

## レビューの新たな品質メトリクス収集「エーオース (Ēōs) 法」の提案 ～レビュー会議の会話データを分析し、発言に関する心理的パラメータを測定～

研 究 員 : 星野 智彦 (株式会社アイシン)  
          村上 薫 (株式会社日立システムズ)  
          芳沢 圭一 (株式会社オーグス総研)  
主 査 : 中谷 一樹 (T I S 株式会社)  
副 主 査 : 上田 裕之 (株式会社D T S インサイト)  
アドバイザー : 安達 賢二 (株式会社H B A)

### 研究概要

ソフトウェアレビューは、有識者や専門家がレビューアとなって問題指摘や指導が行われるのが一般的であるが、参加者全員で協力して欠陥検出を行うことでレビュー効果がさらに高まる。この場合、参加者間の信頼関係の構築、つまり誰もが「発言しやすい場」にすることが課題となる。そのためには、レビューにおける「発言しやすさ」を測定し改善することが必要である。そこで、新たなレビュー品質メトリクス「エーオース (Ēōs : easy to speak) (発言しやすさ)」を提案することにした。そして実験を行い、影響を与えるパラメータを抽出することに成功した。

### 1. はじめに

ウォーターフォール開発では、各工程で作成される成果物の欠陥検出のための手段として、ソフトウェアレビューを実施することが一般的である。従来、ソフトウェアレビューは有識者や専門家がレビューアとなり、問題指摘や指導などを行う場として認識されてきた。これを受けて、レビューイは成果物の修正を行うこととなる。

一方、アジャイル開発では、チームメンバー間に上下関係を設けず、メンバー同士の信頼関係が重視される。レビューの場においても指摘や指導的な側面よりも、共に問題を解決する場として捉えられることが多い。参加者全員で協力して欠陥検出を行うことで、レビュー効果が高められている。最近の動向としては、ウォーターフォール開発においても、レビューにおけるメンバー間の信頼関係が着目されつつある<sup>[1]</sup>。

本研究では、レビューにおけるメンバー間の信頼関係について、誰もが「発言しやすい場」になっているかという視点からみていく。まず「発言しやすさ」を表すパラメータを定義したうえで模擬レビューにおいて測定を行う。次に、発言しやすい場を作るための要因を考察する。そのことにより、レビュー品質向上の一助となる情報を提供する。

### 2. 背景

ソフトウェア開発の現場では、レビューにおける議論が活発に行われず、欠陥が全く検出されないことや、たとえ検出されても有効な指摘ではないことも少なくない。つまり、効果的なレビューが恒常的に行われているとは必ずしも言い切れない。

この問題を引き起こす原因を考えると、必ずといっていいほど挙げられるのは、レビューの場において「発言しやすさ」に問題点があるということである。例えば、レビューアがレビューイに対して高圧的な態度を取ること、参加者同士の関係が劣悪であること、などが挙げられる。そういった心理的環境においては、レビューイが指摘されることを恐れたり、自信を無くしたりすることもあると考えられる。

我々は、この「発言しやすさ」がレビューに及ぼす影響に着目した。レビューが「発言しやすい場」であると、レビュー効果も高まるのではないかと考えた。

### 3. 課題設定

#### 3.1 設定した課題

本研究では、「発言しやすさ」を左右するパラメータを定義したうえで、そのパラメータを実際のレビューの場で測定可能にし、コントロールできるようにすることを課題として設定した。

これにより、レビューを「発言しやすい場」にし、レビュー参加者「全員が建設的に発言」できるようにすることを本研究の目標・ねらいとする。その結果として有効な指摘が増加するといった「レビュー効果の向上」も期待されるが、そこは本研究の対象外とする。

### 3.2 先行研究と本研究の視点

レビューにおける心理的環境に着目した研究として、ソフトウェア信頼性工学における人的モデルにより、人的要因が設計レビュー作業効率に影響を与える効果を明らかにする実験を行っている<sup>[2]</sup>。結果として、レビュー時の議論における人的要因が大事だとしている。BGM などを利用すると、レビューの場が和み「雰囲気」がよくなると考えられるのは確かであろう。

しかし、場が和んでいればレビューの効果は本当に高まるのだろうか。ここで注意しなければならないのは、「和んだ雰囲気＝発言しやすい場」なのかということである。場が和んでいるように見えても、参加者同士お互いに遠慮しあって指摘ができないということもあるのではないだろうか。逆に、一見和やかではない雰囲気でも、強い口調で相手の発言に対する反論をしているような場合でも、実はそれはお互いを理解し認め合ったうえで議論に集中しているという状況かもしれない。これは、エイミー・C・エドモンドソンの「心理的安全性」<sup>[3]</sup>の考え方と似ている。

次に、レビューの可視化に着目した研究として、TMBRI 法<sup>[4]</sup>がある。TMBRI 法は、レビュー会議の目的を参加者で共有し、会議での発言内容毎の時間と発言者を測定、可視化し、全員が参加して分析を行う手法であり、目的の曖昧さを認識し、目的の明確化とレビュー会議の改善を行うというものである。ただし、その発言内容がレビューにおける心理的環境にどのような影響を与えているかまでは考察されていない。

### 3.3 本研究におけるレビュー評価基準

我々は、レビューの場を評価するための基準として「発言しやすさ」に着目した。発言しやすい場とは、先述の通り、見た目の和やかさや居心地のよさのような「雰囲気」とは異なる。

### 3.4 仮説

以上のことを鑑みて我々は以下の仮説を立てることとした。

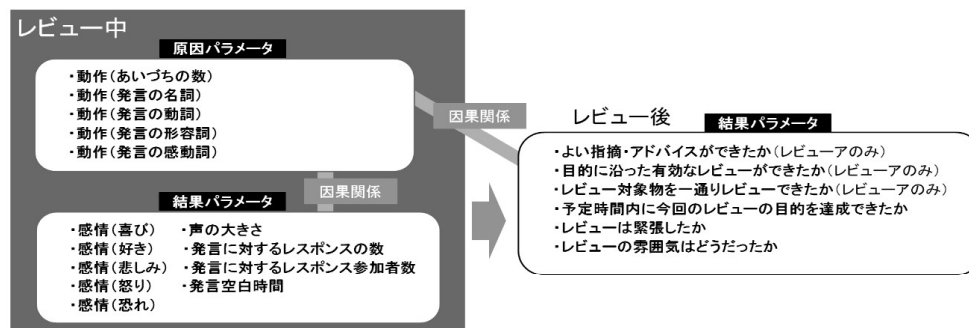
「発言しやすい場で行われるレビューでは議論が活発になり、結果としてレビュー効果が高まる」、メンバー間の信頼関係が構築できており、お互いを理解しあって発言しやすい環境で行われるレビューであるほど、レビューの目的を達成しやすくなると考えた。

### 3.5 「発言しやすさ」を表すメトリクス（「エーオース（Ēōs）」）

レビューにおける発言しやすさを客観的に表すメトリクスとして、我々は「エーオース（Ēōs）」を定義することにした（図 1）。Ēōs はギリシア神話に登場する暁の女神である。「夜明けのごとく明るい心理的環境」では、夜の暗中には見えなかった探し物（欠陥）を見つけやすくなる」という意味を込めて命名した。

「発言しやすさ」の要因は多岐に渡るため一つの公式で示すことは難しいと考え、本研究では影響するいくつかのパラメータとそれらの関連性（因果関係）を探ることとした。

「Ēōs」にはレビュー中に測定するパラメータとレビュー後に測定するパラメータがある。レビュー中に測定するもののうち「動作」は「発言しやすさ」を左右する原因となるため、これを「原因パラメータ」と呼ぶことにする。レビュー中に測定する「感情」等のパラメータやレビュー後のアンケートは結果として「発言しやすさ」がどうだったかを表すパラメータであり、これを「結果パラメータ」と呼ぶ。各パラメータの関連性（因果関係）は図 1 のようになる。



※発言に対するレスポンス = ある発言に対して言葉で反応すること。「うん」「そう」などのうなづきも含む。

図 1 「Ēōs」

#### 4. 実験方法

前項で定義したパラメータを収集するために実験を行った。実験はオンライン会議システム「Zoom」上で行い、Zoomのレコーディング機能を利用して動画を記録し、その動画から音声データを抽出。音声データから文字起こしを行い、各種パラメータを収集した。

本研究では感情分析に AI テキストマイニング<sup>1</sup>を使用した。このツールは、インプットとなるテキストから「喜び」「好き」「怒り」「恐れ」「悲しみ」という感情の大きさ数値化し、時系列でグラフ化する機能を持つ。このツールの使用にあたり、分析結果の妥当性を検証するために予備実験を行うことから始めた。

次に本番実験として、他チーム研究員の協力のもとロールプレイにより「発言しやすいレビュー(以下、ホワイトレビュー)」と「発言しにくいレビュー(以下、ブラックレビュー)」を模擬レビューとして再現し、そこから各種パラメータを収集した。

##### 4.1 予備実験の方法

AI テキストマイニングツールの妥当性を検証するために、ホワイトレビュー(全員が建設的に発言)とブラックレビュー(特定の人が一方向的に発言)のシナリオ<sup>[5]</sup>(別紙 I)を利用し、そのテキストデータを AI テキストマイニングツールで分析した。

##### 4.2 本番実験(ロールプレイによる模擬レビュー)の方法

ここでは、分科会内の他チーム研究員を 2 つのチームに分けたうえでレビューアの役割を担ってもらい、それぞれのチームがホワイトレビューまたはブラックレビューを再現することを意識して、模擬レビューを実施した。レビューイは執筆者の中の 1 名が両方の模擬レビューの中で同じ役割を担った。これは、2 つのレビューのレビューアの態度によるレビューイの感情の違いを見るためである。感情については AI テキストマイニングツールで収集し、その他のパラメータは動画データおよび音声データの文字起こしから収集した。また、各レビュー終了後にアンケートを実施し、レビューへの満足度や緊張感、雰囲気といった定性的パラメータを収集した。

#### 5. 実験結果と考察

##### 5.1 予備実験の結果と考察

図 2 はレビューにおける発言を AI テキストマイニングツールにかけた結果であり、横軸はレビューの経過時間、縦軸はその時点の感情の大きさを表す。ブラックレビューに比べてホワイトレビューの方が「恐れ」や「悲しみ」が少ないことが見て取れる。

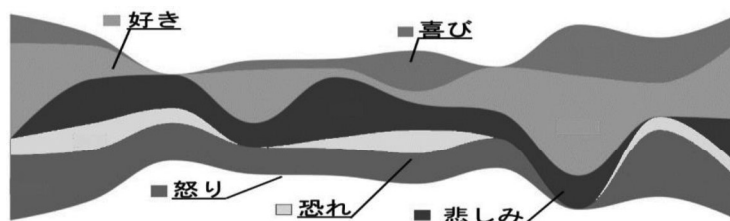


図 2-A 予備実験結果(ブラックレビュー)

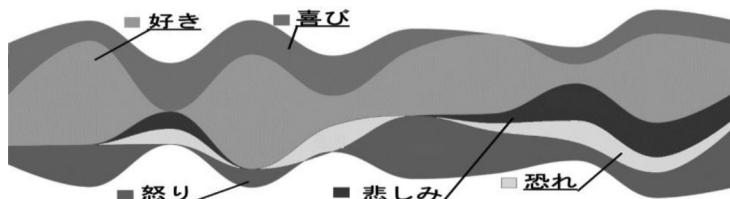


図 2-B 予備実験結果(ホワイトレビュー)

<sup>1</sup> 「ユーザーローカル AI テキストマイニング」による分析  
( <https://textmining.userlocal.jp/> )

このように AI テキストマイニングツールによる分析結果は、ブラックレビューとホワイトレビューで明確に違いが出ており、シナリオが意図した通りの結果が現れている。ゆえに、このツールはレビュー時の感情の定量化という点において、我々が期待する結果を出せると考えた。

## 5.2 本番実験(ロールプレイによる模擬レビュー)の結果と考察

### 5.2.1 原因パラメータ

#### (1) 動作(あいづちの数)

模擬レビュー中にレビューイ、レビューア双方が会話の中であいづち動作を行った回数を収集した。「あいづち」については「はい」「うん」「そう」「そうですね」「なるほど」「確かに」のいずれかの用語が使用されているセンテンスを 1 回としてカウントした。例えば「はい、そうですね」のように 1 センテンスに複数の用語が含まれる場合も 1 回とカウントした。

表 1 をみると、レビューイの回数はブラックレビューとホワイトレビューで大きな違いは見られないが、レビューアの回数はホワイトレビューがブラックレビューの 3 倍超となっている。このことから、レビューアのあいづち回数は発言しやすさに大きく影響すると考えられる。

表 1 あいづち用語出現頻度

あいづち	ブラックレビュー		ホワイトレビュー	
	レビューイ	レビューア	レビューイ	レビューア
はい うん そう そうですね なるほど 確かに	31	10	27	33

#### (2) 動作(発言の名詞・動詞・形容詞・感動詞の出現頻度)

発言に含まれる単語の出現頻度についてはブラックレビューとホワイトレビューで若干の違いが確認できた(別紙Ⅱ)。まず動詞では、ホワイトレビューにのみ出現する「思う」という単語が特徴的である(出現頻度 12)。これは言い切るのではなく表現を弱めることによる相手への気遣いと考えられる。また形容詞をみるとホワイトでは「いい」「すごい」「嬉しい」といった肯定的な単語が出ているのに対し、ブラックレビューでは「面倒くさい」というようなネガティブに感じられる単語が出現している。

もう一つ特徴的なのは、ホワイトレビューでは「わかりやすい」「わかりづらい」をはっきり言葉にしている点である。これは、発言しやすい場であることの表れといえる。

### 5.2.2 結果パラメータ

#### (1) 感情(喜び、好き、悲しみ、怒り、恐れ)

模擬レビュー中の感情の推移をグラフに表したのが図 3 である。「喜び」の感情については、ホワイトレビューの方の値が大きくなっている。一方、「怒り」の感情は、指摘や議論の段階でも増加することがない。「悲しみ」「恐れ」の感情については両者大きな差は見られなかった。

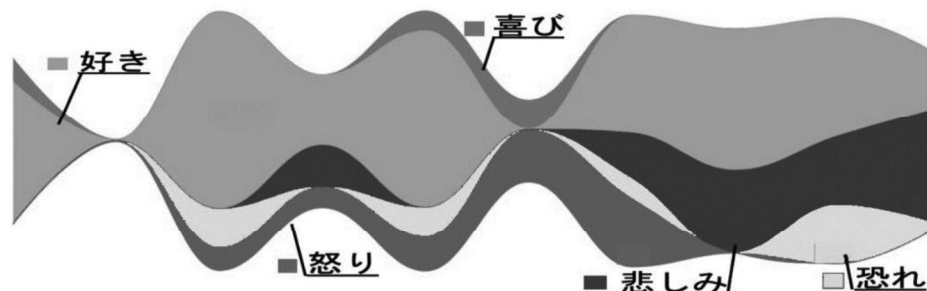


図 3-A 感情と声の大きさ(ブラックレビュー)

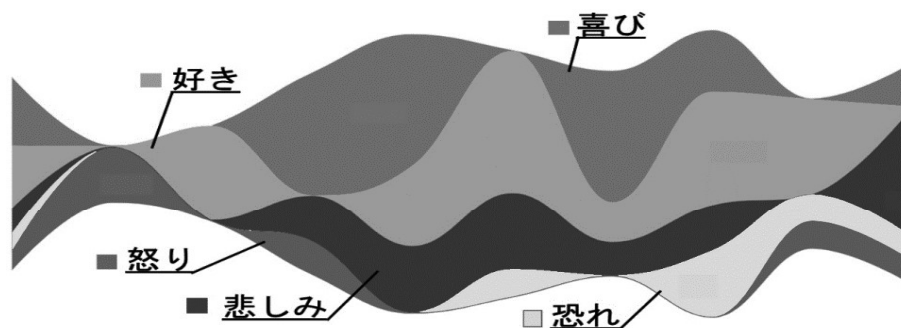


図 3-B 感情と声の大きさ(ホワイトレビュー)

(2) 声の大きさ

図 4 を見ると, 指摘を受けた際に大きく値を下げる (dB が小さくなる) 傾向が見られた. ただし, それはブラックレビュー, ホワイトレビューの双方から見える傾向であるため, 両者の差分となり得るパラメータとは言えない.

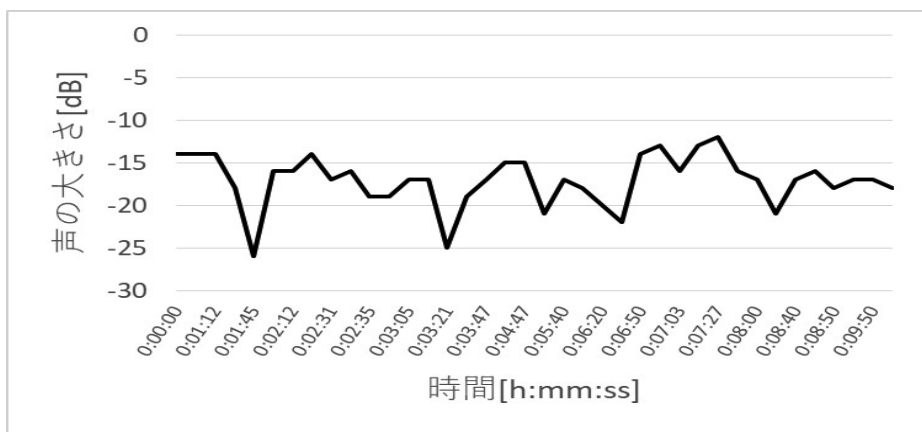


図 4-A 声の大きさ(ブラックレビュー)

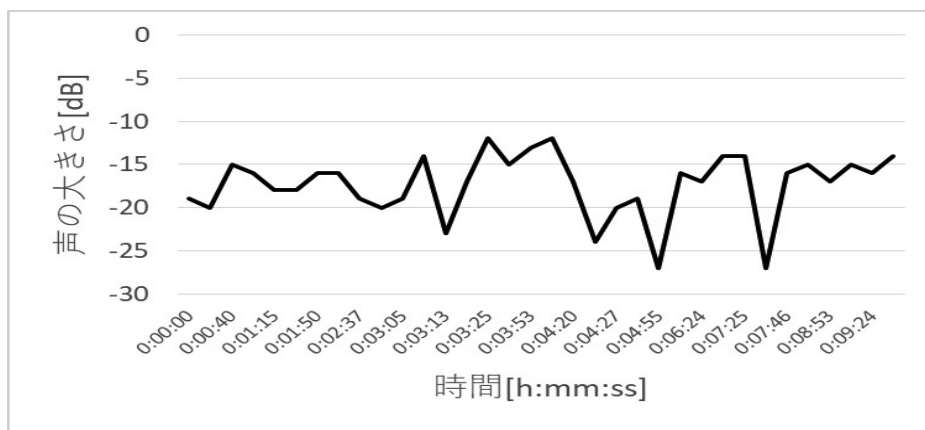


図 4-B 声の大きさ(ホワイトレビュー)

(3) 感情と声の大きさの相関について

感情または声の大きさ, それぞれ単体では一定の効果の確認ができた. その感情のゆらぎが声の大きさに左右されることが言えれば, 感情をコントロールするために, レビューイの声の大きさが関与していると言える.

ただし, 指摘に対する感情のゆらぎと声の大きさの変動に対しての相関は見られなかった.

(4) 発言に対するレスポンスの数と参加者数

レビューイの発言に対するレビューアのレスポンスについては、ブラックレビューとホワイトレビューで明確な違いが見られた。図 5 は、レビュー中の「話題(同一テーマ)」ごとのレスポンス回数とレスポンス相手を表している。ブラックレビューではレスポンス数が少なく(最大 20 回)、レスポンス相手数は基本的に 1 対 1 でのやり取りにとどまっている(図 5-A)。図 5 の見方については別紙Ⅲで補足している。

ブラックレビュー

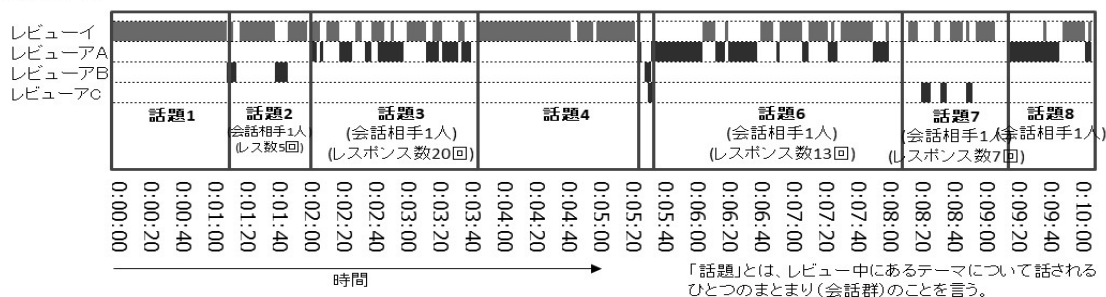


図 5-A 発言に対するレスポンス(ブラックレビュー)

それに対してホワイトレビューではレスポンス数が多く(最大 30 回)、レスポンス相手も複数となっており、活発な会話のやりとりが行われていることがうかがえる(図 5-B)。

ホワイトレビュー

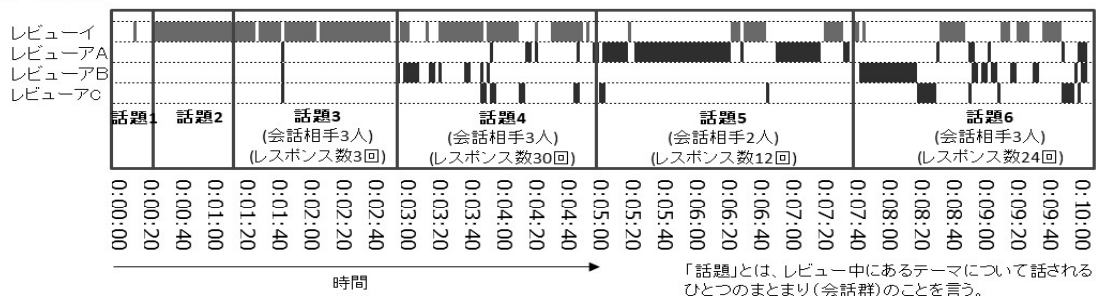


図 5-B 発言に対するレスポンス(ホワイトレビュー)

ただし、今回の実験では「話題」がすべてレビューに関する内容であったが、実際のレビューでは雑談等が行われることもある。この場合、レビューでの「発言しやすさ」を単純に測れるものでなくなる可能性があることも考慮する必要がある。

(5) 発言空白時間

会話における発言と発言の間の空白時間を収集した。ブラックレビューとホワイトレビューそれぞれの、直前発言から該当する行の発言の間に生じた空白時間を表にした(別紙Ⅳ 会話空白時間の比較)。5 秒以上の空白時間はブラックレビュー、ホワイトレビュー各 2 回ずつで差は見られなかった。この空白時間は、返答に困った場合に発生することが考えられるが、単に会話が行き詰った場合にも起こりえることは考慮すべきであろう。

空白時間の合計を比較すると、ブラックレビューのほうが長くなっており、全体的に空白時間が多いことは、発言しにくさから何らかの影響を受けていることも考えられる。

(6) アンケートで収集したパラメータ

レビュー実施後のアンケートの結果を別紙Ⅴにまとめた。ブラックレビューとホワイトレビューで満足度や緊張度に違いが見受けられる。

そこで、この定性的結果を「満足度因子」と「心理的因子」に分類し、順序尺度で表 2 のように数値化した。ここでいう満足度因子とはレビュー効果を表すパラメータであり、心理的因子はレビューの「発言しやすさ」を評価するパラメータである。

表 2 結果パラメータの数値化

満足度因子	よい指摘・アドバイスができたか (レビューアのみ)	できた・・・3点
		どちらかといえばできた・・・2点
		どちらかといえばできなかった・・・1点
		できなかった・・・0点
	目的に沿った有効なレビューができたか (レビューアのみ)	できた・・・3点
		どちらかといえばできた・・・2点
		どちらかといえばできなかった・・・1点
		できなかった・・・0点
レビュー対象物を通りレビューできたか (レビューアのみ)	確認できた・・・3点	
	7割超達成できた (一部達成できなかった)・・・2点	
	半分～7割しか達成できなかった・・・1点	
	達成度は半分未満だった・・・0点	
予定時間内に今回のレビューの目的を達成できたか	確認できた・・・3点	
	7割超達成できた (一部達成できなかった)・・・2点	
	半分～7割しか達成できなかった・・・1点	
	達成度は半分未満だった・・・0点	
心理的因子	レビューは緊張したか	緊張した・・・2点
		少し緊張した・・・1点
		緊張しなかった・・・0点
	レビューの雰囲気はどうだったか	厳粛な会議の雰囲気・・・2点
通常の会議の雰囲気・・・1点		
くだけた・緩い雰囲気・・・0点		

4 つの満足度因子の合計値と「緊張」「雰囲気」の心理的因子それぞれとの相関を見るために、実験参加者ごとにプロットシグラフ化したのが図 6 である。

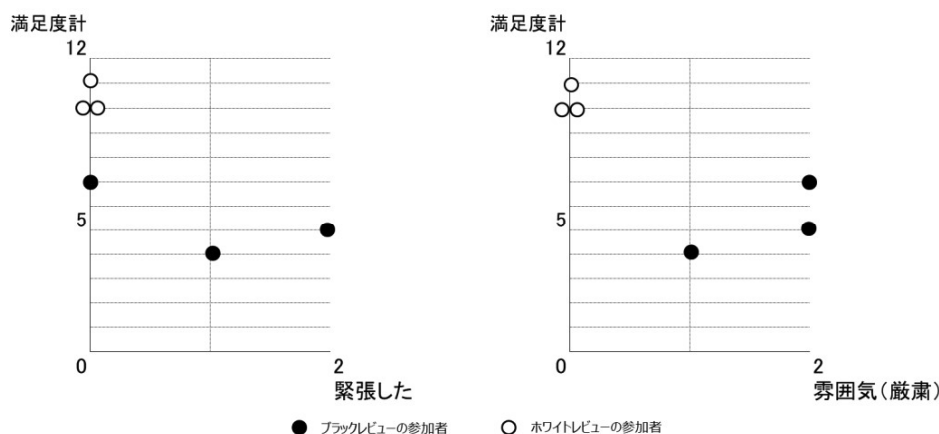


図 6 満足度因子と心理的因子(緊張・雰囲気)

「緊張」「雰囲気」ともに、ホワイトレビューではすべての参加者で満足度の値が大きくなっていることがわかる。一方ブラックレビューでは、満足度は全体に低く、参加者ごとの満足度のばらつきはホワイトレビューと比べて大きい。なお、3.2 節に記載した我々の考え「和んだ雰囲気＝発言しやすい場とは限らない」については、サンプル数が少なく「雰囲気」と「緊張」で大きな差もみられなかったことから、その考えが合っているかを確認することはできなかった。

### 5.2.3 実験結果から分かったこと

以上の実験結果を整理すると以下ようになる。

(1) 「発言しやすさ」に及ぼす影響が大きい「原因パラメータ」

動作(あいづちの数, 発言の動詞・形容詞の出現頻度)は、発言しやすさに影響を及ぼす傾向がある。特に形容詞(「いい」「すごい」「嬉しい」といった肯定的な単語)を与える影響が大きいことが分かった。

(2) 「発言しやすさ」の計測に有効な「結果パラメータ」

感情(喜び), 発言に対するレスポンスの数, 発言に対するレスポンス参加者数, 発言空白時間の合計値, 心理的因子(緊張)は、発言しやすさの計測に有効であることが分かった。

## 6. まとめ

### 6.1 「Eös」の定義

本研究を進めていくにあたり、レビューにおける「発言しやすさ」を測るためのメトリクスとして「Eös」を定義した(図 1)。パラメータには「原因パラメータ」と「結果パラメータ」があり、「原因パラメータ」をコントロールすることで「発言しやすさ」を改善し、「結果パラメータ」で改善結果を測定するという利用方法を想定した。

### 6.2 「Eös」の有効性評価

実験結果から、「Eös」の各パラメータの有効性は以下のように評価できた。

- 「原因パラメータ」では、動作(あいづちの数, 発言の動詞, 形容詞)が有効であった。
- 「結果パラメータ」では、感情(喜び), 発言に対するレスポンスの数, 発言に対するレスポンス参加者数, 発言空白時間の合計値, 心理的因子(緊張)がそれぞれ有効であった。

### 6.3 「Eös」の今後の改善・拡張

今後「Eös」のデータ収集・分析を継続すればパラメータ間の新たな因果関係が見つかり、「発言しやすさ」のコントロール手段を増やすこともできるだろう。

新たなパラメータの追加も考えていくべきである。例えば、以下のようなものが候補となる。

- 原因パラメータ: ジェスチャー(身振り手振り, 身体の動き), 顔の表情(笑顔, 顔色)など
- 結果パラメータ: バイタルデータ(心拍数, 体温, 発汗など), 声の波形(音程, 震え)など

### 6.4 「Eös」の想定できる活用方法の例

今後想定できる「Eös」の活用について、2つの例を示す。

#### 6.4.1 レビューにおけるファシリテーションの質の向上

実際のレビューにおいて、「Eös」の各パラメータに注意しながら進行することで、発言しやすい場となり効率的・効果的なレビューを行うことが出来る。レビューの質を向上させる手段の一つとしてファシリテーションが挙げられる<sup>[6]</sup>が、ファシリテートを行う際にも、明確な観点の下で実施できるようになり、ファシリテーションの質の向上が期待できる。

#### 6.4.2 開発チームの組織的な成熟度の向上

レビューが「発言しやすい場」になっている、つまり良質なコミュニケーションにより高いパフォーマンスを発揮できるチーム状態であることを客観的に示す指標の一つとして活用できる。

## 7. おわりに

本研究では、レビューにおける「発言しやすさ」を測定・改善するためのメトリクス「Eös」を提案した。実験を通じて、影響の大きいパラメータの傾向や、パラメータ間の因果関係を見出すことができた。新たなレビュー品質メトリクス「Eös」の扉は開かれたばかりである。更に有効なパラメータや因果関係が発見される可能性を秘めており継続研究が必要と考えている。

「Eös」をコントロールすることでレビューの質を向上させることが可能になるとすれば、レビューにおけるファシリテーションの方法に関する研究にも大いに役立つものと考えられる。

最後に、論文作成にあたりご指導いただいたソフトウェアレビュー分科会の中谷主査, 上田副主査, 安達アドバイザーおよび実験にご協力いただいた分科会メンバー, また研究活動を支えてくれた各研究員の家族に深く感謝の意を表します。

## 参考文献

- [1] 森崎 修司, 「レビューで失敗しない 8 つのポイント」, 2014
- [2] 山田 茂, 松田 僚太郎, 「ソフトウェア設計レビューにおける信頼性に影響を及ぼす人的要因の品質工学的評価」, 日本経営工学会論文誌. Vol. 54, No. 1, 2003
- [3] エイミー・C・エドモンドソン, 「恐れのない組織——「心理的安全性」が学習・イノベーション・成長をもたらす」, 2021
- [4] 竹森 和哉, 田中 拓也, 中山 匡, 西澤 賢一, 湯川 健「レビュー会議の可視化により目的の曖昧さを明確にする手法の提案」, 2016
- [5] 中谷 一樹, 上田 裕之, 安達 賢二, 「よいレビューと悪いレビューのシナリオ」, 2022
- [6] 金子 聡子, 艸薙 匠, 「ファシリテーション技術を応用したピアレビュー会議の改善—教育の開発と実施—」, 2018