集合 R は集合 P の必要条件でも十分条件でもない，タイプ NS: 集合 R は集合 P の必要十分条件である）

			十分性	
			あり	なし
必要性	あり	タイプ	NS	N
		試行（問題数）	80	140
		正答率	66.3%	62.9%
	なし	タイプ	S	U
		試行（問題数）	20	20
		正答率	95.0%	95.0%

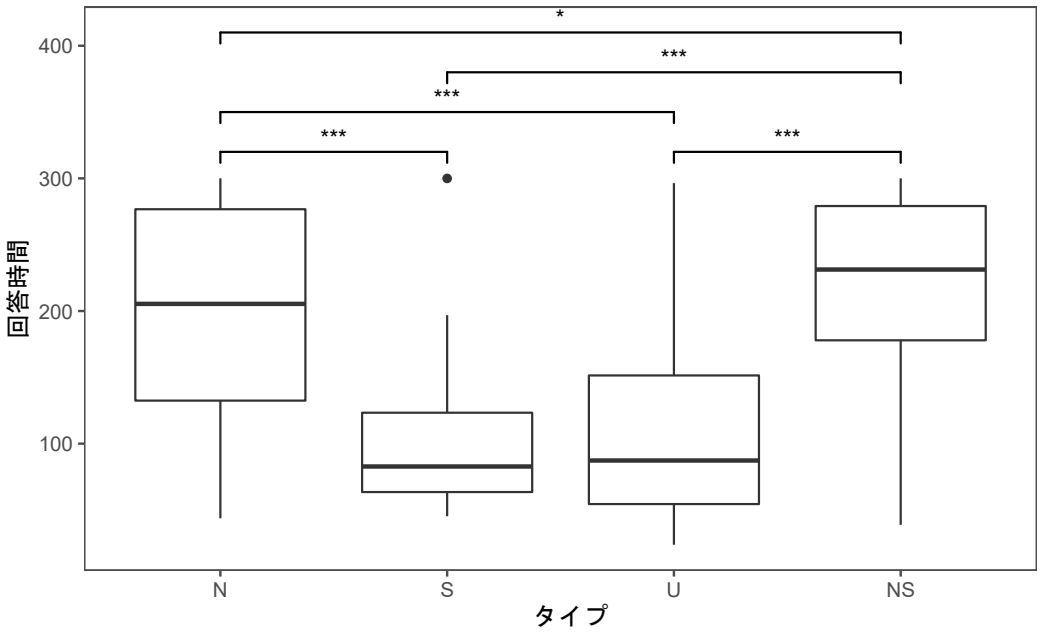


図 4.4　タイプごとの回答時間の箱ひげ図（$*: p < .05, ** : p < .01, *** : p < .005$）

4.3.1.3　タイプ N と NS の正答率の関係

被験者ごとのタイプ N とタイプ NS の正答率の散布図を図 4.5 に示す. 被験者ごとのタイプ N と NS の正答率の間には負の有意な相関がみられた（タイプ N の正答率とタイプ NS の正答率のピアソンの相関係数 $r = -.628, p = .00305$）. タイプ N においては，一致しないと答えた場合に正答であり，タイプ NS においては，一致すると答えた場合に正答（一致しないと答えた場合に誤答）である．そのため，タイプ N の正答率が高く，タイプ NS の正答率が低い被験者は一致しないと答える傾向がある．その場合には，要件定義書が顧客要求を必要十分に満たすにもかかわらず，一致しないと回答しているため，要件定義書に対して欠陥を指摘するバイアスがあるといえる．逆の場合は，一致すると答える傾向にあり，要件定義書が顧客要求を必要かつ十分に満たさないにもかかわらず，欠陥がないと判断するバイアスがあるといえる．

したがって，タイプ N と NS の正答率の偏りは，閲読者の欠陥の指摘バイアスを反映していると考えられる．欠陥を指摘するバイアスがある閲読者は，タイプ N の正答率が高くなる傾向にあり，そうでない閲読者はタイプ NS の正答率が高くなる傾向にある．

そのため，品質の良い閲読者の特徴は，全体の正答率が高いことに合わせてタイプ N と NS の正答率に偏りがないことである．欠陥を指摘するバイアスがある閲読者は，欠陥がない要件定義書に対して不要な欠陥を指摘することで，品質の低下や納期遅れを発生される要因になる．一方で，そうでない閲読者は

重要な欠陥を見逃してしまう可能性があり，品質の低下に繋がる．

本課題におけるタイプ N と NS のそれぞれの正答率を考慮することで，閲読者が欠陥を指摘する傾向にあるのか否かが判別できる可能性が示唆された．合わせて，実施期間を空けて本課題を実行することにより，その傾向の変化を分析できると考えられる．

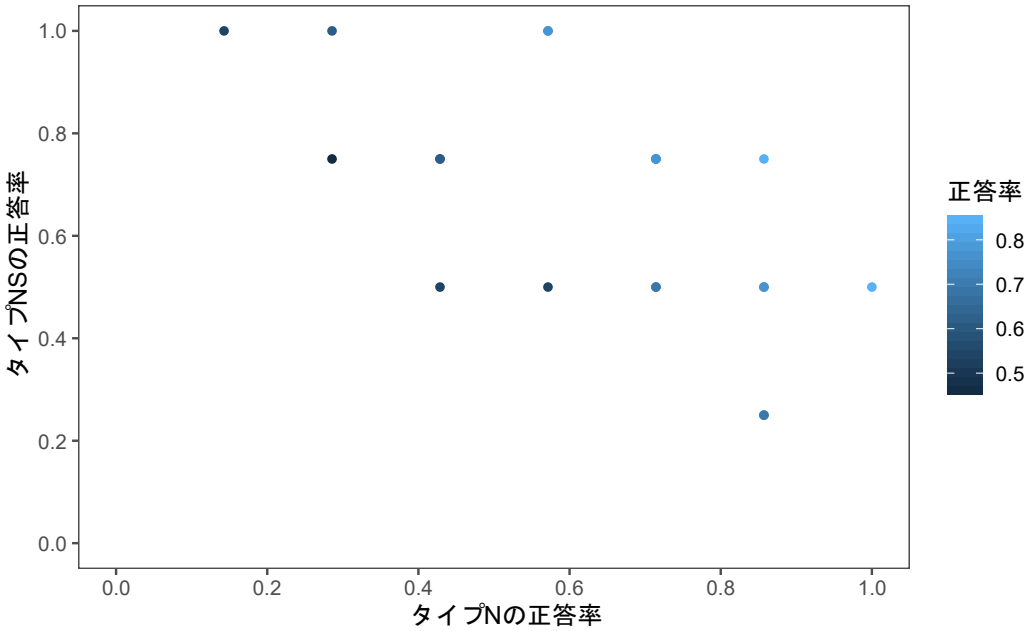


図 4.5　タイプ N とタイプ NS の正答率の散布図

4.3.2　分析 2：正答率のモデルベース分析

2.5.3 の分析手法と同様の手法を用いて，被験者の全試行の平均正答率の予測モデルを構築する．被験者の正答率は 13 問全体での正答率を用いた．

4.3.2.1　特徴量の抽出

2.5.3 で定義した 47 個の特徴量を基に，random forest（以降，RF）を用いて正答率に影響を及ぼす特徴量を抽出した．特徴量の抽出は，RF 重要度を基にして変数増加法を用いた．最初に，全データ（256 問題分のデータ）を用いて RF を実行して特徴量ごとの重要度を算出した．以降，RF における重要度はジニ係数を基にした．最も重要度が高い特徴量を用いて RF を実行して実測値との誤差である RMSE を算出する．RMSE は次式

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - \hat{y}_k)^2} \qquad (4.1)$$

で定義される．ここで，n は問題数，y_kは k 番目の問題の正答率（被験者の 13

問全体の正答率を用いるため，番号によらず同じ正答率を用いる)，$\hat{y}_k$はモデルによる予測値である．その後，次に重要度が高い特徴量を追加して RF を実行し，同様に RMSE を算出する．RMSE が最も小さくなるまで繰り返し，正答率に有効な特徴量を抽出する．

特徴量の抽出によって得られ 26 個の特徴量と重要度を表 4.3 に示す．瞳孔径を z スコア化した最小値が最も重要度が高かった．

表 4.3　RF により抽出された特徴量と重要度

順位	特徴量	重要度
1	瞳孔径の最小	0.442
2	瞳孔径の尖度	0.174
3	fixationの持続時間の中央	0.162
4	瞳孔径の歪度	0.158
5	瞬目の回数	0.135
6	fixationの持続時間の平均	0.124
7	saccadeの距離の平均	0.120
8	saccadeの角度の平均	0.117
9	saccadeの回数	0.111
10	瞳孔径のレンジ	0.109
11	saccadeの距離の標準偏差	0.102
12	saccadeの持続時間の標準偏差	0.099
13	fixationの持続時間の標準偏差	0.099
14	saccadeの角度の標準偏差	0.092
15	saccadeの距離の尖度	0.086
16	瞬目の時間の割合	0.084
17	瞬目率	0.084
18	saccadeの距離の尖度	0.083
19	瞳孔径の平均	0.082
20	saccadeの持続時間の最大	0.076
21	saccadeの距離の中央	0.073
22	saccadeの距離の尖度	0.073
23	saccadeの距離のレンジ	0.069
24	saccadeの距離の最大	0.068
25	瞳孔径の中央	0.062
26	瞳孔径の最大	0.040

4.3.2.2　正答率予測モデル

抽出した 15 個の特徴量を用いて，正答率予測モデルを複数のアルゴリズムで構築した．アルゴリズムは，2.5.3.2 で用いた，RF，decision tree（以降，DT），support vector regression（以降，SVR），multiple linear regression（以降，MLR）である．各アルゴリズムは R のパッケージを用いて，パラメータ等のチューニングは行わずライブラリのデフォルト値を採用した．

各アルゴリズムによる正答率予測モデルの有効性の確認は全データ（236 問分のデータ）で 10-fold cross validation を行い，mean absolute errors(MAE)，RMSE を算出して比較する．MAE は次式

$$\mathbf{MAE} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}}|\boldsymbol{y_k} - \boldsymbol{\hat{y}_k}| \qquad (\mathbf{4.2})$$

で定義される．

正答率予測モデルを構築し，アルゴリズムごとに得られた結果を表 4.4 に，に示す．SVR による閲読品質予測モデルの MAE，RMSE が最も低く，次いで RF であった．問題 1 題を解いている際の視線情報を用いることで，全体の正答率を 0.067 の誤差で予測できることが示唆された．

表 4.4　アルゴリズムごとの正答率予測モデルの誤差（太字が最も誤差が小さい値を示す）

		RF	DT	SVR	MLR
正答率	MAE	0.053	0.064	**0.049**	0.070
	RMSE	0.068	0.091	**0.067**	0.091

4.3.2.3　瞳孔径と平均正答率の関係

正答率を予測するための特徴量において，瞳孔径（z スコア）の最小値が最も重要度が高かった．本論文で用いた瞳孔径は，被験者ごとの全データを用いて z スコア化（平均値をゼロ，標準偏差を 1 に正規化）したものである．

瞳孔径（z スコア）の最小値と正答率の散布図を図 4.6 に示す．両者の間には負の有意な相関がみられた（瞳孔径（z スコア）の最小値と正答率のピアソンの相関係数 $r = -.483, p < .00001$）．正答率が高い場合には，z スコア化した瞳孔径の最小値が低い傾向がある．

課題実施時の瞳孔径を基にして z スコア化しているため，その最小値は課題実施時の瞳孔の挙動に影響を受ける．認知負荷がある状態では瞳孔径が大きくなる傾向がある報告を基にすると，課題を解いている際の瞳孔径も大きくなっていると想定される．したがって，z スコア化した瞳孔径の平均値は瞳孔が散瞳している状態であるといえる．そのため，散瞳している割合が大きければ大

きいほど，その平均値は大きな値を示す．さらに，人の瞳孔径は一般的に 2–8mm の範囲で周期的に変化するといわれており(宮尾他, 1992)，散瞳している状態が多数を占める場合には，その標準偏差も小さくなると想定される．瞳孔が開いている状態が多数を占める場合には，平均値が大きく標準偏差が小さくなり，瞳孔径の最小値は z スコア化した場合に，相対的に小さな値をとる．例えば，瞳孔径の最小値が 2mm と仮定し，平均値が 4mm で標準偏差が 2mm の場合では z スコア化した最小値は-1 であり，平均値が 6mm で標準偏差が 1mm の場合では，z スコア化した最小値は-4 である．

したがって，z スコア化した瞳孔径の最小値が小さいということは，瞳孔径の分布が正規分布から歪んだもしくは鋭いピークを示していると考えられる．瞳孔径は有限の値で変化するため，z スコア化した瞳孔径の最小値が小さい場合には，瞳孔径の平均が大きくかつ標準偏差が小さいと考えられる．つまり，瞳孔径が著しく小さくなるのではなく，常に散瞳している状態が多数を占めるような歪んでいるかつ鋭い分布では，z スコア化した瞳孔径の最小値が相対的に小さい値となる．

課題時の瞳孔径の分布の歪度と尖度と正答率を図 4.7，図 4.8 に示す．歪度は正規分布からの歪みを示す値で，正の場合は左に偏っており，負の場合は右に偏っている分布である．尖度は正規分布の鋭さを示す値で，正の場合は正規分布より鋭く長い裾野を持っており，負の場合は丸みを帯びたピークと短い裾野を持つ分布である．尖度においては，外れ値のような値が存在すると裾野が大きくなるため，尖度が著しく大きい値を示す．そのため，尖度についてはスミノルフ・グラブス検定(Grubbs, 1969)を用いて外れ値を含むデータを削除した（256 問中，尖度が 8 より大きい計 39 問分のデータを削除した）．歪度と正答率の間には負の有意な相関が（瞳孔径（z スコア）の歪度と正答率のピアソンの相関係数 $r = -.302, p < .00001$），尖度と正答率の間には正の有意な相関がみられた(瞳孔径(z スコア)の歪度と正答率のピアソンの相関係数 $r = .319, p < .00005$)．

したがって，正答率が高い被験者の瞳孔径は，正規分布と比較して右に偏った鋭いピークを持つ分布であることがわかった．安静時の瞳孔径は，周期的にその大きさを変化させるため，正規分布に近い分布を示すと想定される．そのため，正答率が高い被験者は課題実施時において，安静時と比較して瞳孔径が散瞳し（右に偏った分布），かつ全体における散瞳割合が高い（鋭いピーク）と考えられる．

そこで，安静時の瞳孔径と課題時の瞳孔径の比較を行った．安静時の瞳孔径の平均と標準偏差を基に，瞳孔径の平均と標準偏差の増減率を次式

$$\textbf{瞳孔径の平均の増減率} = \frac{(\text{課題時の瞳孔径の平均値} - \textbf{安静時の瞳孔径の平均値})}{\text{安静時の瞳孔径の平均値}} \quad \textbf{(4.3)},$$

$$\text{瞳孔径の標準偏差の増減率} = \frac{(\text{課題時の瞳孔径の標準偏差} - \text{安静時の瞳孔径の標準偏差})}{\text{安静時の瞳孔径の標準偏差}} \quad (4.4)$$

にて算出した．さらに，安静時と課題実施時の瞳孔径の分布の差を比較するためにカルバック・ライブラー情報量（KL 情報量）(Kullback, 1997)を算出した．

課題実施時の瞳孔径の平均の増減率と正答率の散布図を図 4.9 に，標準偏差の増減率と正答率の散布増を図 4.10 に，安静時の瞳孔径と課題実施時の瞳孔径の分布における KL 情報量を図 4.11 に示す．正答率の高い人／低い人の安静時と課題時の瞳孔径の分布を図 4.12 に示す．瞳孔径の平均の増減率と正答率の間には相関がみられなかった（ピアソンの相関係数 $r = -.0372, p = .550$）ものの，その多くは正の値を示したことから課題時により散瞳している傾向がある．瞳孔径の標準偏差と正答率の間には有意な負の相関（ピアソンの相関係数 $r = -.371, p < .00001$）がみられた．KL 情報量と正答率の間には有意な相関がみられなかった．KL 情報量に相関がみられなかった理由としては，正答率が低い場合だとしても瞳孔径が安静時と異なる分布を示す可能性が挙げられる．正答率が低い場合において，標準偏差の増減率が正の値をとる被験者がおり，その場合は散瞳と縮瞳を繰返している状態であると考えられる．そのため，瞳孔径が安静時と比較して異なる分布を示すため，正答率が低い場合でも KL 情報量が大きくなり，正の相関がみられなかったと考えられる．

正答率が高いほど，安静時と比較して瞳孔径の標準偏差が小さくなる傾向があり，かつ瞳孔径の平均値が増加している傾向があることから，正答率が高い被験者は課題時には安静時と比較して瞳孔が散瞳し，かつ散瞳割合が多いことが示唆された．特に標準偏差にて有意な負の相関がみられたことから，瞳孔径（z スコア）の最小値へは標準偏差による影響が大きい．正答率が高い人は，散瞳している割合が大きいことが示唆された．

瞳孔の散瞳は，認知負荷や課題の難易度と関連していることが報告されている(M A Just & Carpenter, 1993; Kahneman & Beatty, 1966)．一方で，注意力散漫やマインドワンダリングのような状態では縮瞳することが報告されている(Grandchamp et al., 2014; Kang et al., 2014)．正答率が低い人は，注意力が低下し課題に集中できていない可能性がある．そのため，瞳孔の散瞳と縮瞳を繰り返すことでその標準偏差が大きくなっていると考えられる．一方で，正答率が高い人は課題実施の間，常に認知的努力が大きく，集中して課題に取り組んでいるといえる．

したがって，z スコア化した瞳孔の最小値を用いることで，課題に対する集中度合いのような指標を定量化できる可能性が示唆された．

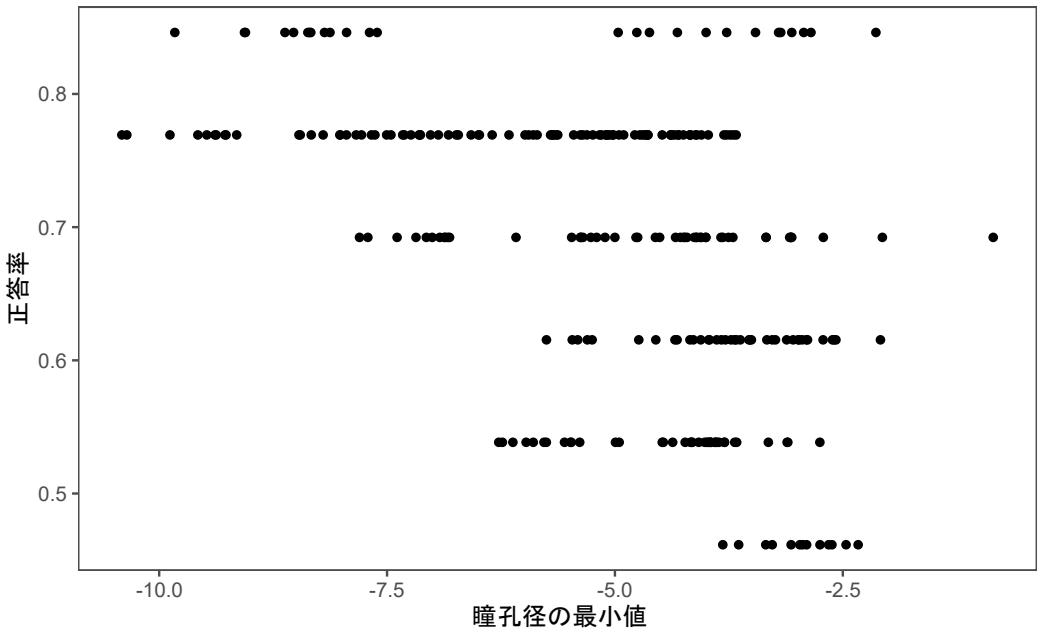


図 4.6　瞳孔径（z スコア）の最小値と正答率の散布図

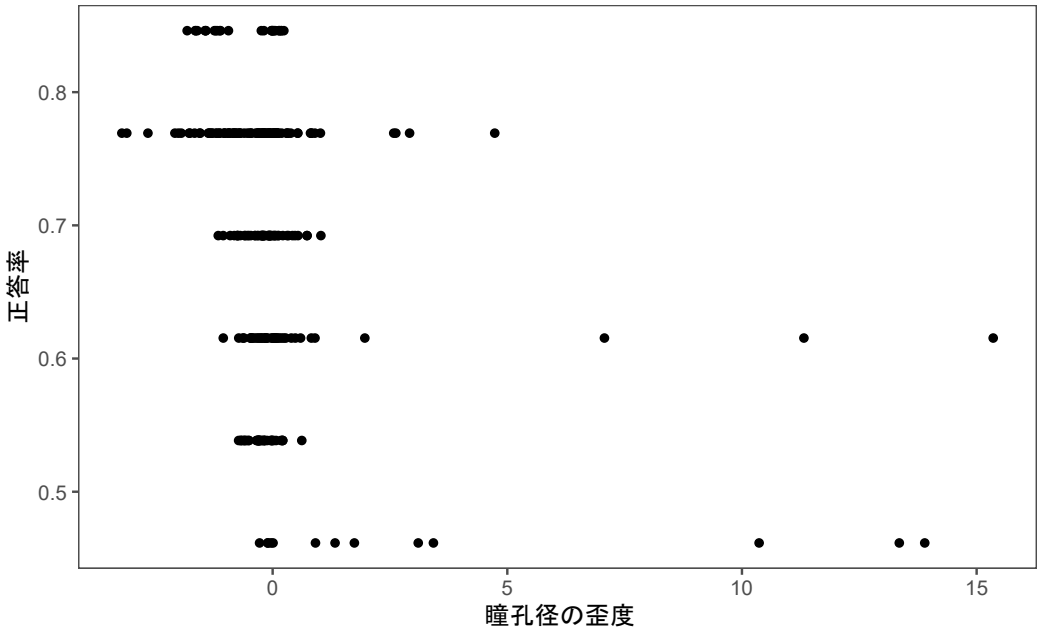


図 4.7　瞳孔径（z スコア）の歪度と正答率の散布図

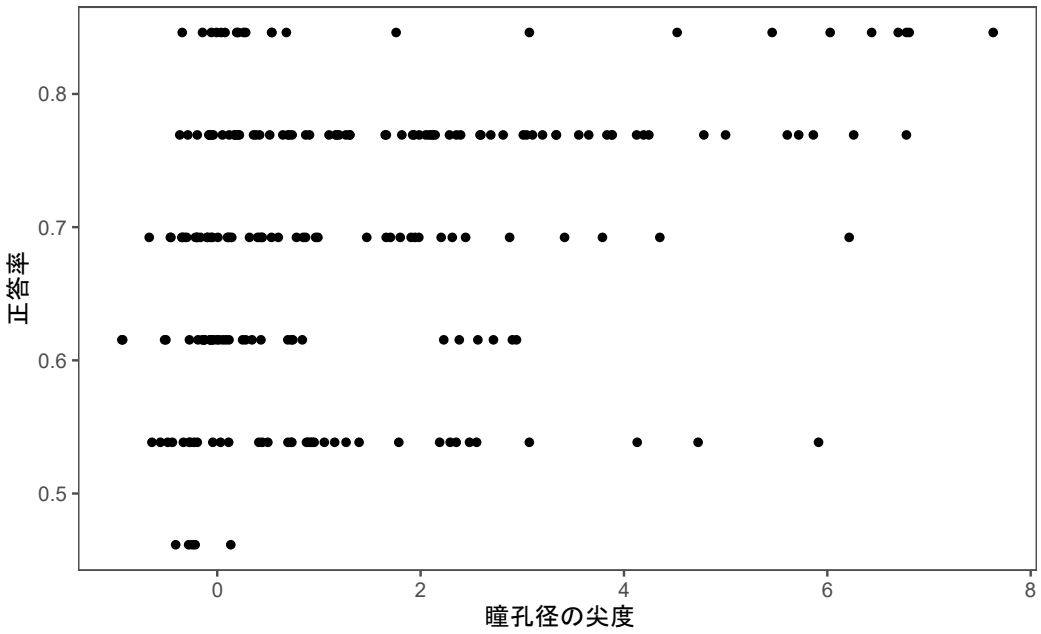


図 4.8　瞳孔径（z スコア）の尖度と正答率の散布図

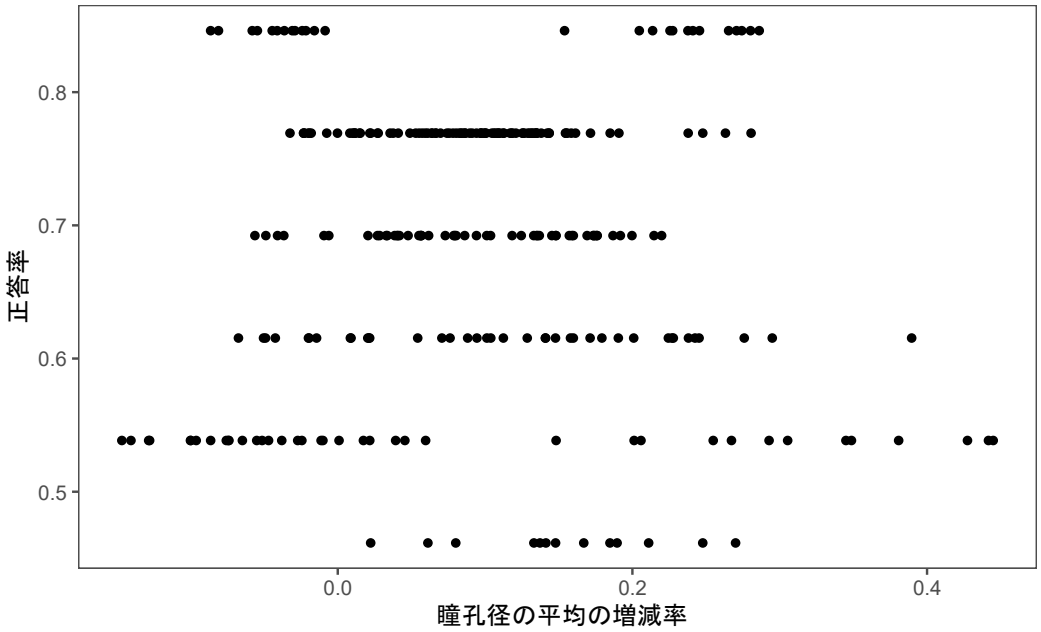


図 4.9　瞳孔径の平均値の増減率と正答率の散布図

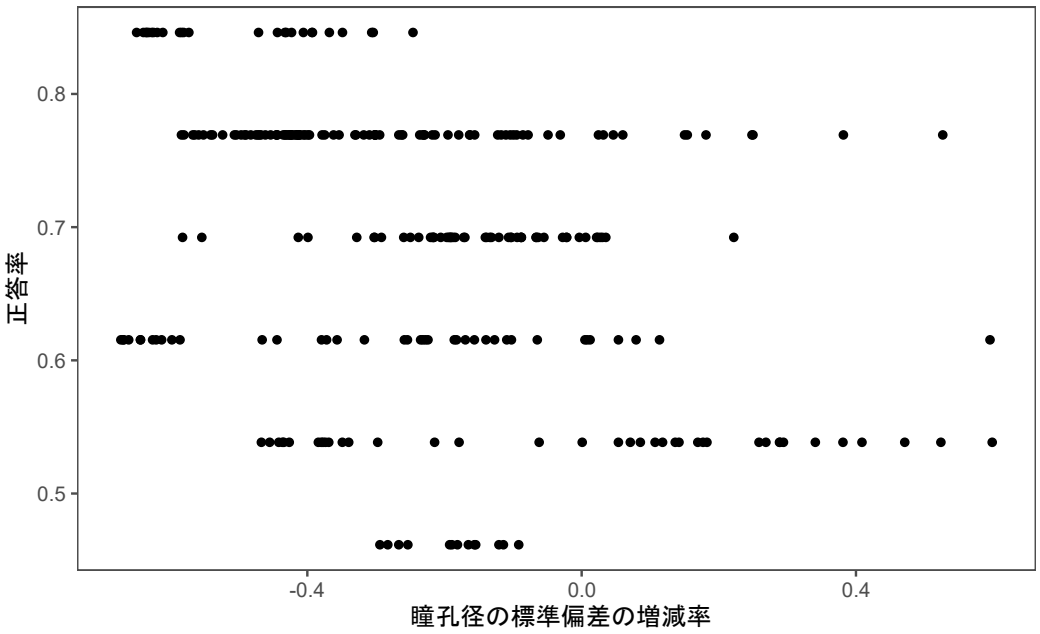


図 4.10　瞳孔径の標準偏差の増減率と正答率の散布図

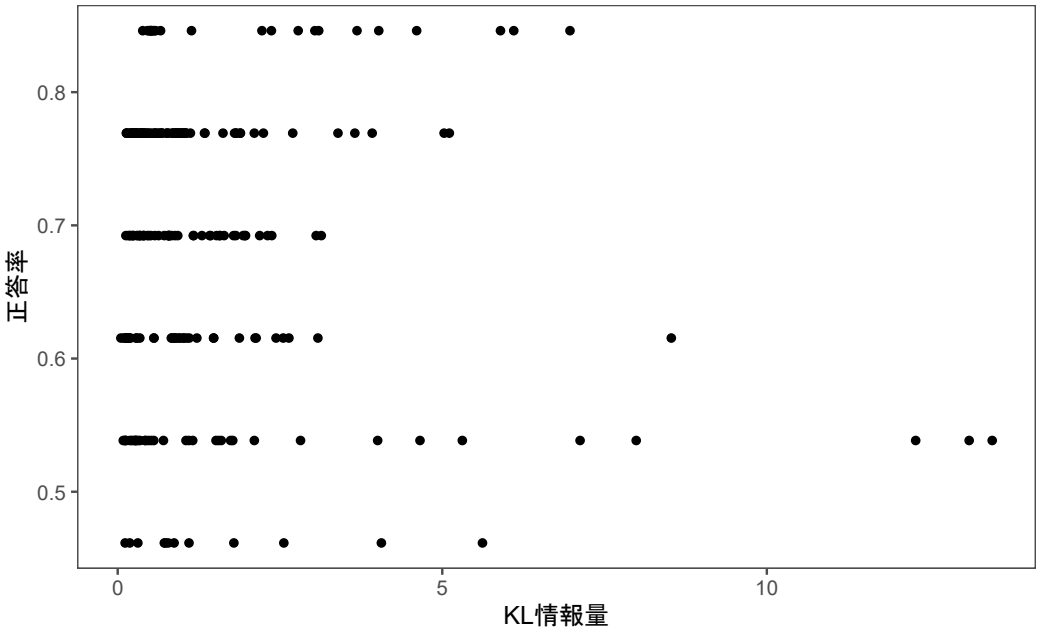


図 4.11　KL 情報量と正答率の散布図

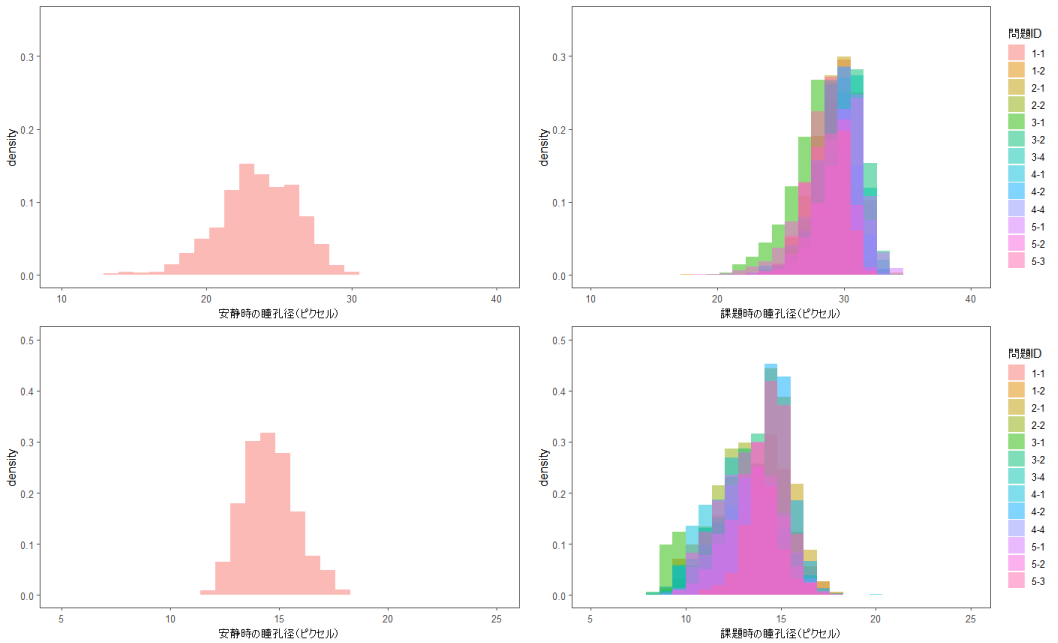


図 4.12　安静時と課題実施時の瞳孔径のヒストグラム（上段：正答率 0.847 の被験者，下段：正答率 0.538 の被験者）

4.3.3　分析 3：正答／誤答の分析

4.3.3.1　正答／誤答と回答時間

タイプごとの正答／誤答での回答時間の箱ひげ図を図 4.13 に示す．タイプ S と U においては 20 人中の 19 人が正答していたため，分析 3–5 では，タイプ N と NS のみ実施する．タイプ N の問題は 7 問，タイプ NS の問題は 4 問で，20 人の被験者で計 220 問である．そのうち，4.3 のスミノルフ・グラブス検定にて削除されたデータがタイプ N にて 2 問分あるため，以降の分析では計 218 問分のデータを用いた．

タイプ N では，正答時には有意に回答時間が短かった一方で，タイプ NS では正答／誤答で回答時間に差は見られなかった．さらに，タイプ N と NS では正答時間に有意な差が見られた．4.3.1.2 で述べたように，タイプ N においては一致しないと判別する材料があるため，顧客要求を基に負例を作成できることがわかった時点で回答できる．したがって，タイプ N の誤答時やタイプ NS の正答時と比べて短い時間で回答できると考えられる．

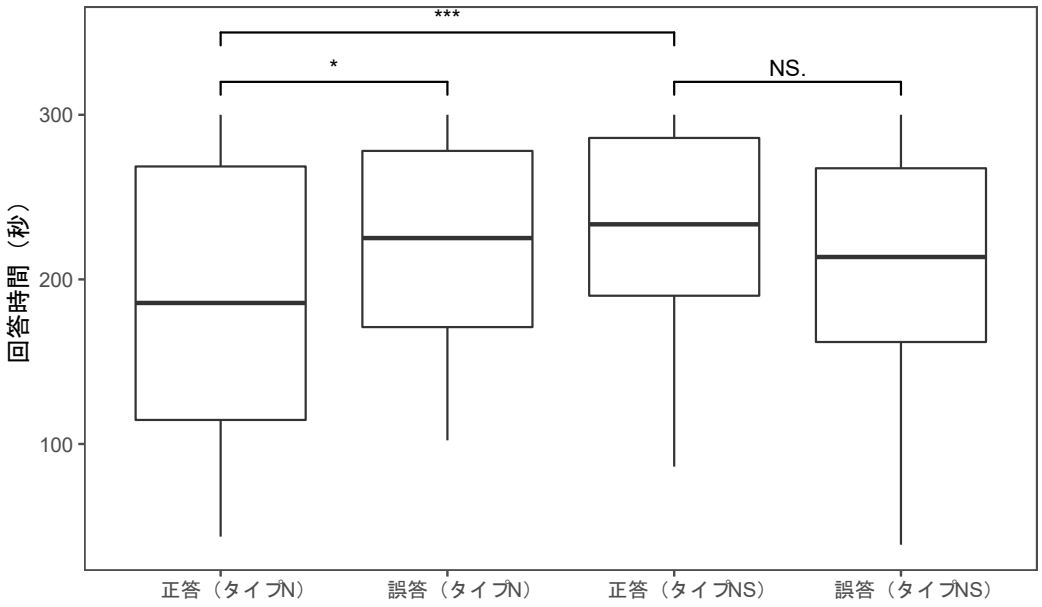


図 4.13　タイプごとの正答／誤答の回答時間の箱ひげ図（$*: p < .05, ** : p < .01, *** : p < .005$）

4.3.3.2　**正答／誤答と** fixation

閲読の正誤と視線情報を分析した．タイプ N における正誤ごと，顧客要求の文番号ごとの fixation 数を図 4.14 に，単位時間あたりの fixation 数を図 4.15 に，タイプ NS における fixation 数を図 4.16，単位時間あたりの fixation 数を図 4.17 に示す．回答時間が長ければ長いほど，全体の fixation 数は増加傾向にあるため，異なる回答時間の問題同士を比較する目的で単位時間当たりの fixation 数を算出した．タイプ N においては，正答者は前半の要求文をより注視し，後半になればなるほど注視回数は共に減る傾向にあった．誤答者も後半になればなるほど注視が減る傾向はあったが，その減少幅は正答者と比べると少なく，全ての要求文で一様に注視が分布する傾向があった．タイプ NS においては，要求文ごとの注視の違いは見られなかったものの，後半の要求文では注視が減る傾向にあった．

タイプ N の問題においては，後半の要求文に比べて前半をよく注視することで正答に繋がった可能性がある．

正答時においては，回答時間も重要な要素であり，閲読効率に関連して顧客要求文の読み方に影響を及ぼすと考えられる．本分析で得られた結果が正答時の特徴だとすると，回答時間が短く閲読効率が高い場合に顕著に現れる可能性がある．そこで，前半後半の注視の違いによって回答時間の予測ができるかを確認するため，予測モデルの構築を行う．

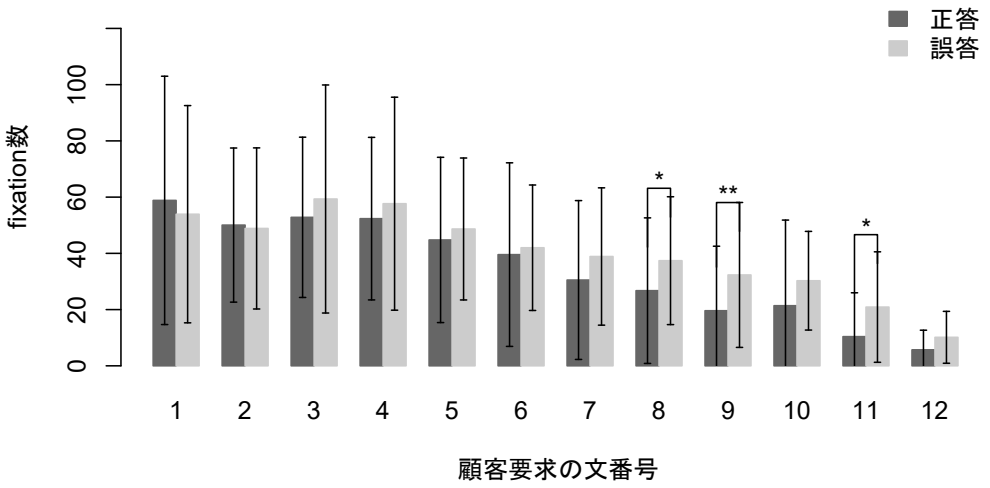


図 4.14　タイプ N における顧客要求の文番号ごとの fixation 数（エラーバーは標準誤差を示す，$*: p < .05, ** : p < .01$）

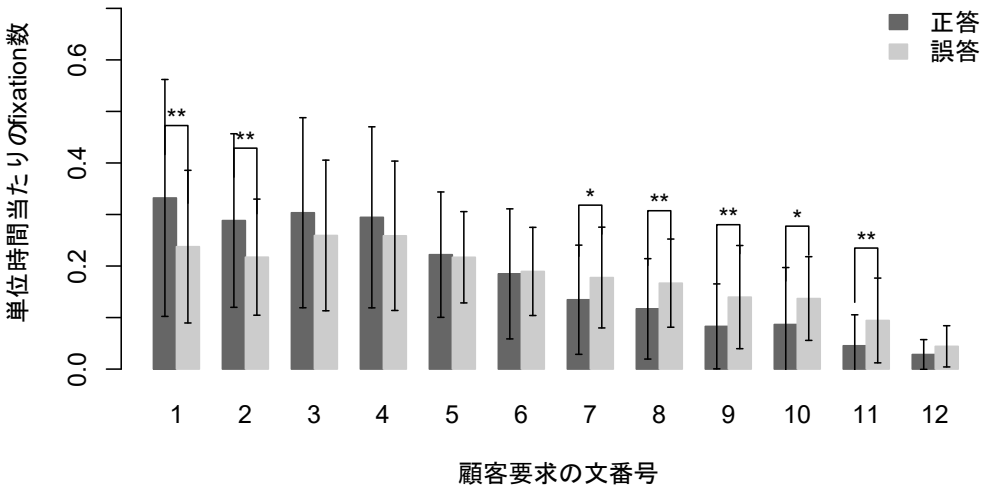


図 4.15　タイプ N における顧客要求の文番号ごとの単位時間あたりの fixation 数（エラーバーは標準誤差を示す，$*: p < .05, ** : p < .01$）

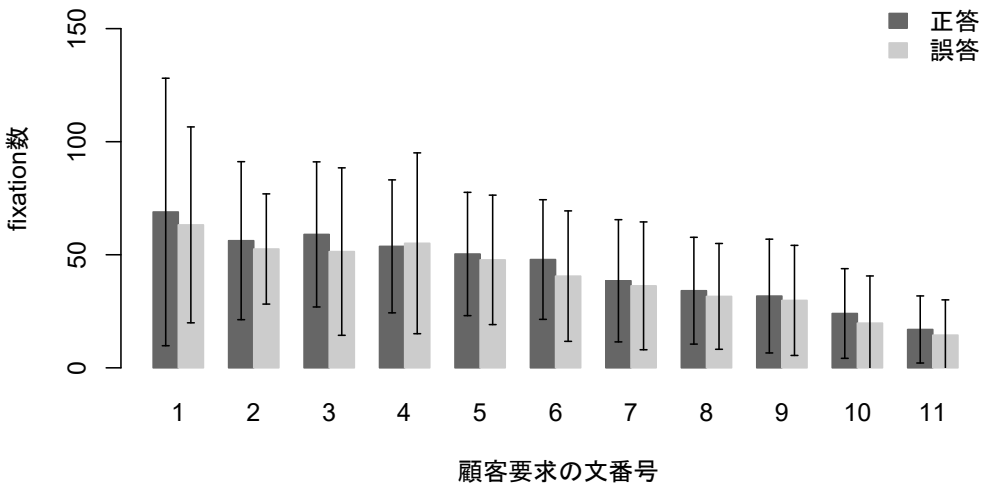


図 4.16　タイプ NS における顧客要求の文番号ごとの fixation 数（エラーバーは標準誤差を示す，タイプ NS の問題では文番号 11 が最大値である）

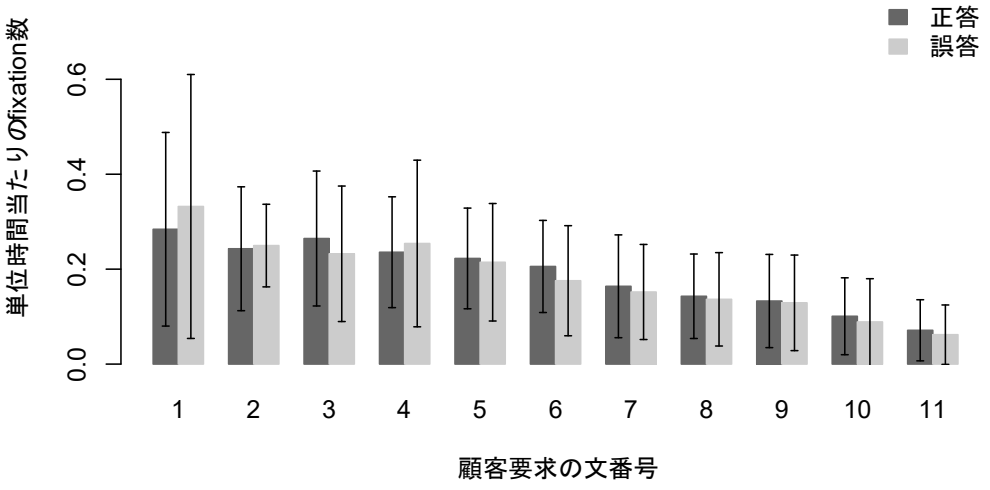


図 4.17　タイプ NS における顧客要求の文番号ごとの単位時間当たりの fixation 数（エラーバーは標準誤差を示す，タイプ NS の問題では文番号 11 が最大値である）

4.3.3.3　回答時間の予測モデル構築

回答時間の予測モデルの構築の特徴量として，顧客要求文における単位時間当たりの fixation 数の他に，正例と負例における単位時間当たりの fixation 数を用いた．被験者は顧客要求文と正例や負例を突き合わせながら，問題を解いているため，正例と負例の注視も特徴量として考慮に入れた．顧客要求文は最大で 12 個，正例と負例はそれぞれ 6 個ずつのため，計 24 個の特徴量を用いた．

タイプ N と NS の全データを用いて RF により，各特徴量の重要度を計算し，回答時間に影響を及ぼす要因を特定した．表 4.5 に上位 3 個の特徴の重要度を示す（24 個全ての特徴量の重要度は付録 B.1 に示す）．最も重要度が高かった単位時間当たりの fixation 数は，7 番目の顧客要求文であり，次いで 6 番目の顧客要求文であった．上位二つの特徴量が 3 位以下と比べて，顕著に重要度が高かった．

24 個の全特徴量を用いて，RF にて 10-fold cross validation を行った結果，RMSE は 58 秒であった．ベースラインとして，問題 ID と正誤の 2 個を特徴量として用いて，RF にて 10-fold cross validation を実施した結果，RMSE は 68 秒であった．Fixation を用いることにより予測精度が向上した．

6 番目と 7 番目の顧客要求文は，後半の要求文にあたり，その fixation 数が回答時間に影響を及ぼすことがわかった．重要度が高かった 6 番目と 7 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図を図 4.18 と図 4.19 に示す．それぞれ正答時には，正の有意な相関がみられた（正答時の 6 番目の顧客要求文における単位時間当たりの fixation 数と回答時間のピアソンの相関係数 $r = .354, p < .00001$，正答時の 7 番目の顧客要求文における単位時間当たりの fixation 数と回答時間のピアソンの相関係数 $r = .445, p < .00001$）．一方で，誤答時には有意な相関はみられなかった．

6 番目と 7 番目の顧客要求文における単位時間当たりの fixation 数の回答時間への影響を調べるため，二つを説明変数として重回帰モデルにより回答時間を予測し，実際の回答時間と比較した結果を図 4.20 に示す．6 番目と 7 番目の顧客要求文における単位時間当たりの fixation 数を用いた場合と同様に，正答時のみ有意な正の相関がみられた（重回帰による予測回答時間と回答時間のピアソンの相関係数 $r = .471, p < .00001$）．

したがって，正答時においては，効率的な閲読ができている場合は後半の要求文を見る fixation 数が減少する傾向にあると考えられる．

さらに，回答時間が短い場合において，6 番目と 7 番目の顧客要求文の fixation 数ゼロの回答者が存在した．タイプ N と NS の問題においては，その性質上全ての顧客要求文を読まない限り正答できない．にもかかわらず，全ての顧客要求文を読むことなく（fixation 数ゼロの顧客要求文が存在する），正答または誤答している人がいる．

全ての顧客要求文を見ずに回答している場合は，タイプ N や NS をタイプ S や U の問題と誤認識して回答していると考えられる．タイプ N の問題をタイプ S や U の問題と誤認識した場合は，回答者は「一致しない」を選択するため，結果として正答となる．一方で，タイプ NS の問題の場合は，「一致しない」を選択した結果，誤答となる．

つまり，タイプ N や NS の問題をタイプ S や U の問題と誤認識した場合には，タイプ N では正答のデータに誤認識時のデータが混入し，タイプ NS では誤答のデータに誤認識時のデータが混入する．さらに，タイプ S や U と誤認識した場合は，二つのタイプの問題の性質上，特定の要求文に達した段階で一致しないことが判別できる．そのため，それ以降の要求文は読まなくても回答できる．そのため，注視は特定の要求文までの文に集中して，それ以降の文では顕著に減少する傾向にある．特定の要求文の位置にもよるものの，前半の要求文で注視が減ることなく，後半の要求文で注視が減少すると想定される．合わせて，タイプ S や U における回答時間は短い傾向があるため（図 4.4），タイプ N にてタイプ S や U と誤認識した場合には，正答時に後半の注視が減少する特徴に合わせて，回答時間も短くなると想定される．それにより，正答時において，後半の要求文の fixation 数と回答時間に正の相関がみられた可能性がある．ただし，タイプ NS にてタイプ S や U と誤認識した場合には，誤答時に後半の注視が減少し，回答時間が短くなる特徴が現れるものの，誤答時には後半の fixation 数と回答時間に正の相関は見られない．したがって，fixation 数の違いを正確に議論するためには，誤認識による影響範囲を排除する必要がある．

表 4.5　RF による重要度上位 3 位の特徴量

順位	特徴量（単位時間当たりのfixation数）	重要度
1	7番目の顧客要求文	138198
2	6番目の顧客要求文	135054
3	1番目の正例	73085.1

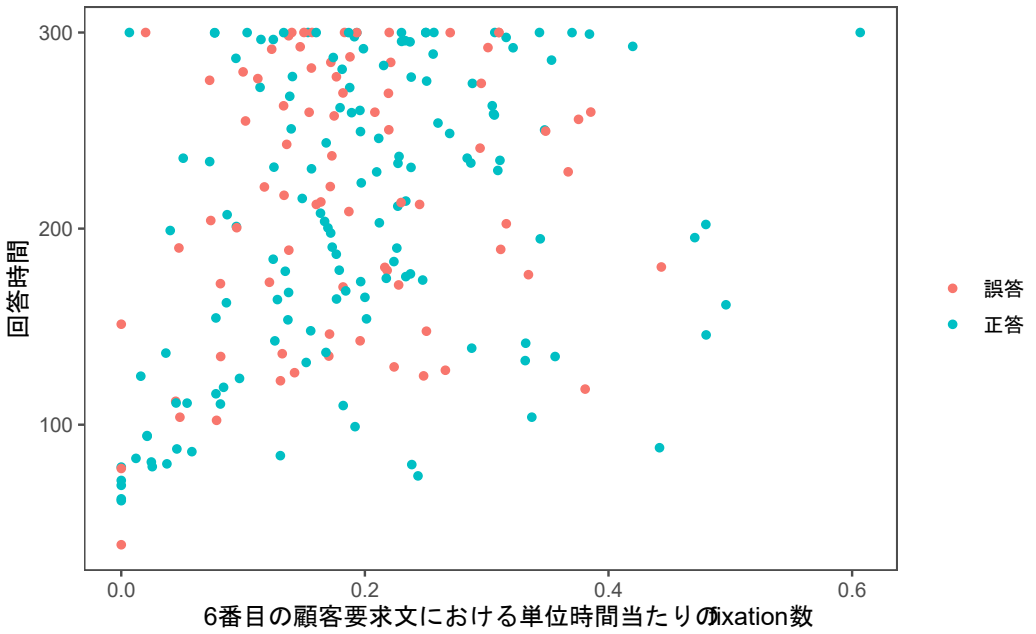


図 4.18　6 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図

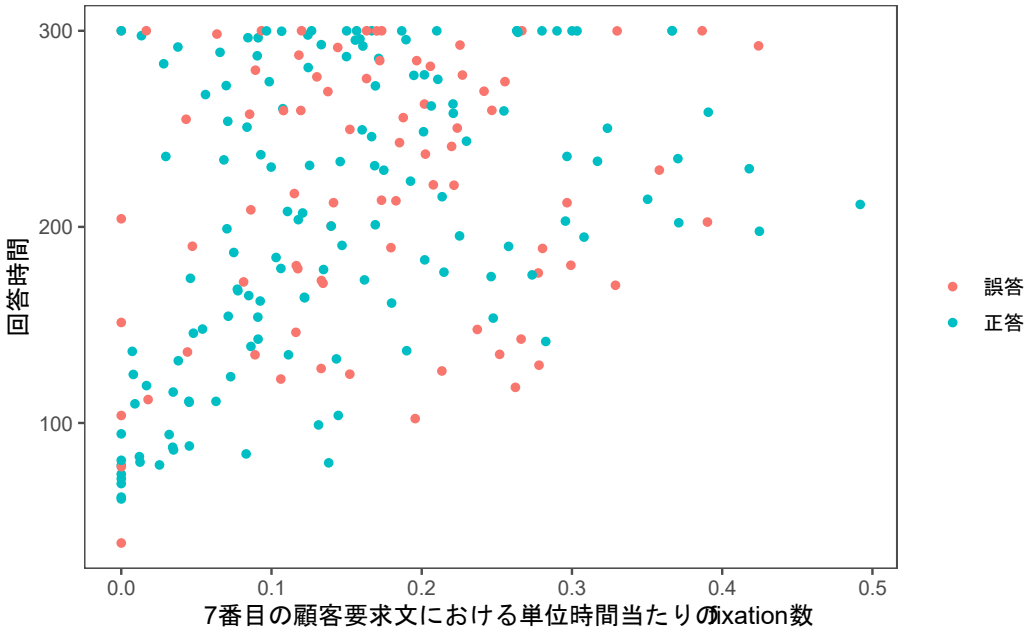


図 4.19　7 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図

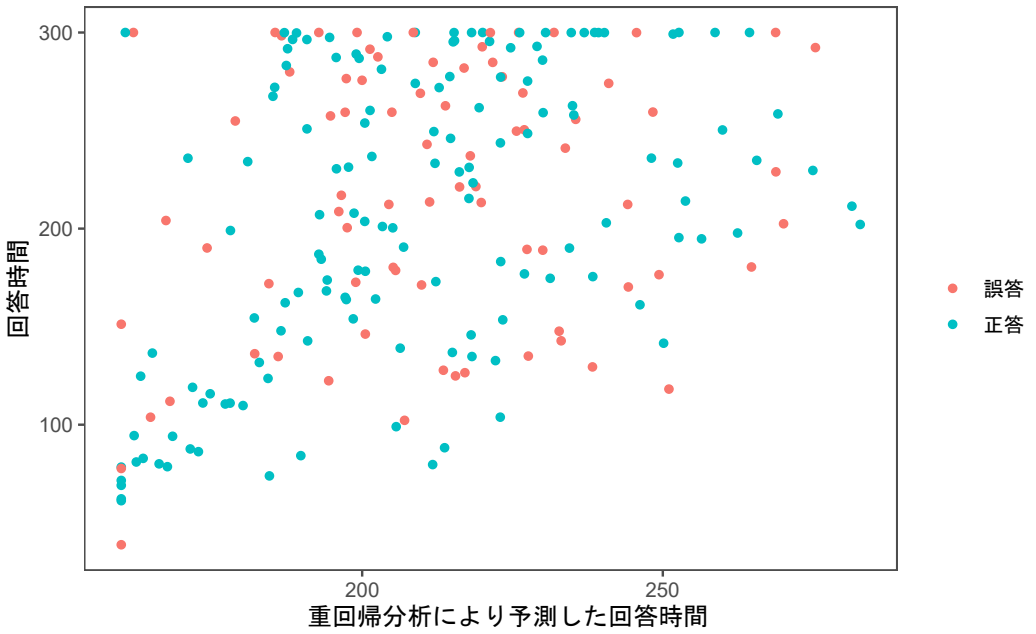


図 4.20　重回帰分析による予測回答時間と回答時間の散布図（6 番目と 7 番目の顧客要求文における単位時間当たりの fixation 数を説明変数として重回帰モデルを構築）

4.3.4　分析 4：誤認識を考慮した正答／誤答の分析

分析 3 での回答時間の予測モデル構築から，全ての顧客要求文を見ていない場合があり，タイプ N や NS の問題をタイプ U や S と誤認識していることが示唆された．それによって，前半後半の fixation 数が影響を受けた可能性がある．そこで，誤認識データを排除した，正答／誤答と fixation の分析を行う．

誤認識した場合には，全ての顧客要求文を見ない特徴があることから，問題ごとに全ての要求文の fixation 数がゼロより多いデータを用いて分析する．その結果，誤認識データは 218 問のうち 55 問であり，有効なデータは 163 問であった．タイプごとの内訳は，タイプ N では 138 問中 100 問，タイプ NS では 80 問中 63 問であった．タイプ N の問題の方が，誤認識している割合が多かった．

163 問のデータを用いて，正答／誤答における顧客要求文ごとの fixation の分析と，回答時間の予測モデル構築を同様に行った．

4.3.4.1　正答／誤答と fixation

タイプ N における正誤ごと，顧客要求の文番号ごとの fixation 数を図 4.21 に，単位時間あたりの fixation 数を図 4.22 に，タイプ NS における fixation 数を図 4.23，単位時間あたりの fixaiton 数を図 4.24 に示す．

タイプ N においては，4.3.3.2 で得られた傾向（正答時は前半の要求文をよ

り注視し，後半になればなるほど注視回数は共に減る傾向）と同様であった．ただし，有意差は 1 番目と 9 番目の顧客要求文における単位時間当たりの fixation 数でのみみられた．誤認識データによって，正答時に後半の注視が減少する傾向が強められていたため，削除することによって一部の文にて有意差がなくなったと考えらえる．しかし，誤認識データを削除した場合でも，4.3.3.2 と同様の傾向が得られたことから，前半の要求をより注視し，後半の注視が減る傾向は，正答時にみられる特徴であると結論付けられる．

タイプ NS においては，後半の要求文では，誤答時の方が fixation 数の平均が大きい傾向にあった．4.3.3.2 では，後半の要求文では誤答時の方が fixation の平均値が小さい傾向にあったため，逆の結果が得られた．有意差は得られなかったものの，誤認識データの影響を削除することで，タイプ N と同じように後半において正答時の注視が減る傾向にあった．

したがって，誤認識データを考慮すると，前半の要求文をより注視して後半の注視が減る傾向は，タイプ N と NS の両方において，正答時の視線の特徴である可能性が示唆された．

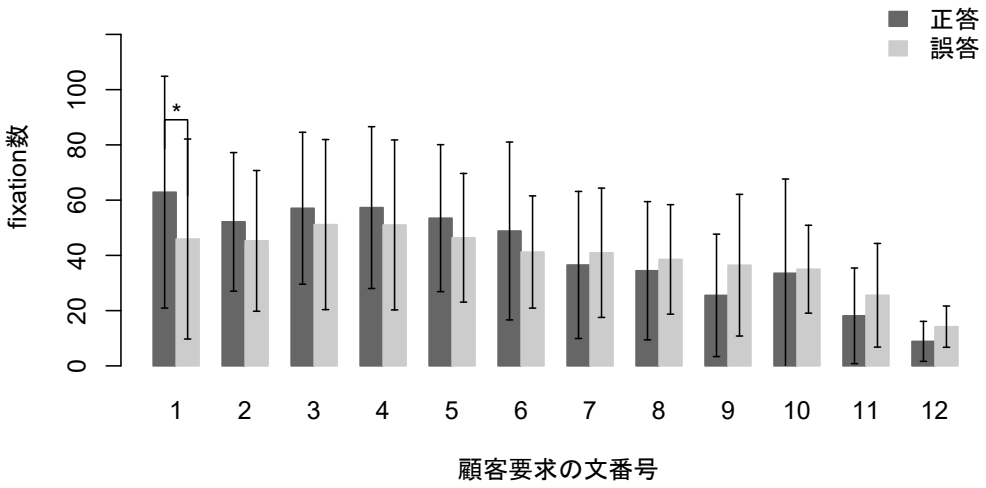


図 4.21　タイプ N における顧客要求の文番号ごとの fixation 数（エラーバーは標準誤差を示す，∗: $p < .05$）

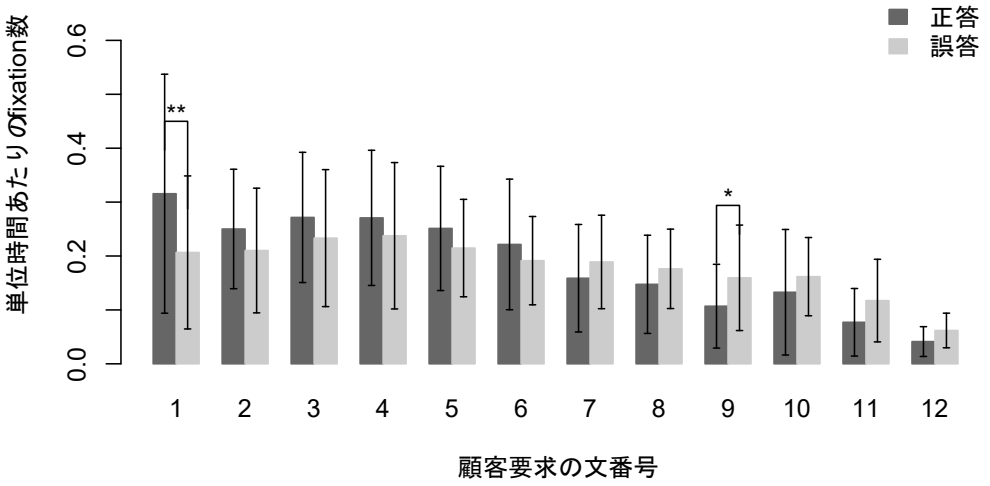


図 4.22　タイプ N における顧客要求の文番号ごとの単位時間あたりの fixation 数（エラーバーは標準誤差を示す，$*: p < .05, ** : p < .01$）

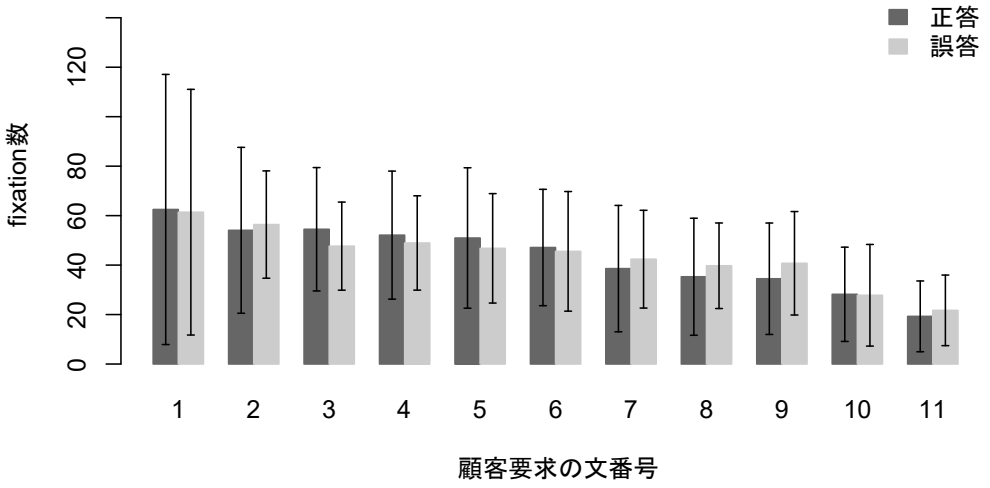


図 4.23　タイプ NS における顧客要求の文番号ごとの fixation 数（エラーバーは標準誤差を示す）

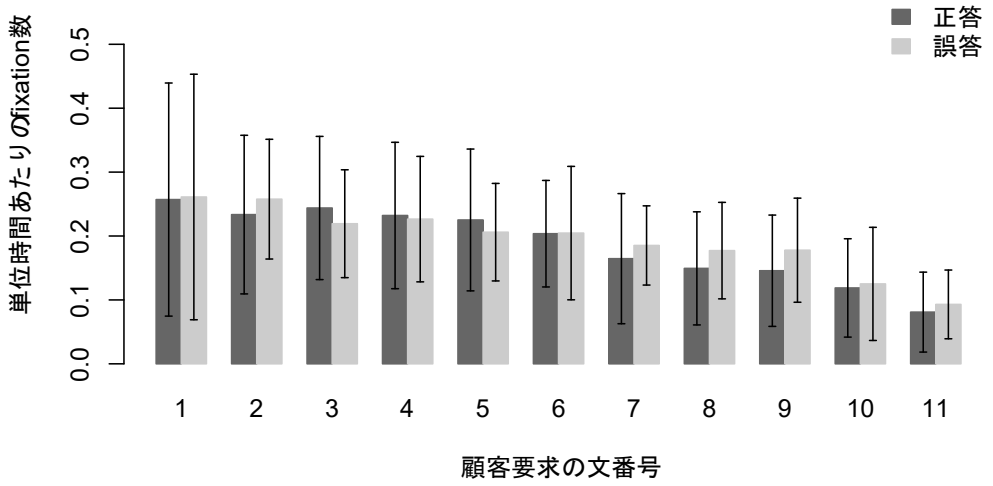


図 4.24　タイプ NS における顧客要求の文番号ごとの単位時間あたりの fixation 数

4.3.4.2　回答時間の予測モデル構築

4.3.3.3 で用いた 24 個の特徴量を用いて，回答時間の予測モデル構築を行った．表 4.6 に上位 5 個の特徴の重要度を示す（24 個全ての特徴量の重要度は付録 B.2 に示す）．最も重要度が高かった単位時間当たりの fixation 数は，7 番目の顧客要求文であり，次いで 3 番目の顧客要求文，第四位に 9 番目の顧客要求文であった．誤認識データを削除した場合でも，7 番目の顧客要求文は重要であることが示された．

24 個の全特徴量を用いて，RF にて 10-fold cross validation を行った結果，RMSE は 57 秒であった．誤認識を考慮しない差異の RMSE とほぼ同じ値であり，誤認識データによる回答時間の予測精度への影響はない．ベースラインとして，問題 ID と正誤の 2 個を特徴量として用いて，RF にて 10-fold cross validation を実施した結果，RMSE は 61 秒であった．誤認識を考慮するとベースラインの精度は向上したものの，fixation を用いたモデルの精度の方が高い値を示した．

7 番目と 9 番目との顧客要求文は，後半の要求文にあたり，その fixation 数が回答時間に影響を及ぼすことがわかった．さらに，前半の要求文である 3 番目の fixation 数も同様に回答時に影響を与えることがわかった．重要度が高かった 7 番目，3 番目，および 9 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図を，図 4.25，図 4.26，図 4.27 に示す．7 番目と 9 番目の顧客要求文ではそれぞれ正答時には，正の有意な相関がみられた（正

答時の 7 番目の顧客要求文における単位時間当たりの fixation 数と回答時間のピアソンの相関係数 $r = .265, p < .01$, 正答時の 9 番目の顧客要求文における単位時間当たりの fixation 数と回答時間のピアソンの相関係数 $r = .270, p < .01$). 3 番目の顧客要求文では，負の有意な相関がみられた（正答時の 3 番目の顧客要求文における単位時間当たりの fixation 数と回答時間のピアソンの相関係数 $r = -.226, p < .05$）．一方で，誤答時には有意な相関はみられなかった．

したがって，正答時において，効率的な閲読ができている場合は前半の要求文の注視が増え，後半の要求文への注視が減少する傾向にあると考えられる．これは，誤認識データを含めた場合と同様の結果であり，さらに，前半の要求文の注視が増えることも示唆された．

誤認識データを含めた場合と同様の結果が得られたため，後半より前半の要求文の注視が多い傾向は，正答時の特徴であると考えられる．なぜならば，タイプ N と NS の問題をタイプ S や U だと誤認識した場合は，図 4.4 より回答時間が有意に減少し，かつ全ての要求文を読むことなく回答できるため後半の注視が減少する．つまり，この注視の特徴は，正答時の特徴ではなく誤認識データの特徴である可能性があった．しかし，誤認識データを削除した場合においても同様の注視の傾向が得られたため，正答時にもこの特徴を持っていることが示唆された．さらに，回答時間が短いほど，後半より前半の要求文を注視する傾向が強められていた．短い時間で回答できる人は，閲読の効率が高いため，閲読能力が高いといえる．そのため，より閲読能力が高い人は，後半より前半の要求文を注視する傾向があることが示唆された．

表 4.6　RF による重要度上位 5 位の特徴量

順位	特徴量（単位時間当たりのfixation数）	重要度
1	7番目の顧客要求文	40770
2	3番目の顧客要求文	38709
3	3番目の正例	38572
4	9番目の顧客要求文	37249
5	1番目の正例	34780

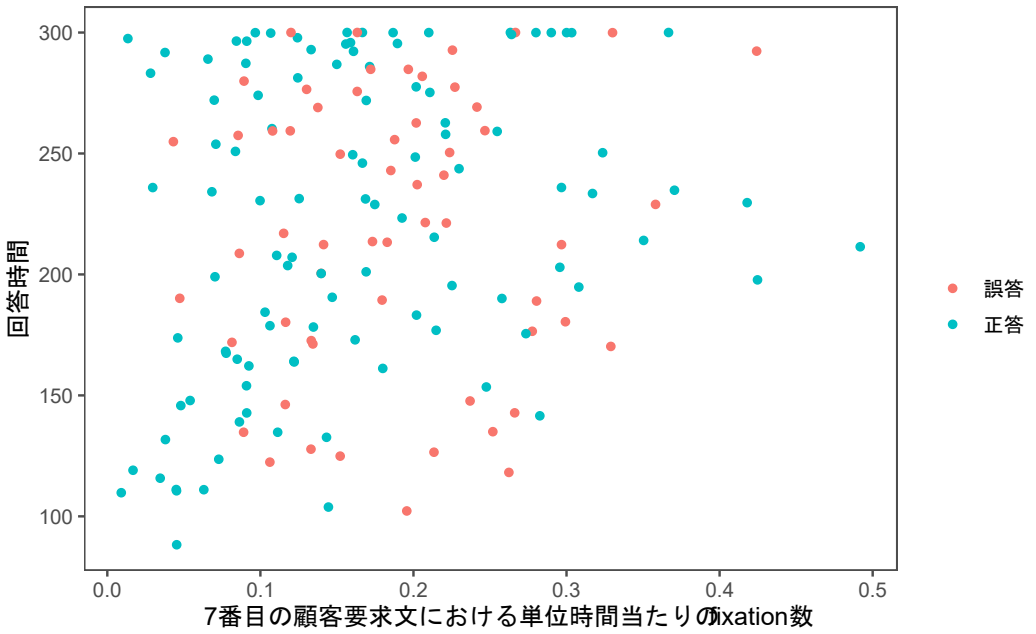


図 4.25　7 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図

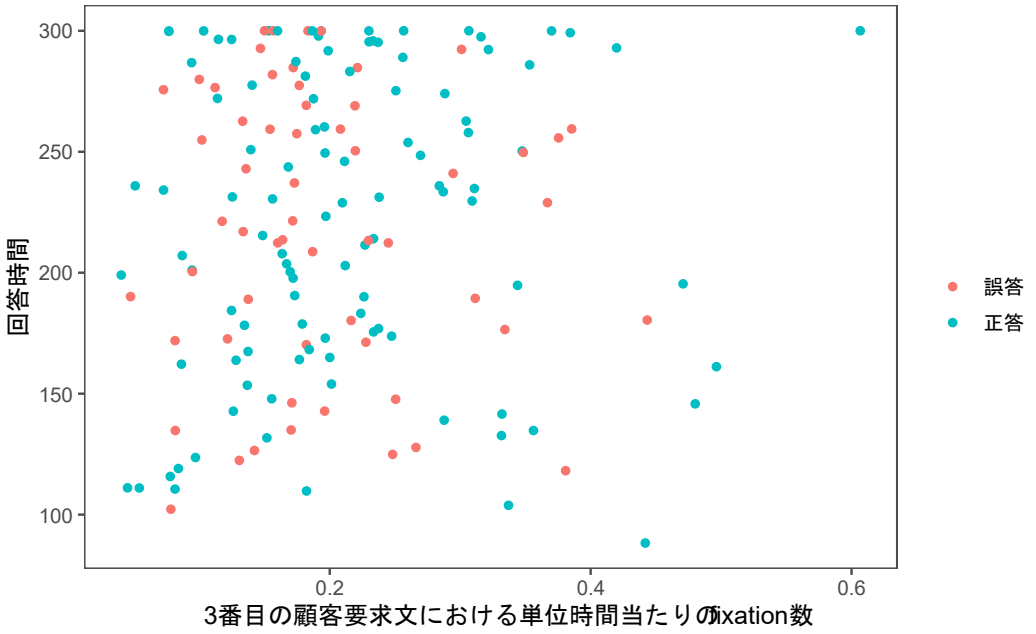


図 4.26　3 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図

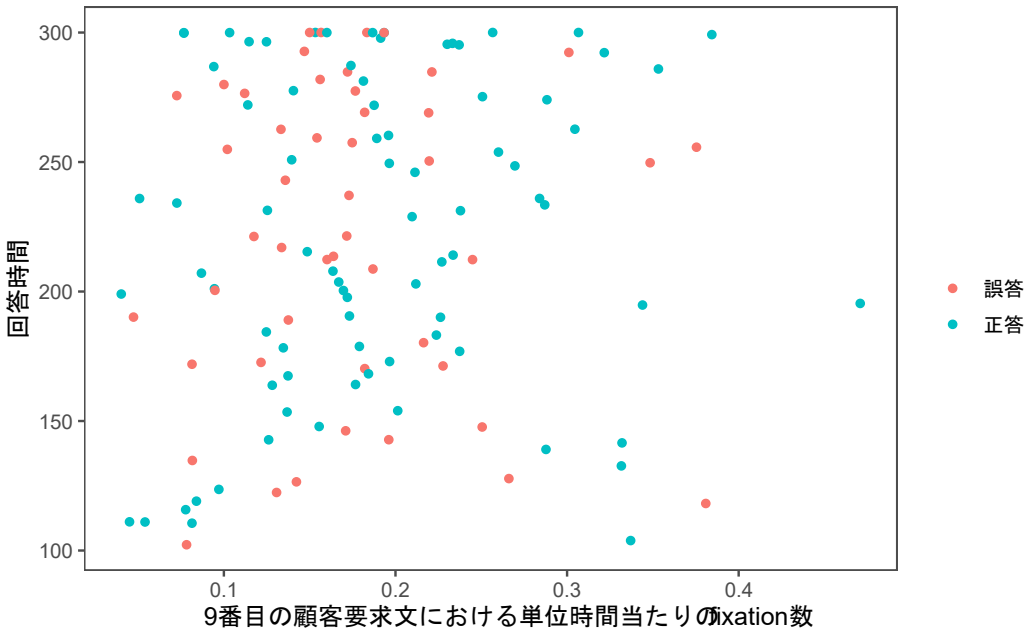


図 4.27　9 番目の顧客要求文における単位時間当たりの fixation 数と回答時間の散布図

4.3.5　分析 5：正答／誤答のモデルベース分析

本章では今までの分析で得られた正答／誤答の視線の特徴を基にして，正誤を分類するモデルを構築する．使用するデータは 4.3.4 にて誤認識していると想定されるデータを除いた計 163 問を用いる．ただし，正答と誤答のデータが不均衡であり，誤答データの方が少ないため，オーバーサンプリングにて誤答データの疑似的に増やす．

4.3.5.1　誤答データのオーバーサンプリング

誤認識と想定されるデータを削除した，タイプ N と NS の問題は計 163 問であり，そのうち正答は 105 問，誤答は 58 問であった．誤答データが正答データと比べて少ないため，機械学習モデルにおいて正答と予測される傾向が強くなる．不均衡データに対してはいくつか手法があり，本論文では少数のクラスのデータを増やすオーバーサンプリングを用いる．ただし，少数クラスのデータを単に複製するだけでは，過学習を起こす恐れがある．

そこで，SMOTE（synthetic minority over-sampling technique）(Chawla et al., 2002)を用いてオーバーサンプリングを行った．SMOTE は少数クラスのあるデータを基に，k 近傍法を用いて近傍データを探し，それとの差分に乱数をかけたものを元データに加えることで新しいデータを生成する手法である．SMOTE を用いて，誤答データを 58 問から，正答データと同じ 105 問まで増加させた．以降のモデル構築では，正答／誤答それぞれ 105 問ずつの計 210 問

分のデータを用いる.

4.3.5.2　特徴量の重要度

モデル構築に用いる特徴量は, 回答時間の予測モデル構築に用いた 24 個（顧客要求文，正例，負例それぞれにおける単位時間当たりの fixation 数）を用いる.

210 問の全データを用いて RF により，各特徴量の重要度を計算した結果の上位 3 位を表 4.7 に示す（24 個全ての特徴量の重要度は付録 B.3 に示す）．特徴量の上位 3 位に 7 番目と 8 番目の顧客要求文があり，後半の要求文に対する fixation 数は，正答／誤答分類において重要であることがわかった.

この結果は，正答時には前半の要求文の注視が増え，後半の要求文への注視が減少する傾向があるという事実を支持し，正答／誤答の違いは後半の要求文に特に顕著に表れると考えられる.

表 4.7　RF による重要度上位 3 位の特徴量

順位	特徴量（単位時間当たりのfixation数）	重要度
1	7番目の顧客要求文	8.35
2	6番目の正例	6.33
3	8番目の顧客要求文	5.71

4.3.5.3　正答／誤答分類モデル

RF にて 10-fold cross validation を行った結果，accuracy は 0.713，recall は 0.726，precision は 0.695 であった． Accuracy，recall，precision は次式

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.5}$$

で定義される．ここで，TP，TN，FP，FN は表 2.9 に示す混合行列の各要素である．Recall と precision 共に 7 割前後の精度であるため，SMOTE による誤答データ増加によって過学習は起こっていないと考えられる.

Accuracy は 7 割強にとどまっており，十分な分類精度とはいえない．モデルの精度に関する議論は考察の本実験の限界にて記述する.

4.3.6　分析 6：情報量と視線の関係

単位時間当たりの fixation 数は認知負荷と関連しており，共に高ければ高い

ほど，認知負荷も高くなる(Zagermann et al., 2016)．この事実を考慮すると，注視が前半の要求文に集中したことは，前半の要求文に含まれる情報量が他よりも高いもしくは後半に含まれる情報量が少ない可能性が示唆される．それにより，より情報量の多い文に着目して注視することで，正答率が上がった可能性がある．

集合 **P** は顧客が求めるソフトウェア製品を象徴化しているため，ソフトウェアごとに要素数は一定である．一方で，集合 **R** は，顧客要求文を読み進めるにつれてその要素数が減少するという特徴を持つ．したがって，顧客要求文がもつ情報量は，要素数の変化に起因するものであると考えられる．そこで，本研究では顧客要求文が進むにつれて変化する要素数などの特徴量に着目し，情報量の候補として採用した．それらの候補と視線情報の関係を分析することで，正答時の閲読者特性を明らかにする．

4.3.6.1　情報量の定義と分析手法

閲読課題にて，全ての被験者で文番号の若い順に読んでいるため，累積の文の部分集合で図形の集合を小さく絞ると考え，顧客要求の文番号が進むにつれて変化する集合 R に着目し，その要素数を基に三つの特徴量を情報量の候補として定義した．文番号をiとし，文 1 からiまでの文の集合を満たす図形の集合をR_iとする．三つの特徴量は，集合R_iの要素数，要素の減少数および減少率とした．ここで，要素数の初期値は組合せ可能な図形事例の最大値である 267 とし，最初の顧客要求文においては初期値からの要素数の変化を減少数／減少率とした．2 番目以降の顧客要求文においては，一つ前の文で作成される集合を基として要素数を計算し，その変化を減少数／減少率とした．

三つの特徴量を基に，顧客要求文との相関と fixation との関連を分析し，三つの特徴量が顧客要求文の情報量として妥当性を調査した．最初に，顧客要求文ごとの特徴量の推移を確認する目的で，顧客要求文番号と特徴量の相関関係を算出した．第二に．正答時の閲読者特性を調査する目的で，正答／誤答それぞれで特徴量における単位時間当たりの fixation 数のヒストグラムを算出した．単位時間当たりの fixation 数は，各要求文を注視している際の fixation の総数を注視時間で割ったものである．一般に fixation の総数は，注視時間に比例して増加するため，文章の長さや読みやすさなどの影響を受ける．それらの影響を排除するため，注視時間で割った単位時間当たりの fixation 数を用いた．ヒストグラムを基に，正答／誤答で特徴量による fixation 数の違いを分析し，三つの特徴量が情報量としての妥当性を評価した．

4.3.6.2　特徴量と顧客要求文との相関関係

13 問において，顧客要求の文番号ごとの三つの特徴量の平均，標準偏差を表 4.8 に，文番号ごとの要素数の箱ひげ図を図 4.28 に，減少数の箱ひげ図を図

4.29 に，減少率の箱ひげ図を図 4.30 に示す．三つの特徴量は，顧客要求の文番号との間に負の有意な相関がみられ（要素数と顧客要求文番号のスピアマンの相関係数 $r=-0.950, p<0.001$, 減少数と顧客要求文番号のスピアマンの相関係数 $r=-0.853, p<0.001$，減少率と顧客要求文番号のスピアマンの相関係数 $r=-0.443, p<0.001$），要素数が最も相関が強く，次いで減少数，減少率であった．

実験 3 では，正答の場合には前半にてより fixation が多い傾向にあったため，三つの特徴量の値の大小が情報量を示していないことが示唆された． fixation が大きいほど，それに費やす認知負荷が大きいことから正答／誤答で有意に差がみられた前半の文において最も情報量が高いと想定される，しかし，三つの特徴量はすべて有意な負の相関がみられたため，単に三つの特徴量の大きさが情報量を示すのではなく，特定の値にて情報量が多いと考えられる．そこで，三つの特徴量と fixation の関係を分析し，正答／誤答にて特徴量による fixation の違いを調査した．

表 4.8 顧客要求の文番号ごとの三つの情報量の平均／標準偏差の推移

文番号	データ数	要素数		減少数		減少率	
		平均	標準偏差	平均	標準偏差	平均	標準偏差
1	13	153.0	23.8	114.0	23.8	0.427	0.089
2	13	80.7	23.6	72.3	26.5	0.469	0.141
3	13	58.5	16.0	22.2	19.1	0.253	0.145
4	13	43.5	15.8	15.0	6.4	0.265	0.125
5	13	29.0	10.4	14.5	8.9	0.320	0.148
6	13	23.9	11.5	5.1	3.8	0.195	0.161
7	13	15.1	4.4	8.8	9.7	0.290	0.230
8	10	13.3	4.6	2.7	1.9	0.176	0.123
9	10	10.2	2.2	3.1	2.7	0.207	0.097
10	8	7.9	1.1	2.9	1.9	0.253	0.122
11	7	6.6	1.8	1.6	1.1	0.203	0.170
12	1	7.0	−	1.0	−	0.125	−

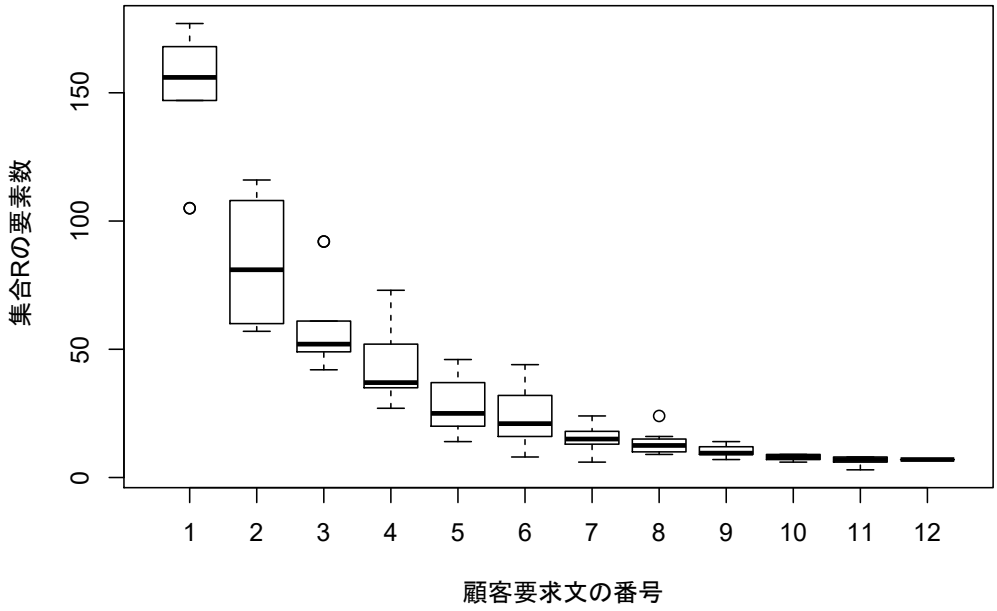


図 4.28　顧客要求の文番号ごとの要素数の箱ひげ図

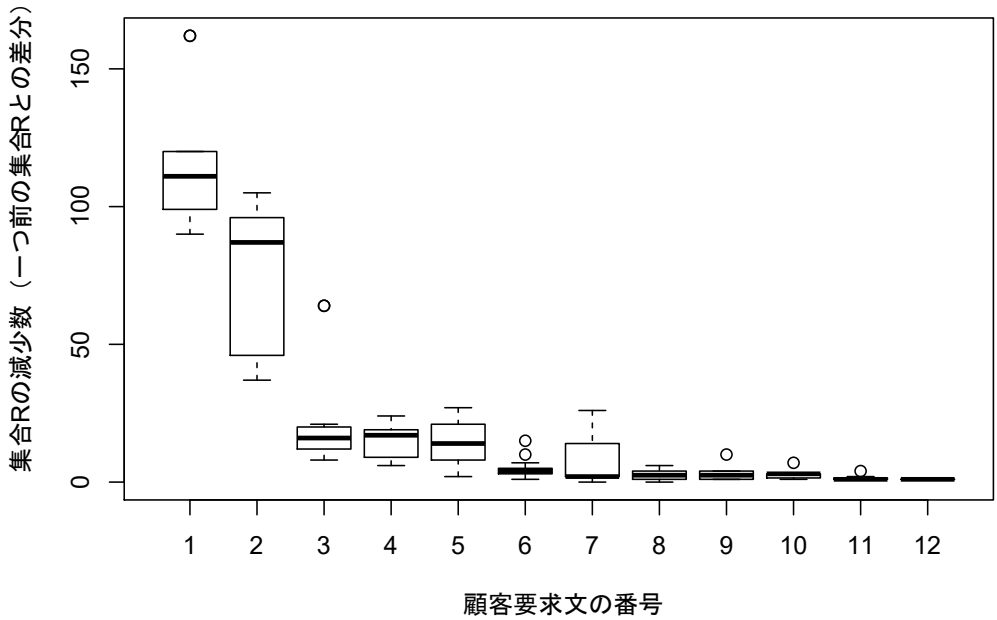


図 4.29　顧客要求の文番号ごとの減少数の箱ひげ図

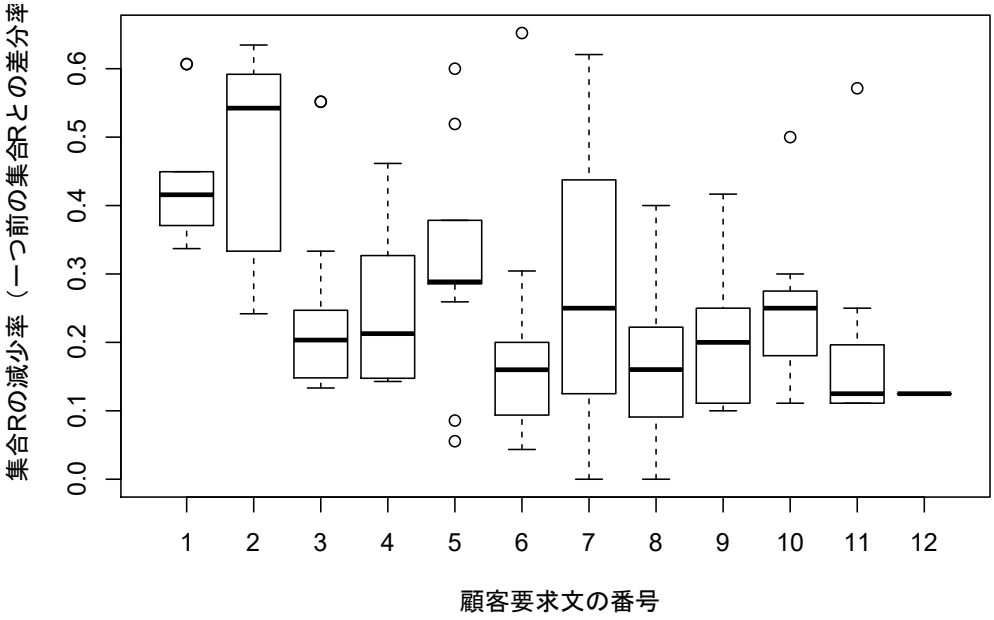


図 4.30　顧客要求の文番号ごとの減少率の箱ひげ図

4.3.6.3　文集合のもつ情報に関連する特徴量と fixation の関係

特徴量ごとの，単位時間当たりの fixation 数のヒストグラムを基に，正答／誤答での違いを分析したところ，減少数でピーク位置に違いがみられた（図 4.31）．要素数と減少率においては，正答／誤答での単位時間当たりの fixation 数に大きな違いはみられなかった．減少数では，正答の場合には誤答と比較して減少数が 15–25，95–100，110–115 の幅にてピークがみられた．これらの減少数をもつ顧客要求文にて情報量が大きいと考えられる．そこで，減少数と顧客要求の文番号との対応関係を調査した．

その結果，fixation 数のピークがみられた減少数に対応する顧客要求の文番号の多くは文番号 1–4 であり（表 4.9），それらの文にて情報量が多いことが示唆された．さらに，減少数 0–5 では誤答時の fixation 数が多く，これに対応する顧客要求文の後半の番号であり，後半の文は情報量が少ないことが示唆された．

したがって，情報量としては減少数が最も適していると考えられる．正答時には情報量の多い文（減少数 15–25，95–100，110–115 に対応する文）により多くの認知資源を費やしており，一方で情報量の少ない文（減少数 0–5 に対応する文）には少ない資源を費やしていることが示唆された．

以上を整理すると，顧客要求の必要十分性を判別する上での重要な情報量は，顧客要求文における減少数であり，正答時の閲読者は特定の値の減少数（15–25，95–100，110–115）に着目し閲読することで正答している可能性が示唆された．

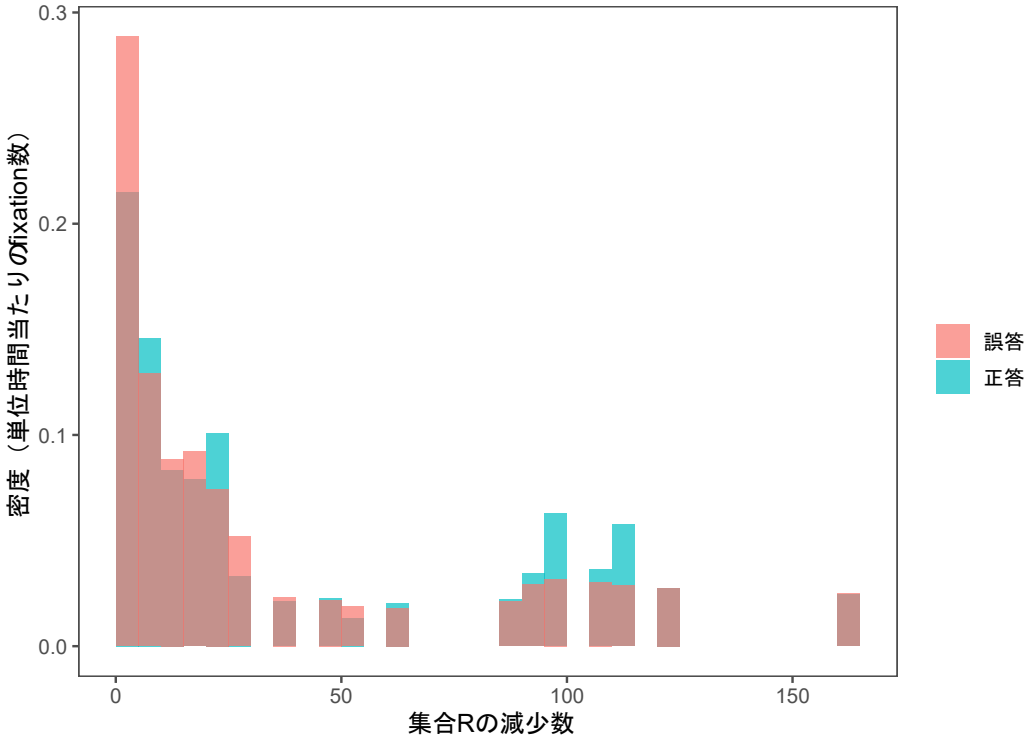


図 4.31　減少数ごとの単位時間当たりの fixation 数のヒストグラム

表 4.9　問題単位での顧客要求の文番号ごとの減少数の推移

問題ID	顧客要求の文番号											
	1	2	3	4	5	6	7	8	9	10	11	12
1-1	105	54	**16**	**19**	27	2	26	2	4	3	1	–
1-2	105	54	**16**	**19**	27	2	26	4	4	3	1	–
2-1	**114**	37	64	**17**	**21**	1	2	1	1	–	–	–
2-2	**114**	37	64	**17**	3	3	**18**	1	1	1	2	–
3-1	**111**	**99**	8	**22**	7	4	2	–	–	–	–	–
3-2	120	87	8	**24**	8	4	1	6	2	–	–	–
3-4	120	87	8	**24**	8	4	2	5	1	2	–	–
4-1	90	**96**	**20**	9	**15**	5	14	3	3	3	1	–
4-2	90	**96**	**20**	9	**15**	5	14	3	3	3	1	1
4-4	90	**96**	**20**	9	27	4	8	2	2	1	1	–
5-1	**99**	105	**21**	6	2	10	0	0	10	7	4	–
5-2	162	46	12	10	14	7	–	–	–	–	–	–
5-3	162	46	12	10	14	**15**	2	–	–	–	–	–

4.4 考察

4.4.1 本課題の有用性

本実験では，CD ゲームで得られた結果を基に，言語で記述された要件定義書と非言語的な顧客要求の突合能力を図る課題を作成した．分析 1 の結果，被験者によってタイプごとの正答率に偏りがみられ，その偏りは被験者の欠陥の検出バイアスを示すことが示唆された．

したがって，実際のソフトウェア開発において，本課題を適用して閲読者の欠陥検出バイアスを事前に把握することで，品質向上に寄与できると考えられる．閲読者の欠陥の検出バイアスを定量化することにより，最適な要員を配置できるようになる．合わせて，欠陥検出数の高低が要件定義書の品質に依存しているのか，それとも閲読者の検出バイアスに依存しているかの判別もできるようになる．本課題を要件定義書の閲読者に対して事前に実施することで，その特性を定量化し，品質の良い開発に繋げられる．

さらに，本課題を定期的に実施することで閲読能力の時系列的な変化を評価できる．しかし，本課題を実施するだけでは閲読能力の向上が図れないため，視線情報から得られた知見を基に能力向上のための訓練を検討する必要がある．

ただし，本課題では顧客要求自体は幾何学図形で与えられているため，顧客要求を正確に把握する能力は測れない．実際のソフトウェア開発においては，顧客要求を正確に把握することは難しく，顧客要求を誤って認識したことによってソフトウェア開発の品質が低下することもある．本課題では，閲読者の顧客要求に対する認識誤りは評価できないため，課題の正答率が高い閲読者が必ずしも品質の良いソフトウェア開発を導けるとは限らない．本課題の正答率に基づく閲読能力の他に，顧客要求を把握できているか否かは別途評価する必要がある．このような課題はあるものの，本課題の正答率が低い場合には閲読能力が低いと想定されるため，閲読能力を測る課題としては有用性があると考える．

4.4.2 閲読品質が高い人の特徴

分析 1～6 にて得られた結果を表 4.10 に整理し，それぞれの結果を基に閲読品質が高い人の特徴を議論する．

分析 1 と 2 では，問題のタイプごとの正答率を中心に分析し，タイプ N と NS の両方の正答率が高い人の閲読能力が高いと考えられ，かつそのような人は安静時と比較して課題時には瞳孔が散瞳し，かつ散瞳した状態が多数を占めていることが示された．さらに，タイプ N と NS での正答率の偏りは，欠陥の検出バイアスと関連していた．瞳孔は，認知負荷が高い場合や興味関心がある場合に散瞳することから(M A Just & Carpenter, 1993; Kahneman & Beatty, 1966)，正答率が高い人は課題中，常に高い負荷がかかっており，かつ課題時に

集中できていると考えられる．一方で，注意散漫やマインドワンダリングなどでは縮瞳することから(Grandchamp et al., 2014; Kang et al., 2014)，正答率が低い人は課題時に集中できていない可能性がある．先行研究では，課題実施時の瞳孔径の時系列変化を観察する研究が多く(Laeng et al., 2012; van der Wel & van Steenbergen, 2018)，本実験のように瞳孔径の分布に着目した研究は少ない．本研究では安静時と課題実施時の瞳孔径の分布の違いによって閲読能力を明らかにしているため，様々な課題に対しても有用性があると考えられる．

したがって，閲読能力の違いは，閲読に取り組む姿勢や態度に依存する可能性がある．瞳孔径の分布の特徴から，正答率が高い人は常に認知負荷がかかっており認知的な努力が高く，集中して課題に取り組んでいる傾向があると考えられる．

得られたこの知見を活用することで，実際の現場における閲読にも活用できる．1 頁単位での閲読の視線情報をリアルタイムで計測することで，閲読に対する姿勢（認知的な努力や集中度合い）が定量化でき，十分な閲読が行われたかという指標となりえる．

分析 3–6 では，正答／誤答の違いを中心に分析し，タイプ N と NS 共に正答時には後半より前半の要求文をよく注視しており，認知的資源の配分が文ごとに異なっていることが示唆された．Fixation は認知負荷を反映しているため，正答時には，単に顧客要求文を一律に同じ認知的資源の配分で読んでいるのではなく，資源を分配しながら文を読んでいると考えられる．正答時には，特に前半の文に対して資源を多く配分する傾向があると考えられる．戦略的に認知資源を配分したことによって，正答できた可能性がある．

文が持つ情報量が異なっていることを考慮すると，課題が持つ潜在的な構造を理解して，それに適合する戦略をとることが正答に繋がる可能性がある．顧客要求文から作成できる集合の要素の減少数に着目すると，文番号によって異なる構造を示す．正答時には，特定の減少数において fixation が増加している傾向があるため，集合の要素を特定するためにより重要な文に認知的資源を割いていると考えられる．

以上をまとめると，必要十分性を判別できる人は，課題の取り組む姿勢が異なっており，かつ課題の持つ潜在的な構造に適合した戦略をとっていることが視線から示唆された．

表 4.10　分析 1～6 にて得られた結果の整理

No	分析対象	得られた主な結果
分析1	正答率	タイプNとNSの正答率に関して
		・共に正答率が低い（難易度が高い）
		・それぞれの正答率の偏りは欠陥の指摘バイアスを反映
分析2	全試行の平均正答率	瞳孔径に関して
		・瞳孔径（zスコア）の最小値が予測には最も重要
		・正答率が高い場合には安静時と比較し右に偏った鋭いピークを持つ分布
分析3	正答／誤答	文ごとのfixationの傾向（後半より前半を注視する傾向）に関して
		・タイプNの正答時に，上記fixationの傾向
		・正答時には回答時間が短い場合に，上記fixationの傾向がより顕著
分析4	正答／誤答	文ごとのfixationの傾向（後半より前半を注視する傾向）に関して
	（誤認識を考慮）	・タイプNの正答時に，上記fixationの傾向
		・タイプNSの正答時に有意差はないが，上記fixationの傾向
		・正答時には回答時間が短い場合に，上記fixationの傾向がより顕著
分析5	正答／誤答	文ごとのfixation（後半の要求文）に関して
	（誤認識を考慮）	・後半の要求文における単位時間当たりのfixation数が予測には重要
分析6	文がもつ情報量	文の情報量（図形集合の減少数）とfixationに関して
		・正答時に特定の情報量をもつ文（その多くが前半の要求文）を注視する傾向

4.4.3　実際の閲読との比較

本節では，本実験で得られた知見における，実際のソフトウェア開発での適用可能性を議論する．実際のソフトウェア開発における要件定義書の閲読と，CD ゲームにより作成された課題の閲読の大きな違いは，幾何学図形で象徴化された顧客要求が提示されているか否かである．その違いにより，視線が影響を受けた可能性は否定できない．

実際のソフトウェア開発の閲読では，顧客要求は明示的には提示されないため，CD ゲームのように要件定義書と顧客要求のペアを交互に確認しながら閲読するとは限らない．閲読者は断片的な情報を基に顧客要求をイメージし，要件定義書を閲読している．閲読の過程で，断片的な情報である業務フローなどを基に要件定義書と突き合わせることは考えられるものの，常にそのような突合対象が用意されているわけではない．

そのため，実際のソフトウェア開発の閲読においては CD ゲームと比べて，異なる視線の特徴を持っている可能性は否定できない．例えば，CD ゲームでは要件定義書と顧客要求である幾何学図形の間を行き来する視線の特徴がみられたものの，実際の閲読では対応する顧客要求が明示的には存在しないため，そのような視線の特徴は現れない可能性がある．

しかし，本実験にて得られた瞳孔径の分布や文ごとの fixation の違いは，顧客要求の提示によって生じた特徴ではないと考えられる．なぜなら，文ごとの fixation は文が持つ潜在的な構造に適合した結果として違いが生じ，顧客要求の提示によって現れた視線の特徴ではないと考えられるからである．ただし，

要件定義書の文が持つ潜在的な構造は，文書ごとに異なっていると想定されるため，文ごとの fixation の偏りはそれに伴い変化すると想定される．瞳孔径の分布は，顧客要求である幾何学図形を見ている際の瞳孔径も含むものの，幾何学図形を見ることによって瞳孔径が影響を受けた可能性は低い．

したがって，本実験で得られた知見は実際のソフトウェア開発においても適用できる可能性があり，閲読能力の高い人の一つの特徴として採用できると考える．ただし，CD ゲームは要件定義工程を単純化したゲームであり，実際の要件定義書はより複雑な構造を含んでおり，幾何学図形で置き換えられないような要件が含まれる可能性もある．合わせて，顧客要求を幾何学図形で象徴化しているため，図形の認知能力にも依存している可能性も排除できない．さらに，CD ゲームの“ソフトウェア”は図形の集合として象徴化されており，その集合の各要素間の依存関係はないと想定されている．しかし，実際のソフトウェアはそうした独立したソフトウェアの部分だけではなく，互いに機能が依存した関係をもつ部分もある．このような依存関係がある場合，他より重要な要件に関する文が存在する．したがって，CD ゲームではそうした機能的な重要度の違いを表現できておらず，過度に単純化されており，この点により実際的な性質を導入した新たなゲームの開発は今後の課題といえる．これらの残された課題については今後の課題として，本課題の正答率と既存の指標の比較などにより，CD ゲームによって測ることができる閲読能力の精緻化を行いたい．

4.4.4 本実験の限界

分析 3 の回答時間の予測モデル構築にて，タイプ N と NS にて全ての顧客要求文を見ていないにも関わらず回答している場合がみられ，タイプ S や U と誤認識した可能性があった．タイプ N と NS は問題の性質上，全ての顧客要求文を見ることなしに正答できない．一方で，タイプ S や U の場合は，特定の顧客要求文において，作成できない正例が生じた時点で回答できる．結果として，タイプ N においては正答データにタイプ S や U と誤認識したデータが混入し，タイプ NS では，誤答データにタイプ S や U と誤認識したデータが混入した．

分析 5 においては，そのような誤認識データと想定されるデータを削除したものの，全ての誤認識データは削除できていない．なぜならば，誤認識データとして全ての顧客要求文を見ていないこと（いずれかの顧客要求文にて fixation がゼロを含む）を基準にしたものの，タイプ S や U と誤認識したとしても全ての文を一通り読む可能性があるためである．問題や人によって傾向は異なるものの，制限時間に余裕がある場合には回答の確認を兼ねて再度文を読み返す場合もある．

誤認識データを完全に排除できなかったことによって，結果に影響を与えた可能性を否定できない．例えば，タイプ NS で正答と誤答では単位時間当たりの fixation 数では有意差は得られなかった理由として，誤認識データを排除し

きれていないことが挙げられる．誤認識データがタイプ NS の誤答に混在していることで，誤答の場合でも後半より前半の要求文を注視するデータが混入してしまっていると考えられる．分析 5 の正答／誤答分類モデルにおいても，誤認識データが排除しきれなったことによって，精度が十分に向上しなかったと考えられる．

合わせて，要件定義書の閲読では基本的には全文を読んでいるため，全ての顧客要求文を読まずに回答できてしまうことは，課題として相応しくない可能性もある．問題において，「一致する」もしくは「一致しない」の二択で回答していたため，問題のタイプを特定することなく回答できてしまう形式となっていた．タイプ S や U においては，顧客要求文が正例（顧客要求）を満たさない問題であるもの，実際の要件定義においてはその意味合いが異なってくる．タイプ S では，特定の顧客要求が抜けているだけである一方で，タイプ U では，顧客要求が抜けているだけではなく余分な要求も定義されてしまっている．

そのため，今後の課題としては二択ではなく，四つのタイプを特定するような回答形式へと変更し，より実際の要件定義書の閲読に近づけ，かつタイプごとの視線情報を分析できるように検討する．

今回，タイプ N や NS において誤認識が見られたことは新たな知見であり，実際の要件定義書の閲読において同様の現象が起こっていると想定される．タイプ N や NS では，顧客要求文は正例の必要条件は満たしている．にもかかわらず，タイプ S や U と誤認識するということは，正例の必要条件を満たしていないと判断している．実際の要件定義書の閲読においては，必要条件を満たした定義書になっているにもかかわらず，誤った指摘が行われている可能性がある．そのような閲読者は，欠陥を検出したとしても誤った欠陥を検出している可能性があり，むしろ閲読によって品質の低下を招く恐れがある．回答の選択肢を工夫することで，誤認識の有無を分かれば，それを視線から予測できる可能性がある．この点については今後の課題とする．

第5章　総合議論

5.1　閲読能力と視線情報の関係

本論文では，閲読能力を測る指標として二つの指標を提案し，視線情報との関係性を分析した．実験 1 では，欠陥検出率の代わりに，信号検出理論を適用して欠陥の検出感度を導入し，視線情報との関連を議論した．実験 2 では，閲読を顧客要求と要件定義書の突合であると捉え，突合により必要十分性を判別する能力を測る課題を作成するゲームを考案し，象徴化した顧客要求に対する要件定義書と製品の対を作成した．続く実験 3 では，作成した課題を用いて，要件定義書の潜在的な欠陥のタイプを考慮して，閲読能力と視線情報との関連を議論した．

実験 1–3 の結果を基に，閲読における三つのプロセス（(1) 読み，(2) 構造

の理解，(3) 欠陥の検出／修正)，本実験にて導入した二つの指標（欠陥検出感度，必要十分性判別能力)，および視線情報との関連を整理する．

実験 1 では，三つのプロセスにおいて，閲読品質が高い人は，平常時よりも瞬目が増加しており，(2) (3) のプロセスに多くの認知的資源を費やすことが示唆された．欠陥検出感度が低い人は，平常時よりも瞬目が減少し，(1) 読みのプロセスが多数を占め，要件定義書を読んでいるだけの可能性が高い．一方，感度が高い人は，(2) (3) のプロセスに認知的資源を費やし，要件定義書から読み取れる顧客要求を基に適切な突合を行っていると考えられる．したがって，瞬目を用いることで，(1) 読みにとどまっているのか，(2) (3) のプロセスの認知処理がなされているのかを判別できることが示唆された．

しかし，(2) (3) のプロセスでの認知処理がなされていることは瞬目から読み取れるものの，そのプロセスにおける閲読品質の違いは測れない．そこで，実験 2 と 3 では閲読の (2) (3) のプロセスに焦点をあて，実験 2 にて顧客要求と要件定義書のペアを作成する実験を行い，実験 3 にてそれらのペアを用いてプロセス (2) (3) において必要十分性を図る能力と視線の関係を分析した．

実験 3 において，全体の正答率が高い閲読者は，瞳孔径が散瞳している割合が多数を占めており，課題中の認知的努力が大きく，集中して課題に取り組んでいたことが示唆された．つまり，(2) (3) のプロセスにおける閲読能力の違いは，瞳孔径に現れることが示された．さらに，課題に正答している場合には，顧客要求の文番号ごとの fixation 数が異なっており，後半より前半の要求文をよく注視する傾向がみられた．文番号ごとの情報量が異なることを考慮すると，正答時には文の潜在的な構造に適合する，認知的資源の配分を行っている可能性が示唆された．つまり，(2) (3) のプロセスにおける閲読能力の違いは，顧客要求文ごとの fixation の分布に現れることが示唆された．

以上をまとめたものを図 5.1 に示す．閲読のプロセス (1) にとどまっている場合は，閲読品質が低く，瞬目にその特徴が現れる．閲読のプロセス (2) (3) に至っている場合だとしても閲読品質には違いがあり，瞳孔径や文ごとの fixation 数の違いに特徴が現れることがわかった．要件定義書の閲読に視線を活用した研究は少なく(Sharafi, Guéhéneuc, et al., 2015; Sharafi, Shaffer, et al., 2015)，本研究にて閲読に影響を及ぼす視線の特徴を明らかにしたことは有用であると考える．

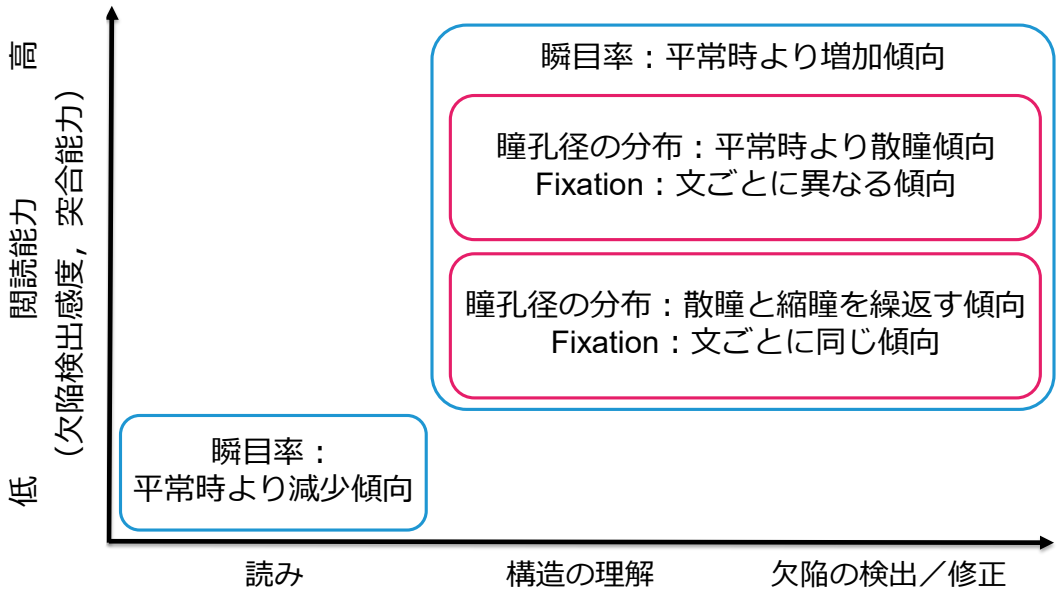


図 5.1　閲読のプロセスと閲読能力における視線の特徴

5.2　本研究の貢献

本研究の貢献の一つは，比較的取得が簡便な瞬目を用いて，閲読における読みのプロセスにとどまっているのか否かが判別できるという基礎的な知見を示したことである．序論でも述べたように，瞬目は専用のアイトラッカを用いることなく，カメラ映像から取得できる．そのため，モニタ上での閲読であればカメラを用いることで，瞬目の情報が得られ，本論文の知見が活用できる．したがって，本知見をすぐに実際の業務の場へ活用することが検討できる．実験 1 にて，欠陥検出感度が低く平常時より瞬目が検証した人が一定数いたことから，実際のソフトウェア開発においても読みにとどまっている閲読者がいると想定される．実験では制限時間は設定されていなかった．実際の開発では限られた時間内での閲読となり，読みのプロセスが優位となっている場面が多い可能性がある．したがって，本知見を適用することによって，ソフトウェア開発の品質向上へ貢献できる．

二つ目の貢献は，閲読のプロセス（2）（3）において，閲読能力を測る CD ゲームを作成し，潜在的な欠陥のタイプごとの閲読過程の違いを示したことである．実験 2 にて，ソフトウェアを幾何学図形で象徴化することで，ソフトウェアと要件定義書のペアを作成し，両者を突合して必要十分性判別する課題を作成した．要件定義書の閲読能力を測る課題はほとんど存在しないため，本課題は有用性が高いと考えられる．閲読能力の他にも，閲読者の欠陥の指摘傾向を定量化できる可能性が示唆された．タイプごとの正答率を考慮することで，閲読者が欠陥を指摘する傾向あるのか否かが判別できる．合わせて，期間を空け

て本課題を複数回実施することで，閲読者の能力向上などの時系列変化をとらえることもできる．本課題によって個人の閲読能力を定量化することで，ソフトウェア開発における最適な閲読者の配置や教育に活かすことができ，全体の品質向上に寄与できる．

三つ目の貢献は，開発した課題における閲読能力に影響を及ぼす視線の特徴を明らかにし，閲読能力をリアルタイムで推定できる可能性を示唆した点である．閲読能力に影響を及ぼす要因として，瞳孔径の平常時からの変化と文ごとの fixation があることが明らかになった．さらに，この瞳孔径の特徴は，本課題だけではなく多くの課題やテストなどでも活用ができると考える．安静時の瞳孔径の分布が分かっていれば，課題時の瞳孔径の分布を用いて，リアルタイムで被験者の集中度合いや認知的な努力などを定量化でき，学習者への最適なフィードバックに繋げられる可能性がある．さらに，文ごとの fixation の違いを活用することで,要件定義書などの文章の改良にも応用できる可能性がある．つまり，閲読者が注意を向ける傾向がある文章もしくは閲読者の注視の傾向に応じた，文章校正とすることで閲読のしやすい要件定義書を作成できると考える．実験 3 において，問題ごとの正答率の違いはなかったものの，インタビューにて解きやすい問題とそうでない問題があったという意見があり，人が理解しやすい文章構成は暗黙的に存在する．人の視線の傾向を分析することで，理解しやすい文章構成の提案にも繋げられる．

知識科学への貢献は，本研究の知見が顧客要求という暗黙知を形式知化する認知プロセスを解明する足掛かりとなりえることである．ここでの暗黙知は，野中ら(1996)が提唱する特定の状況に関する個人的知識で，他者に伝えるのが困難な類の知識を示す.ソフトウェア開発において,顧客要求は暗黙的であり，要件定義書の作成では暗黙知を形式知化していると考えられる．閲読においても顧客要求が明示されないため，暗黙的な顧客要求を基に，形式知化された要件定義書を閲読している．暗黙的な顧客要求を実験参加者に伝えることが難しいため，実験 2 では顧客要求を幾何学図形で象徴化した．幾何学図形そのものは形式化されたものであるものの，幾何学図形がどのような顧客要求に基づいているのかは暗黙的である．つまり，ある顧客要求を基に幾何学図形の集合が作成されており，幾何学図形を基にして一意な顧客要求は対応付かないため，顧客要求は暗黙的である.したがって,実験 3 の参加者にとって幾何学図形は，図 3.1 において断片的な情報を基に顧客要求をイメージした結果に該当すると想定される．閲読において，被験者は要件定義書と顧客要求を突合可能な状態に変換しており，幾何学図形で象徴化された暗黙知を形式知化していると想定される．実験 3 における回答時間の予測モデルや正答／誤答の分類モデルにおいて，正例の図形を見ている fixation の重要度も上位に位置しているため，暗黙的な顧客要求を象徴化した図形を基にして形式知化された情報も重要であると考えられる．実験 3 における正答／誤答の違いは，文における fixation を

対象として分析したものの，図形ごとの fixation の違いも重要である可能性がある．つまり，閲読能力の高い閲読者は，単に要件定義書を基にして図形へと変換しているだけではなく，図形を基にして文と比較できるような状態に変換している可能性がある．後者のプロセスが暗黙知を形式知化するプロセスにあたる．しかし，実験 3 では図形のパターンが全ての問題で同じではないため，統計的な分析が困難であった．実験の被験者増やして，図形パターンごとの fixation などの視線情報を活用することで，暗黙知を形式知化する認知プロセスを明らかにできると考えられる．合わせて，品質の良い閲読に必要な能力も暗黙的であり，本知見を活用することで良い閲読者が持つ暗黙知自体を形式知化でき，閲読者の教育に活かすことができる．

5.3　今後の課題

最も早急に対処すべき課題は，第 4 章 実験 3 の本実験の限界で述べたように，問題のタイプを誤認識している人が一定数いたことへの対処である．実験において，タイプ NS か否かを判別しているかを回答する形式となっていたため，被験者がどのタイプと認識していたのかを知ることはできなかった．合わせて，タイプによっては顧客要求文をすべて読まなくても回答できる構造となっており，実際の要件定義書の閲読と異なってしまっている可能性もある．したがって，誤認識によって視線が影響を受けた可能性を否定できない．そのため,被験者がどのタイプの問題と認識しているかを回答から読み取れるように，回答の選択肢を変更して再度実験を行うことで，タイプごとの視線の正確な分析を行うことを検討する．合わせて，構築したモデルにおいて実用可能な精度へより近づけるためにも，被験者を増やして分類モデルの精度向上を目指す．

上記対応と合わせて，第 4 章の分析 6 で実施した文のもつ情報量と視線の関係について詳細な分析を試みる．第 4 章の結果から，閲読能力の高い人は複数の文を均一に読んでいるのではなく，戦略的に認知資源を配分して文を読んでいる可能性が示唆された．その際，文によって作成される図形集合の減少数を文の情報量として着目し，特定の情報量を持つ文をより認知資源を費やす傾向がみられた．CD ゲームにおいては，作成できる文（内包的定義）と図形（外延的定義）はそれぞれ有限であり，それぞれ有限な空間を持つ．さらに，文と図形には対応関係が存在し，一つの文に対して一つ以上の図形が対応づく．課題において，被験者は文を読み進めながら，それに対応する図形の関係性を確認し,必要十分に対応付いているのかを判断している.文を読み進めるに応じて，対応づく図形が絞り込まれていく.その認知プロセスは明らかになっておらず，なぜ文による図形の減少数に着目すると正答できる可能性があるのかはわかっていない．このような対応付けは人の理解とも関連している可能性があり，認知プロセスを明らかにすることは重要である．今後詳細な分析を行い，文と図形の対応関係と，人の認知との関係を視線情報により明らかにしたい．

実験 3 の瞳孔径については個人差を考慮した分析であるものの，それ以外の分析のおいては個人差が考慮されていないため，個人差を考慮した分析やモデル構築を行いたい. 視線研究では, 個人差を考慮することが重要であるものの，多くの研究では個人差について考慮されていない(平田他, 2011). 本論文においては，実験 3 の瞳孔径の分析については平常時との比較を行っており，個人差が考慮されている．しかし，実験 1 における瞬目や，各モデル構築に用いた視線の特徴量では個人差は考慮されていない．精度の高いモデルや，個人に最適化されたモデルを作成するためには，個人差を考慮したモデル構築を行うことが重要である．さらに，個々の視線情報だけではなく閲読スタイルも個人によって異なる可能性がある．ソフトウェアの内包的定義と外延的定義を基に，突合可能な状態に変換する際に，内包的定義を基に外延的定義に変換して突合する場合や，その逆の場合などいくつか変換の種類が考えられる．その変換の種類によって視線の傾向が異なる可能性があり，視線情報はそのような個人の閲読スタイルを明らかにできる可能性がある. 本研究で得られた知見を基にして，より個人差を考慮したモデル構築や分析を行いたい.

合わせて，より実際の要件定義書の閲読の環境に近づけるように，実験方法を検討する．要件定義書はモニタ上に提示され，視線計測の精度を高めるために頭をできるだけ動かさないように指示した．あご台を用いず，閲読方法に制限はなかったため，比較的実際の閲読の状況を示していると考えられる．しかし，実際のソフトウェア開発では，複数の文書を行き来することや，紙とモニタ上の文書が混在して閲読することも想定される．そのような状況を再現するために，装着型のアイトラッカの利用し，閲読者に負担がない自然な形での閲読における視線情報の収集を検討する.

将来的には，本論文にて得られた知見を活用し，閲読だけではなく読解力との関係性についても調査したい．要件定義書の閲読では，要件定義書を読んで内容を理解し，顧客要求を満たすかを確認する課題であり，文章理解と関係があると考えられる．つまり，閲読能力とは文章の読解力と関連し，閲読能力の高い人は読解力も高い可能性がある．その場合，本論文で得られた良い閲読者の特徴をさらに分析することで，読解力の高低に影響を及ぼす要因が特定できる．したがって，視線を活用することで読解力の予測ができ，教育への応用が期待できる．学習者に対して，その読解度に応じたリアルタイムなフィードバックなどの適切な教育も可能となる.

5.4　おわりに

本論文は，ソフトウェア開発における要件定義書の閲読能力を測る指標を新たに定義し，視線の関係を分析することで閲読能力に影響を及ぼす要因を明らかにした．閲読能力を測る指標として，欠陥の検出感度と突合における必要十分性判別能力の二つを定義し，視線との関連を調査した．その結果，閲読のプ

ロセスと関連して，瞬目，瞳孔径，文ごとの fixation が関連していることがわかった．このように視線情報は，要件定義書の閲読に影響を及ぼし，閲読能力を特徴付ける重要な要因であることが示された．閲読時の視線情報を用いることで，定義した二つの閲読能力を予測でき，リアルタイムでその品質を評価できることも示唆された．2 章と 4 章の本実験の限界で述べたように，実験の方法には課題が残されているため，実験の改良を行い，閲読能力と視線の関係を今度も調査していきたい．合わせて，実際のソフトウェア開発へ本知見を適用し，実開発における有用性も検討していきたい．

参考文献

Augereau, O., Kunze, K., Fujiyoshi, H., & Kise, K. (2016). Estimation of english skill with a mobile eye tracker. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16*, 1777–1781. https://doi.org/10.1145/2968219.2968275

Bednarik, R. (2012). Expertise-dependent visual attention strategies develop over time during debugging with multiple code representations. *International Journal of Human-Computer Studies*, *70*(2), 143–155. https://doi.org/https://doi.org/10.1016/j.ijhcs.2011.09.003

Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., & Albanese, A. (1997). Analysis of blink rate patterns in normal subjects. *Movement Disorders*, *12*(6), 1028–1034. https://doi.org/10.1002/mds.870120629

Bernardez, B., Genero, M., Duran, A., & Toro, M. (2004). A controlled experiment for evaluating a metric-based reading technique for requirements inspection. *10th International Symposium on Software Metrics, 2004. Proceedings.*, 257–268. https://doi.org/10.1109/METRIC.2004.1357908

Binkley, D., Davis, M., Lawrie, D., Maletic, J. I., Morrell, C., & Sharif, B. (2013). The impact of identifier style on effort and comprehension. *Empirical Software Engineering*, *18*(2), 219–276. https://doi.org/10.1007/s10664-012-9201-4

Bixler, R., Blanchard, N., Garrison, L., & D'Mello, S. (2015). Automatic Detection of Mind Wandering During Reading Using Gaze and Physiology. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15*, 299–306. https://doi.org/10.1145/2818346.2820742

Bixler, R., & D'Mello, S. (2015). Automatic Gaze-Based Detection of Mind Wandering with Metacognitive Awareness. In F. Ricci, K. Bontcheva, O. Conlan, & S. Lawless (Eds.), *User Modeling, Adaptation and Personalization* (pp. 31–43). Springer International Publishing.

Boehm, B, & Basili, V. R. (2001). Top 10 list [software development]. *Computer*, *34*(1), 135–137. https://doi.org/10.1109/2.962984

Boehm, Barry. (2004). *Balancing Agility and Discipline: A Guide for the Perplexed BT - Software Engineering Research and Applications* (C. V Ramamoorthy, R. Lee, & K. W. Lee (eds.); p. 1). Springer Berlin Heidelberg.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Busjahn, T., Bednarik, R., Begel, A., Crosby, M., Paterson, J. H., Schulte, C., Sharif, B., & Tamm, S. (2015). Eye Movements in Code Reading: Relaxing the Linear Order. *2015 IEEE 23rd International Conference on Program Comprehension*, 255–265. https://doi.org/10.1109/ICPC.2015.36

Campbell, C. S., & Maglio, P. P. (2001). A Robust Algorithm for Reading Detection. *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, 1–7. https://doi.org/10.1145/971478.971503

Cantone, G., Colasanti, L., Abdulnabi, Z. A., Lomartire, A., & Calavaro, G. (2003). *Evaluating Checklist-Based and Use-Case-Driven Reading Techniques as Applied to Software Analysis and Design UML Artifacts BT - Empirical Methods and Studies in Software Engineering: Experiences from ESERNET* (R. Conradi & A. I. Wang (eds.); pp. 142–165). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-45143-3_9

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A Library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–39. https://doi.org/10.1145/1961189.1961199

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Chen, S., Epps, J., Ruiz, N., & Chen, F. (2011). Eye Activity as a Measure of Human Mental Effort in HCI. *Proceedings of the 16th International Conference on Intelligent User Interfaces*, 315–318. https://doi.org/10.1145/1943403.1943454

Cho, P., Sheng, C., Chan, C., Lee, R., & Tam, J. (2000). Baseline blink rates and the effect of visual task difficulty and position of gaze. *Current Eye Research*, *20*(1), 64–70. https://doi.org/10.1076/0271-3683(200001)20:1;1-H;FT064

Ciolkowski, M., Laitenberger, O., & Biffl, S. (2003). Software reviews, the state of the practice. *IEEE Software*, *20*(6), 46–51. https://doi.org/10.1109/MS.2003.1241366

Crosby, M., Scholtz, J., & Wiedenbeck, S. (2002). The Roles Beacons Play in Comprehension for Novice and Expert Programmers. *PPIG*.

Cruz, A. A. V., Garcia, D. M., Pinto, C. T., & Cechetti, S. P. (2011). Spontaneous Eyeblink Activity. *The Ocular Surface*, *9*(1), 29–41. https://doi.org/10.1016/S1542-0124(11)70007-6

D'Mello, S., Kopp, K., Bixler, R. E., & Bosch, N. (2016). Attending to Attention. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, 1661–1669. https://doi.org/10.1145/2851581.2892329

独立行政法人情報処理推進機構（IPA）社会基盤センター. (2018). ソフトウェア開発データ白書2018-2019. 独立行政法人情報処理推進機構.

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology*, *50*(9), 833–859. https://doi.org/https://doi.org/10.1016/j.infsof.2008.01.006

Eraslan, S., Yesilada, Y., & Harper, S. (2016). Scanpath Trend Analysis on Web Pages: Clustering Eye Tracking Scanpaths. *ACM Trans. Web*, *10*(4). https://doi.org/10.1145/2970818

Fritz, T., Begel, A., Müller, S. C., Yigit-Elliott, S., & Züger, M. (2014). Using Psycho-Physiological Measures to Assess Task Difficulty in Software Development. *Proceedings of the 36th International Conference on Software Engineering*, 402–413. https://doi.org/10.1145/2568225.2568266

Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., & Menke, L. (2016). Which Eye Tracker Is Right for Your Research? Performance Evaluation of Several Cost Variant Eye Trackers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*, 1240–1244. https://doi.org/10.1177/1541931213601289

Goldberg, J., & Kotval, X. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, *24*, 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7

Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind wandering. *Frontiers in Psychology*, *5*, 31. https://doi.org/10.3389/fpsyg.2014.00031

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. In *Signal detection theory and psychophysics.* John Wiley.

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, *11*(1), 1–21. https://doi.org/10.1080/00401706.1969.10490657

Halling, M., Biffl, S., Grechenig, T., & Kohle, M. (2001). Using reading techniques to focus inspection performance. *Proceedings 27th EUROMICRO Conference. 2001: A Net Odyssey*, 248–257. https://doi.org/10.1109/EURMIC.2001.952461

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349–350.

平田乃美・本多麻子・田多英興. (2011). 内因性瞬目研究小史：19世紀末から21世紀に向けて. 白鴎大学教育学部論集 = Hakuoh Journal of the Faculty of Education, 5(1), 115–154.

Jaccard, P. (1912). THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. *New Phytologist*, *11*(2), 37–50. https://doi.org/https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Jacob, R. J. K. (1995). Eye Tracking in Advanced Interface Design. In W. Barfield & T. A. Furness (Eds.), *Virtual Environments and Advanced Interface Design* (pp. 258–288). Oxford University Press, Inc.

Just, M A, & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *47*(2), 310—339. https://doi.org/10.1037/h0078820

Just, Marcel A, & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329.

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kang, O. E., Huffer, K. E., & Wheatley, T. P. (2014). Pupil Dilation Dynamics Track Attention to High-Level Information. *PLOS ONE*, *9*(8), e102463.

Karson, C. N., Berman, K. F., Donnelly, E. F., Mendelson, W. B., Kleinman, J. E., & Wyatt, R. J. (1981). Speaking, thinking, and blinking. *Psychiatry Research*, *5*(3), 243–246. https://doi.org/10.1016/0165-1781(81)90070-6

片岡雅憲・小原由紀夫・光藤昭男. (2017). アジャイル開発への道案内. 近代科学社.

Knorr, A. (1928). *Beiträge zut Kenntnis des Lidschlages unt. normalen u. pathol. Verhältnissen*. Verlag nicht ermittelbar.

Konishi, M., Brown, K., Battaglini, L., & Smallwood, J. (2017). When attention wanders: Pupillometric signatures of fluctuations in external attention. *Cognition*, *168*, 16–26. https://doi.org/https://doi.org/10.1016/j.cognition.2017.06.006

Kullback, S. (1997). *Information theory and statistics*. Courier Corporation.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. https://doi.org/10.1177/1745691611427305

Ledger, H. (2013). The effect cognitive load has on eye blinking. *The Plymouth Student Scientist*, *6*(1), 206–223.

Li, J., Ngai, G., Leong, H. V, & Chan, S. C. F. (2016). Your Eye Tells How Well

You Comprehend. *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, *2*, 503–508. https://doi.org/10.1109/COMPSAC.2016.220

Lima, C., Id, S., Id, O. A., & Id, K. K. (2018). *Estimation of reading subjective understanding based on eye gaze analysis*. 1–16.

松川文一. (2002). UMLで記述された設計仕様書を対象としたレビュー手法CBRとPBRの比較評価実験. オブジェクト指向最前線2002.

宮尾克・石原伸哉・近藤高明・榊原久孝・古田真司・山中克己・山田信也. (1992). VDTの極性および視標の大きさが瞳孔面積に与える影響. 日本衛生学雑誌, 47(2), 658–662. https://doi.org/10.1265/jjh.47.658

西山潤平. (2011). 瞳孔ゆらぎ・眼球運動によるヒトの眠気予兆検出に関する研究. 中部大学.

西山潤平・谷田公二・楠見昌司・平田豊. (2008). 瞳孔ゆらぎを指標とした覚醒度状態評価. 生体医工学, 46(2), 212–217. https://doi.org/10.11239/jsmbe.46.212

野中郁次郎・竹内弘高・梅本勝博. (1996). 知識創造企業. 東洋経済新報社.野中誠. (2004). 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,ソフトウェア工学研究会報告, 146, 25–31.

Okoso, A., Toyama, T., Kunze, K., Folz, J., Liwicki, M., & Kise, K. (2015). Towards Extraction of Subjective Reading Incomprehension: Analysis of Eye Gaze Features. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1325–1330. https://doi.org/10.1145/2702613.2732896

大社綾乃・Augereau, O.・黄瀬浩一. (2017). 視点情報と単語の出現頻度を用いた主観的高難易度単語の推定. 電子情報通信学会技術研究報告, パターン認識メディア理解研究会 (pp. 187–192).

Ponder, E., & Kennedy, W. P. (1927). ON THE ACT OF BLINKING. *Quarterly Journal of Experimental Physiology*, *18*(2), 89–110. https://doi.org/10.1113/expphysiol.1927.sp000433

Poole, A., & Ball, L. (2006). Eye tracking in human-computer interaction and usability research: Current status and future prospects. In *Encyclopedia of Human Computer Interaction* (pp. 211–219).

Porras, G., & Guéhéneuc, Y.-G. (2010). An empirical study on the efficiency of different design pattern representations in UML class diagrams. *Empirical Software Engineering*, *15*, 493–522. https://doi.org/10.1007/s10664-009-9125-9

Porter, A., & Votta, L. (1998). Comparing Detection Methods For Software Requirements Inspections: A Replication Using Professional Subjects.

Empirical Software Engineering, *3*(4), 355–379. https://doi.org/10.1023/A:1009776104355

Records, R. E. (1979). Eyebrows and eyelids. *Physiology of the Human Eye and Visual System*, *13*.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., & Sabeti, P. C. (2011). Detecting Novel Associations in Large Data Sets. *Science*, *334*(6062), 1518–1524. https://doi.org/10.1126/science.1205438

Royce, W. W. (1970). Managing the development of large software systems. proceedings of IEEE WESCON. *Los Angeles*, 328–388.

斉藤功樹. (2019). 文書レビュー時の視線計測情報によるレビュー品質の解析（特集 生命科学). Unisys 技報: Unisys Technology Review, 38(4), 283–294.

斉藤功樹・土肥拓生. (2018). アイトラッキングを利用した，次世代の要件定義書レビュー評価手法. 日本認知科学会第35回大会論文集, 661–668.

Sanches, C. L., Kise, K., & Augereau, O. (2017). Japanese reading objective understanding estimation by eye gaze analysis. *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers on - UbiComp '17*, 121–124. https://doi.org/10.1145/3123024.3123092

Sawyer, P., & Kotonya, G. (2001). Software requirements. *SWEBOK*, 9.

Sharafi, Z., Guéhéneuc, Y.-G., & Soh, Z. (2015). A Systematic Literature Review on the Usage of Eye-tracking in Software Engineering. *Information and Software Technology*, *67*(C), 79–107. https://doi.org/10.1016/j.infsof.2015.06.008

Sharafi, Z., Shaffer, T., Sharif, B., & Gueheneuc, Y.-G. (2015). Eye-Tracking Metrics in Software Engineering. *2015 Asia-Pacific Software Engineering Conference (APSEC)*, 96–103. https://doi.org/10.1109/APSEC.2015.53

Sharif, B., Jetty, G., Aponte, J., & Parra, E. (2013). An empirical study assessing the effect of seeit 3D on comprehension. *2013 First IEEE Working Conference on Software Visualization (VISSOFT)*, 1–10. https://doi.org/10.1109/VISSOFT.2013.6650519

Sharif, B., & Maletic, J. I. (2010). An eye tracking study on the effects of layout in understanding the role of design patterns. *2010 IEEE International Conference on Software Maintenance*, 1–10. https://doi.org/10.1109/ICSM.2010.5609582

Sibert, L. E., & Jacob, R. J. K. (2000). Evaluation of Eye Gaze Interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems, 281–288. https://doi.org/10.1145/332040.332445

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated Learning: A Framework for the Emergence of Language. *Artificial Life*, *9*(4), 371–386. https://doi.org/10.1162/106454603322694825

Tada, H. (1986). Eyeblink rates as a function of the interest value of video stimuli. *Tohoku Psychologica Folia*, *45*, 107–113.

田多英興. (1986). 課題および心的負荷の違いと自発性瞬目. 日本心理学会第50回大会発表論文, 63.

田多英興・山田冨美雄・福田恭介. (1991). まばたきの心理学：瞬目行動の研究を総括する. 北大路書房.

Tecce, J. . (1989). Eyeblinks and psychological functions：A two-process model. *Psychophysiology (Abstract)*, *26*, 5–6.

Thelin, T., Andersson, C., Runeson, P., & Dzamashvili-Fogelstrom, N. (2004). A replicated experiment of usage-based and checklist-based reading. *10th International Symposium on Software Metrics, 2004. Proceedings.*, 246–256. https://doi.org/10.1109/METRIC.2004.1357907

Thelin, T., Runeson, P., & Wohlin, C. (2003). An experimental comparison of usage-based and checklist-based reading. *IEEE Transactions on Software Engineering*, *29*(8), 687–704. https://doi.org/10.1109/TSE.2003.1223644

Uema, Y., & Inoue, K. (2017). JINS MEME algorithm for estimation and tracking of concentration of users. *UbiComp/ISWC 2017 - Adjunct Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 297–300. https://doi.org/10.1145/3123024.3123189

上野秀剛・中道上・井垣宏・門田暁人・中村匡秀・松本健一. (2005). プログラマの視線を用いたレビュープロセスの分析. 信学技報ss2005 12-17, 105(128), 21–26.

Uwano, H. (2011). Measuring and Characterizing Eye Movements for Performance Evaluation of Software Review. *Eye Movement: Theory, Interpretation, and Disorders*.

Uwano, H., Nakamura, M., Monden, A., & Matsumoto, K. (2007). Exploiting Eye Movements for Evaluating Reviewer's Performance in Software Review. *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, *E90-A*(10), 317–328. https://doi.org/10.1093/ietfec/e90-a.10.2290

Uwano, H., Nakamura, M., Monden, A., & Matsumoto, K. (2006). Analyzing Individual Performance of Source Code Review Using Reviewers' Eye

Movement. *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*, 133–140. https://doi.org/10.1145/1117309.1117357

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, *62*, 1–10. https://doi.org/https://doi.org/10.1016/j.dss.2014.02.007

Wong, Y.-K. (2009). Software quality and group performance. *AI & SOCIETY*, *23*(4), 559–573. https://doi.org/10.1007/s00146-007-0174-6

Wong, Y.-K. (2003). An exploratory study of software review in practice. *PICMET '03: Portland International Conference on Management of Engineering and Technology Technology Management for Reshaping the World, 2003.*, 301–308. https://doi.org/10.1109/PICMET.2003.1222807

Yoshimura, K., Kise, K., & Kunze, K. (2015). The eye as the window of the language ability: Estimation of English skills by analyzing eye movement while reading documents. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, *2015-Novem*, 251–255. https://doi.org/10.1109/ICDAR.2015.7333762

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring Cognitive Load Using Eye Tracking Technology in Visual Computing. *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 78–85. https://doi.org/10.1145/2993901.2993908

付録

付録 A 実験 1 のデータ

付録 A.1 被験者ごとの d-prime と属性情報

No	d.prime	性別	年代	要件定義書レビュー経験	文書レビュー頻度
1	-0.84	男	40	0	ほとんどない
2	-0.76	男	30	0	1月に1度程度
3	-0.28	男	30	2	1月に1度程度
4	-0.28	男	50	8	ほとんどない
5	-0.27	男	30	15	1週間に1度程度
6	-0.22	男	30	0	ほとんどない
7	-0.20	男	50	8	1週間に1度程度
8	-0.09	男	40	10	1月に1度程度
9	-0.08	男	30	2	1月に1度程度
10	0.02	男	50	3	2週間に1度程度
11	0.16	男	30	0	1月に1度程度
12	0.16	男	50	10	2週間に1度程度
13	0.19	女	30	0	1月に1度程度
14	0.24	男	40	0	ほとんどない
15	0.31	女	30	0	1月に1度程度
16	0.33	男	30	0	ほとんどない
17	0.34	男	50	4	ほとんどない
18	0.84	男	40	3	1週間に1度程度
19	0.91	女	50	0	ほとんどない

付録 A.2 47 個の特徴量

特徴量			詳細
fixation	duration	minimum	fixationの持続時間の最小
		maximum	fixationの持続時間の最大
		mean	fixationの持続時間の平均
		median	fixationの持続時間の中央値
		standard deviation	fixationの持続時間の標準偏差
		skew	fixationの持続時間の歪度
		kurtosis	fixationの持続時間の尖度
		range	fixationの持続時間のレンジ
		dispersion	fixation中の視点の分散
saccade	distance	minimum	saccadeの距離の最小
		maximum	saccadeの距離の最大
		mean	saccadeの距離の平均
		median	saccadeの距離の中央値
		standard deviation	saccadeの距離の標準偏差
		skew	saccadeの距離の歪度
		kurtosis	saccadeの距離の尖度
		range	saccadeの距離のレンジ
	duration	minimum	saccadeの持続時間の最小
		maximum	saccadeの持続時間の最大
		mean	saccadeの持続時間の平均
		median	saccadeの持続時間の中央値
		standard deviation	saccadeの持続時間の標準偏差
		skew	saccadeの持続時間の歪度
		kurtosis	saccadeの持続時間の尖度
		range	saccadeの持続時間のレンジ
	angle	minimum	saccadeの角度の最小
		maximum	saccadeの角度の最大
		mean	saccadeの角度の平均
		median	saccadeの角度の中央値
		standard deviation	saccadeの角度の標準偏差
		skew	saccadeの角度の歪度
		kurtosis	saccadeの角度の尖度
		range	saccadeの角度のレンジ
	other	count	saccadeの回数
		proportion of horizontal	水平なsaccadeの割合(水平軸から30度未満の角度)
pupil	diameter (z-scoreで標準化)	minimum	pupilの直径の最小
		maximum	pupilの直径の最大
		mean	pupilの直径の平均
		median	pupilの直径の中央値
		standard deviation	pupilの直径の標準偏差
		skew	pupilの直径の歪度
		kurtosis	pupilの直径の尖度
		range	pupilの直径のレンジ
blilnk		count	blinkの回数
		count / reviewtime	blinkの回数の割合 (瞬目率)
		time / reviewtime	blinkの時間の割合
other		fixation duration / saccade duration ratio	fixationの持続時間/saccadeの持続時間の比率

付録 A.3 被験者のノイズに対する指摘欠陥分類と件数

No	d-prime	抜け漏れ	誤り		その他				
			曖昧	誤ったセクション	標準化	校正	確認／提案	冗長	不明
1	-0.84	4	3	0	1	2	5	2	9
2	-0.76	37	35	0	2	11	4	0	1
3	-0.28	1	3	0	1	0	10	0	0
4	-0.28	0	2	0	1	0	4	0	0
5	-0.27	0	2	0	1	18	0	0	0
6	-0.22	3	1	1	5	5	1	1	1
7	-0.20	8	4	9	6	3	4	2	7
8	-0.09	0	0	1	2	1	1	0	0
9	-0.08	0	1	0	2	0	0	0	11
10	0.02	0	0	0	0	0	0	0	4
11	0.16	0	0	0	0	2	0	0	1
12	0.16	0	2	0	0	1	0	0	0
13	0.19	0	1	0	0	1	0	2	2
14	0.24	0	1	0	1	3	1	0	6
15	0.31	0	0	0	0	1	0	0	0
16	0.33	7	3	0	1	1	4	0	1
17	0.34	0	0	0	0	1	0	0	1
18	0.84	1	1	0	1	5	0	0	0
19	0.91	4	1	0	0	0	0	0	2

付録 B 実験 3 のデータ

付録 B.1 回答時間の予測モデル構築における 24 個の特徴量の重要度

順位	特徴量（単位時間当たりのfixation数）	重要度
1	7番目の顧客要求文	138198
2	6番目の顧客要求文	135054
3	1番目の正例	73085
4	9番目の顧客要求文	61780
5	3番目の正例	57620
6	5番目の顧客要求文	55317
7	4番目の顧客要求文	54304
8	3番目の顧客要求文	50702
9	2番目の正例	49143
10	5番目の正例	46338
11	1番目の負例	42198
12	4番目の正例	41440
13	8番目の顧客要求文	41414
14	6番目の正例	40388
15	2番目の負例	36254
16	1番目の顧客要求文	31589
17	3番目の負例	29810
18	2番目の顧客要求文	29531
19	10番目の顧客要求文	22434
20	5番目の負例	16683
21	4番目の負例	15603
22	6番目の負例	14667
23	11番目の顧客要求文	9531
24	12番目の顧客要求文	888

付録 B.2 誤認識を考慮した回答時間の予測モデル構築における 24 個の特徴量の重要度

順位	特徴量（単位時間当たりのfixation数）	重要度
1	7番目の顧客要求文	40770
2	3番目の顧客要求文	38709
3	3番目の正例	38572
4	9番目の顧客要求文	37249
5	1番目の正例	34780
6	5番目の正例	33859
7	4番目の顧客要求文	29997
8	1番目の負例	29568
9	4番目の正例	27187
10	2番目の正例	27083
11	6番目の顧客要求文	24823
12	2番目の負例	23569
13	1番目の顧客要求文	22492
14	6番目の正例	20814
15	3番目の負例	20316
16	5番目の顧客要求文	20115
17	2番目の顧客要求文	19233
18	8番目の顧客要求文	19230
19	4番目の負例	14089
20	6番目の負例	13103
21	10番目の顧客要求文	11715
22	5番目の負例	11693
23	11番目の顧客要求文	7928
24	12番目の顧客要求文	1096

付録 B.3 正答／誤答分類モデル構築における 24 個の特徴量の重要度

順位	特徴量（単位時間当たりのfixation数）	重要度
1	7番目の顧客要求文	8.35
2	6番目の正例	6.33
3	8番目の顧客要求文	5.71
4	3番目の顧客要求文	5.45
5	5番目の正例	5.32
6	4番目の正例	5.23
7	9番目の顧客要求文	4.94
8	3番目の正例	4.83
9	5番目の顧客要求文	4.71
10	6番目の顧客要求文	4.66
11	2番目の顧客要求文	4.55
12	2番目の正例	4.29
13	1番目の正例	4.17
14	10番目の顧客要求文	4.1
15	1番目の顧客要求文	4.01
16	4番目の負例	3.95
17	4番目の顧客要求文	3.78
18	1番目の負例	3.67
19	3番目の負例	3.65
20	2番目の負例	3.57
21	5番目の負例	3.38
22	11番目の顧客要求文	3.24
23	6番目の負例	2.19
24	12番目の顧客要求文	0.4

研究業績

学術誌掲載論文

1. 斉藤功樹・日髙昇平, 視線計測情報に基づいた要件定義書レビュー品質の予測，認知科学，査読有，27(3)，280–294，2020

国際学会口頭発表論文（ポスター発表を含む）

2. Saito, K., & Hidaka, S., Analysis of review quality by using gaze data during document review. Proceedings of The 41st Annual Meeting of the Cognitive Science Society (CogSci2019)，査読有, 2695–2701, 2019 年 9 月 24 日 – 27 日, Montreal, Canada
3. Saito, K., & Hidaka, S., Gaze behavior in a review-a-definition task. Proceedings of The 42nd Annual Meeting of the Cognitive Science Society (CogSci2020)，査読有, 1140, 2020 年 7 月 29 日 – 8 月 1 日, Toronto, Canada

国内学会口頭発表論文（ポスター発表を含む）

4. 斉藤功樹・日髙昇平，定義の必要十分性判別能力を試す課題の構築，日本認知科学会第 36 回大会論文集，査読有，188–192，2019 年 9 月 5 日 – 7 日，静岡大学　浜松キャンパス
5. 斉藤功樹・日髙昇平，レビュータスクにおける情報量と視線の関係分析，日本認知科学会第 37 回大会論文集，査読有，538–541，2020 年 9 月 17 日 – 19 日，オンライン開催

謝辞

本研究を進めるにあたり，終始適切な助言を賜り，丁寧かつ熱心に指導して頂きました日髙昇平准教授に深い感謝の意を表します．仮説の立案，実験の進め方からデータ分析手法など，研究に必要な様々なことを教えていただきました．本当にありがとうございました．

本学位論文の審査を引き受けて下さった，内平直志教授，西本一志教授，藤波努教授，および早稲田大学の鄭顕志准教授に感謝致します．お忙しい中，本研究に対して建設的なコメントを頂き，ありがとうございました．鄭先生には，北陸先端科学技術大学院大学に入学以前から，ソフトウェア工学の観点で助言を頂き，大変お世話になりました．心より感謝致します．

日髙研究室のゼミにて，様々助言を頂いた鳥居拓馬助教，早稲田大学の布山美穂講師，および大学院生の方々に感謝致します．実験の進め方やデータ分析に対する新たな洞察を与えていただきました．

業務の忙しい時間にも関わらず快く実験に参加してデータを提供いただいた日本ユニシス株式会社の皆様に感謝致します．合わせて，博士課程に在籍することを承諾し，ご支援して頂きました日本ユニシス株式会社の羽田昭裕 CTO，香林愛子センター長，石原英里室長，佐藤一広室長にも心より感謝致します．

最後に，本研究を進めるにあたり陰ながら支えてくれた妻の麻由に深い感謝の意を表します．