パスワード生成アシスト技術の有効性評価: 異なる言語圏のユーザを対象とした追試研究

森 啓華[1,a)]　長谷川 彩子[2]　渡邉 卓弥[1,2]　笹崎 寿貴[1]　秋山 満昭[2]　森 達哉[1,3]

概要：本研究は，言語や文化の違いがパスワード生成のアシスト技術に与える影響を理解することを目的する．そのためのアプローチとして，日本および英国在住の参加者に対して，Ur ら [9] が作成した 3 種類のパスワード生成アシスト技術（パスワード強度メータ，生成パスワードに対するフィードバック，パスワード改善案）の有効性を評価する追試研究を実施した．3 種類のアシスト技術を組み合わせて提示した状態で参加者にパスワード生成を依頼し，各技術のユーザビリティや生成パスワードのパスワード強度を評価した．その結果，日本と英国の参加者で各アシスト技術の有効性が異なることが示された．ユーザビリティの観点では，日本の参加者ではフィードバックが，英国の参加者ではパスワード強度メータが役に立ったと回答する参加者が多かった．しかし日本での参加者は，フィードバックに対するユーザビリティを低いスコアで回答した．パスワード強度の観点では，日本の参加者はパスワード強度メータを提示した場合に推測されにくいパスワードを生成し，英国の参加者はどのアシスト技術にも同程度の効果があった．本調査の結果より，適切なアシスト技術を提供するためにサービス利用者の言語や文化を考慮する必要があるといえる．

On the Effectiveness of Password Assist Tools: A Replication Study with the Participants from Different Language Spheres

Keika Mori[1,a)]　Ayako A. Hasegawa[2]　Takuya Watanabe[1,2]　Toshiki Sasazaki[1]
Mitsuaki Akiyama[2]　Tatsuya Mori[1,3]

Abstract: This study aims at understanding the influence of linguistic/cultural differences on the effectiveness of the password assistant tools. As a baseline of such tools, we make use of the tool developed by Ur et al. [9], which assists users to create secure passwords with the data-driven password suggestions. Using the internationalized version of the tool, we attempt to replicate their user study with the participants from different linguistic/cultural spheres. We recruited participants from Japan and the United Kingdom and asked the participants to create password using three types of assistant tools (a password meter, text feedbacks, and suggested improvements) and evaluated the security and the usability of them. As a result, we found that the effectiveness of each assist tools is different between Japanese and the UK participants. In terms of usability, the most Japanese participants answered that the text feedback was useful, and the most UK participants answered that the password meter was useful. In terms of security, the password meter was effective for Japanese participants and the 3 assistant tools had the same impact on the UK participants. Our findings imply that developers of password assistant tools need to consider the linguistic/cultural background of the targeted users when they try to obtain the optimized outcomes.

1. はじめに

ユーザに安全なパスワード生成を促す手段として，パスワードポリシーの強化に加えて，ユーザのパスワード生成をリアルタイムでアシストする技術が提案されている．ア

1 早稲田大学 (Waseda University)
2 NTT セキュアプラットフォーム研究所 (NTT Secure Platform Laboratories)
3 情報通信研究機構 (NICT)
a) keika@nsl.cs.waseda.ac.jp

シスト技術に関しては，パスワード推測攻撃に対して頑健なパスワード生成にユーザを導き，かつ，ユーザが理解・操作しやすいような，セキュリティとユーザビリティの両立が求められる．

Ur ら [9] は 3 種類のパスワード生成アシスト技術（パスワード強度メータ，生成パスワードに対するフィードバック，パスワード改善案）の適切なパラメータ設定を調査した．その結果，1class8 のパスワードポリシー下では，パスワード強度メータと共に詳細なフィードバックを提示することでより安全なパスワード生成を促せることを示した．一方パスワード改善案の影響はさほど大きくないことが明らかになった．

最近の研究により，ユーザの言語や文化の違いがパスワード生成方法に影響を与えることが明らかになっている [4,11]．これを踏まえると，アシスト技術の有効性も言語や文化により異なる可能性がある．そこで本研究では，ユーザの言語や文化の違いがパスワード生成アシスト技術の有効性に影響を与えるのかを明らかにすることを目的とする．そのために，米国在住の参加者を対象に調査された Ur らの研究 [9] に基づいて，日本および英国在住の参加者を対象とする追試調査を実施する．

本調査では，クラウドソーシングサービスの Lancers [2] および Prolific [6] を用いて日本および英国在住の参加者を募集した．調査は Ur らの研究の手順に則って実施した．具体的には，Ur らが作成した 3 種のアシスト技術を組み合わせて参加者に提示し，パスワード生成およびログイン試行を依頼した．また，各アシスト技術のユーザビリティについてのアンケートを実施した．そこで，生成パスワードの強度，および，ユーザビリティの観点で，日本および英国の参加者に対する各アシスト技術の有効性を検証した．

本研究の貢献は以下の通りである．

- 日本および英国の参加者で各アシスト技術のユーザビリティが違うことを明らかにした．特に，日本の参加者ではフィードバックが，英国の参加者ではパスワード強度メータが役に立ったと回答する参加者が多かった．しかし日本での参加者は，フィードバックに対するユーザビリティを低いスコアで回答した
- 日本および英国の参加者で各アシスト技術のパスワード強化の効果が異なることを示した．日本の参加者はパスワード強度メータを提示した場合に推測されにくいパスワードを生成し，英国の参加者はどのアシスト技術を用いても同等の効果が得られた．
- 日本と英国の参加者ではパスワード入力方法が異なり，ログイン成功率に差が出ることを示した．日本では紙や電子媒体に記録する参加者が多い一方で，英国では記憶力に頼る参加者が多くログイン成功率が低かった．

本研究で明らかにした言語や文化の違いによるパスワード生成アシスト技術の有効性への影響は，サービス提供者がそれぞれのユーザに適したアシスト技術を実装し，パスワード生成画面のセキュリティとユーザビリティを高めることに役立てられる．

本論文の構成は以下の通りである．2 章では本研究の背景としてパスワード攻撃やパスワードポリシー，パスワード生成のアシストに関わる技術や文化間のパスワード作成方法の違いについて述べる．3 章では調査デザイン，参加者募集方法，評価手法について述べる．続いて 4 章では調査結果を示す．5 章では本研究の制約と課題，研究倫理について論じ，最後に 6 章で本研究の結論を述べる．

2. 背景と関連研究

2.1 パスワード攻撃

パスワード認証への攻撃はオンライン攻撃とオフライン攻撃の二種類に分けられる．オンライン攻撃は，運用されている実サービスに対する攻撃である．多くのサービスは安全のためにパスワード認証試行回数を制限している．よって少ない試行回数で攻撃を成功させるための手法として，既に漏洩した ID とパスワードの組を用意して順に試すパスワードリスト攻撃がある．2019 年 7 月には 7pay [15]，コーナン Pay [14]，および，クロネコメンバーズ [16] がパスワードリスト攻撃を受けたと言われている．

しかしながら，ウェブサービスに保管されるパスワードはハッシュ化されていることが一般的であり，パスワードハッシュを窃取したとしてもそのまま実サービスへの攻撃に使用することはできない．このパスワードハッシュを平文パスワードに戻す攻撃がオフライン攻撃である．平易なパスワードを利用している場合は実時間でパスワードを推測されるため，ユーザには高い強度のパスワードを利用することが求められている．

2.2 パスワードポリシー

NIST をはじめとして JPCERT/CC や総務省などは安全なパスワード生成方法および管理方法を公開している．NIST は Special Publication 800-63B（以降，SP 800-63B とする）にて，パスワードは 8 文字以上とし，漏洩パスワードに含まれる文字列や辞書単語，同じ文字の繰り返しやユーザ名等の使用を避けるべきとしている [5]．NIST の SP 800-63B はサービス提供者向けに推奨しているのに対し，JPCERT/CC や総務省はユーザ向けに安全なパスワードの生成・管理方法を公開している [1,17]．このようなパスワードに使用できる文字数や文字種類（大文字，小文字，記号，数字）に関する条件をパスワードポリシーという．

パスワードポリシーがユーザのパスワード生成に与える影響について調査されている．Shay らは従来推奨されてきた 4class8（4 種の文字種類を含み 8 文字以上）や basic16（16 文字以上）よりも，3class12（3 種の文字種類を含み 12 文字以上）や 2word16（数字や記号で区切られたアルファ

ベット列が 2 つ以上あり，16 文字以上）の方が安全なパスワード生成を促すことを示した [8].

2.3 パスワード生成アシスト技術

NIST の SP 800-63B では，安全なパスワード生成方法を公開するだけでなく，サービス提供者はユーザのパスワード生成をアシストする技術を適用すべきであると述べられている．パスワード生成アシスト技術としてパスワード強度メータ，生成パスワードに対するフィードバック，パスワード改善案が知られている．パスワード強度メータは入力されたパスワード強度を算出し，そのスコアをユーザに視覚的に提示する．フィードバックは入力されたパスワードをさらに強固にするための方針を提示する．同様に，パスワード改善案は入力されたパスワードをもとにユーザが設定すべきより強固なパスワードを自動生成して提示する．

Ur らはパスワード強度のスコアリングや，リスクを示す表示のサイズや色を変化させた 14 種類のパスワード強度メータの有効性を評価した．その結果，スコアが厳しく算出されるほどユーザはより長いパスワードを生成することが示された [10].

Shay ら [8] はパスワード生成時にリアルタイムでポリシーに合致したかどうかをチェックし，フィードバックを表示する場合のユーザビリティとセキュリティを評価した．フィードバックを表示することで正確にポリシーを満たすことができただけでなく，セキュリティへの意識も高まり，ユーザビリティが向上したことを示した．またそれによりセキュリティを損なうことはないと結論づけた．

Sean ら [7] はユーザに提示するパスワード改善案の数や種類（挿入案および代替案）による影響を調査し，パスワード改善案が 0，1，3 個と異なる場合でもユーザビリティおよびセキュリティへの影響に差がないことを示した．文字の挿入による改善を提案した場合と文字の代替による改善を提案した場合においてもユーザビリティに差は見られないことを明らかにした．

Ur らは，3 種類のパスワード生成アシスト技術（パスワード強度メータ，生成パスワードに対するフィードバック，パスワード改善案）の適切なパラメータ設定を調査した [9]．結果，1class8 のポリシー下ではパスワード強度メータと共に詳細なフィードバックを提示することでより安全なパスワード生成を促せることを示した．一方パスワード改善案の影響はさほど大きくないことを明らかにした．

2.4 パスワード生成における異文化間研究

最近の研究により，ユーザの言語や文化がパスワード生成方法に影響を与えることが明らかになっている．Li ら [3] や Zeng ら [13] は中国におけるパスワードを解析し，中国のユーザは数字やピンイン（中国語の発音をアルファベットで表記する仕組み），およびポジティブなイメージをもつ単語を好んで使用する傾向があることを示した．Wang ら [11] は，中国のユーザのパスワード生成の傾向として `1q2w3e4r` のように数字と文字を交互に組み合わせやすいことを発見し，この傾向を利用することで英語圏のユーザのパスワードよりも容易にパスワードを推測できることを明らかにした．Mori ら [4] は日本，英国，中国におけるパスワード生成の傾向を調査し，日本および中国の参加者は個人情報，英国の参加者は一般的な単語を好んで使用する傾向があることを示した．また各国のパスワードの傾向を利用することで攻撃者はパスワードを推測しやすくなることを明らかにした．

ユーザの言語や文化によりパスワード生成方法に違いがあることを踏まえ，我々はパスワード生成アシスト技術の有効性に関しても同様に言語や文化により違いが生じる可能性があると考えた．よって本研究では，パスワード生成アシスト技術の有効性の異文化間調査を実施する．

3. アプローチ

本研究は言語や文化の違いがパスワード生成アシスト技術のセキュリティとユーザビリティに与える影響を明らかにすることを目的とする．Ur ら [9] の研究では米国在住の参加者を対象としてパスワード生成アシスト技術の実験を行なっている．我々は，言語や文化の違いによる影響を明らかするために日本および英国在住の参加者を対象として，Ur らの手法を用いて追試調査を実施する．

3.1 追試元研究

Ur ら [9] はニューラルネットワークとヒューリスティックを組み合わせて正確なパスワード強度のスコアを算出するパスワード強度メータを提案し，加えて，それをもとにしたフィードバックおよびパスワード改善案を含めたパスワード生成画面を作成した．そして，これら 3 種類のパスワード生成アシスト技術（メータ，フィードバック，改善案）に関して各々の最適なパラメータ設計を明らかにするためのユーザ調査を実施した．ユーザ調査においては米国在住の参加者を対象とし，各々のアシスト技術のパラメータおよびパスワードポリシー（1class8，3class12）の条件を変えたパスワード生成画面を参加者に提示して，参加者にパスワードの生成を依頼した．参加者が生成したパスワードの強度およびユーザビリティの観点で評価し，各アシスト技術の最適なパラメータ設計を導き出した．

3.2 調査デザイン

本研究では，3.1 節で述べた Ur らの研究 [9] の追試調査を行う．Ur らは米国の参加者を対象としたのに対し，本研究では日本および英国の参加者を対象にユーザ調査を実施する．本研究では Ur らの方法と同様に調査を二段階に分け，参加者には両方の調査に参加するように依頼した．

調査 1 ではパスワード生成タスクおよびユーザビリティ評価アンケートを実施し，2 日後に調査 2 のログイン試行タスクを実施した．

3.2.1 調査 1：パスワード生成およびユーザビリティ評価

調査 1 には，調査の説明と同意の取得，参加者属性を問う質問，パスワード生成タスク，アシスト技術のユーザビリティに関する質問が含まれる．調査 1 は 5 分程度で完了するよう設計した．

調査説明と同意取得：初めに調査目的と回答の利用方法を記載し，本調査で収集されたデータは安全に保管されること，いつでも参加を辞退できることなどを説明した．また，調査 1 で生成したパスワードを用いたログイン試行タスク（調査 2）を 2 日後に実施することを明記し，普段通りパスワードの生成および管理をするよう依頼した．パスワード生成においては主要メールアカウントのような重要なアカウントを想定するよう説明した．

パスワード生成タスク：パスワード生成タスクにおいては，3 種類のパスワード生成アシスト技術であるパスワード強度メータ，生成パスワードに対するフィードバック，パスワード改善案を組み合わせて，表 1 に示す 5 つの提示条件を設定した．なお，各々のアシスト技術は [9] の調査で最適だと導き出されたパラメータを採用した．また，全ての提示条件において同一のパスワードポリシーとして 1class8 を参加者に要求した．参加者は調査に参加した順番にどれか 1 つの提示条件に割り当てられ（参加者間実験），各参加者はその条件のパスワード生成画面においてパスワードを生成した．3 種類のアシスト技術を全て提示する Basic 条件のパスワード生成画面を図 1 に示す．また図 2 はパスワード改善案の表示例を示す．パスワード生成完了後，ログイン画面でログイン試行をするよう参加者を促した．

ユーザビリティ評価アンケート：ユーザビリティ評価のためのアンケートでは，パスワード生成・アシスト技術・パスワード強度メータ，フィードバックに関する合計 9 種類の項目について 5 段階のリカート尺度（“とてもそう思う”から “まったくそう思わない”）で度合いを評価するよう依頼した．具体的には，表示された画面でのパスワード生成について “いらいらした”・“楽しかった”・“難しかった”，提示されたアシスト技術について “役立った”・“分かりにくかった”・“アシストがあった場合となかった場合でパスワードに違いがあったと思う”，パスワード強度メータについて “スコアが高いことは大切だと思う” ・“スコアは正しかった”，フィードバックについて “フィードバックから新しい知見を得た” という項目である．

3.2.2 調査 2：ログイン試行

調査 1 から 2 日後（24–48 時間以内）に参加者にメールを送信し，調査 1 で生成したアカウントに再度ログインするよう依頼した．ログイン試行に加えて，パスワード入力方法を尋ねた．選択肢として，パスワード管理ソフトやブラウザ機能による自動入力，覚えていたものを手動で入力，紙媒体に記録していたものを見ながら手動で入力，電子媒体に記録していたもの見ながら手動で入力，電子媒体に記録していたものコピー&ペーストした，その他，を提示した．調査 2 は 2 分程度で完了するよう設計した．なお，本調査では依頼メールを受信してから 3 日以内にログインをした参加者のデータを分析した．

表 1　参加者に対する各アシスト技術の提示条件と参加者数（うち，調査 2 参加者）

条件	メータ	フィードバック	改善案	日本 (人)	英国 (人)
Basic	✓	✓	✓	49 (44)	45 (33)
NoMeter		✓	✓	35 (32)	47 (37)
NoFeedback	✓		✓	47 (43)	50 (37)
NoImprovement	✓	✓		44 (37)	45 (35)
None				44 (42)	43 (31)

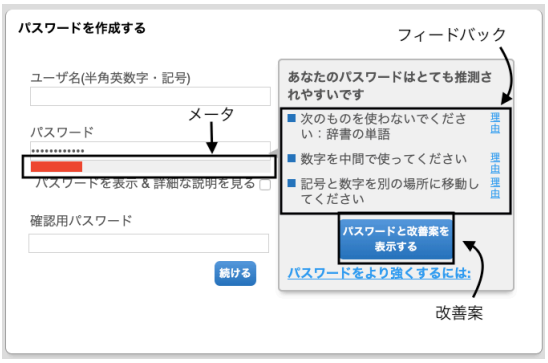

図 1　Basic 条件におけるパスワード生成画面（日本語版）

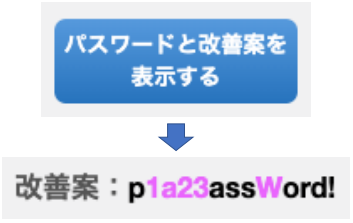

図 2　パスワード改善案の表示例（日本語版）

3.3 参加者募集

日本および英国在住の参加者を募集するために，それぞれ Lancers [2]，Prolific [6] を用いた．Lancers は日本語で，Prolific は英語で運用されているクラウドソーシングサービスである．調査 1 で日本および英国で 250 人ずつ参加者を募集し，調査 2 には 239 人および 201 人が参加した．調査 1 の参加者には，5 分に相当する各国の最低賃金を優に超える 100 円および 0.82 ポンドを報酬として支払った．調査 2 の参加者には再度調査に参加する負荷を考慮し，調査 1 の報酬に追加で 100 円および 0.82 ポンドを支払った．

3.4 パスワード強度の評価方法

調査 1 のパスワード生成タスクで収集したパスワードの強度を PCFG [12] を用いたオフライン攻撃で評価した．PCFG は学習データのパスワードの構成や頻出する文字列

を元に，使われる可能性の高いパスワードを順に生成する．本調査では，対象国を絞った攻撃に対するパスワード強度を評価する．具体的には，日本と英国における漏洩パスワード 20 万件と Rockyou から漏洩したパスワード 1,434 万件からランダムにサンプリングした 20 万件を使って学習した PCFG を使用し，推測パスワードを 10^9 件生成した．日本と英国における漏洩パスワードを利用する理由は，攻撃対象国を絞り，かつ，容易に手に入るその国の漏洩パスワードを使用する攻撃者を想定している．10^9 件の推測パスワードから，評価対象のパスワードが何回の試行でクラックされるかを算出する．

4. 結果

本調査で得られた結果と Ur ら [9] が明らかにした結果を用いて，日本・英国・米国における各アシスト技術のユーザビリティとセキュリティの結果を比較評価する．

4.1 参加者

調査 1 では，各国 250 人分の回答から重複や矛盾を含む回答を除いて，日本では 219 人，英国では 230 人分の有効回答を得た．調査 1 の参加者のうち，調査 2 に参加した人数は日本 239 人，英国 201 人であり．そのうち有効回答は各々 198 回答，173 回答であった．参加者の各提示条件への割り当て状況を表 1 に示す．

日本の参加者の男女比は 56:44 である一方で，英国の参加者の男女比は 28:72 と女性が多く占めた．また，日本および英国ともに幅広い年齢層から回答を得た．参加者のコンピュータサイエンスの専攻状況は日本と英国で大きな差はなく，それぞれ 5%と 7%であった．

4.2 ログイン成功率とパスワード入力方法

各提示条件ごとのログイン成功率を表 2 に示す．調査 1 においてパスワード生成直後のログイン試行を実施した結果，日本の 94%，英国の 96%の参加者は 1 回の試行でログインに成功した．調査 2 において期間をあけてログイン試行を実施した結果，日本の 91%，英国の 75%の参加者は 1 回の試行でログインに成功した．Ur らの研究 [9] では，パスワード生成日のログイン成功率は 98%でほとんどの米国の参加者が 1 回目の試行で成功，2 日後のログイン成功率は 78%でそのうちほとんどの参加者が 1 回目の試行で成功していたため，英国と米国ではほぼ同様の結果が得られた．

日本と比べ，英国や米国では調査 2 の参加者のログイン成功率が低い．これは，パスワード入力方法の違いが影響していると考えられる．調査 2 参加者のパスワード入力方法の回答結果を表 3 に示す．記憶力に頼る参加者の割合が英国では 56%，米国 [9] では 57%を占めていた．一方日本では紙や電子媒体を参照する参加者の割合が高かった．

また，日本・英国どちらにおいても None 条件下でパスワードを生成した参加者は 1 回目の試行での成功率がやや低い傾向にあり，パスワード生成画面の条件もログイン成功率に影響を与えた．

表 2 ログイン試行の成功率

		日本 (%)			英国 (%)		
		1 回	2–5 回	失敗	1 回	2–5 回	失敗
Basic	生成日	92	6	2	95	2	2
	2 日後	91	5	5	79	6	15
NoMeter	生成日	97	0	3	95	2	2
	2 日後	94	3	3	86	5	8
NoFeedback	生成日	94	6	0	91	4	4
	2 日後	95	5	0	78	5	16
NoImprovement	生成日	93	7	0	100	0	0
	2 日後	95	5	0	63	20	17
None	生成日	93	7	0	100	0	0
	2 日後	83	2	14	65	23	13

表 3 パスワード入力方法

	日本	英国
自動入力 (%)	14	24
記憶力に頼り手動入力 (%)	31	56
紙媒体を見ながら手動入力 (%)	19	6
電子媒体を見ながら手動入力 (%)	14	5
電子媒体からコピー&ペースト (%)	20	5
その他 (%)	3	3

4.3 ユーザビリティ

4.3.1 提示条件ごとのユーザビリティ評価

5 つの提示条件におけるパスワード生成画面のユーザビリティについての回答結果を表 4 に示す．ここでは，各質問において，“全くそう思わない” および “そう思わない” と回答した参加者をまとめて “思わない”，“強くそう思う” および “そう思う” と回答した参加者をまとめて “思う” とした．マン・ホイットニーの U 検定の結果，日本と英国で統計的な有意差があった（p<0.01）項目を太字で表した．日本の参加者は，パスワード生成画面上に表示される情報量が多い（つまり複数のアシスト技術を併用する）ほど煩わしさをわずかではあるが感じ，また難しいと感じる傾向にあった．特に，フィードバックがない画面（NoFeedback 条件および None 条件）を提示された参加者は “イライラする” と回答した割合が低い，つまり煩わしさがないと感じた．よって，日本においてはフィードバックがユーザビリティを低下させていると考えられる．一方で，英国の参加者に関しては，アシスト技術と煩わしさに関連性は見られなかった．また 5 つの提示条件全てにおいて，パスワード生成が “難しかった” と回答した参加者の割合は日本の方が高かった．

4.3.2 アシスト技術ごとのユーザビリティ評価

パスワード生成アシスト技術 3 種に関するユーザビリティの回答結果を表 5 に示す．3 種全てにおいて，“役立った” と回答した参加者の割合は英国よりも日本の方が高い．また日本・英国ともに，“役立った”，“アシストがあった

場合となかった場合でパスワードに違いがあった” という質問への回答において，パスワード強度メータ・フィードバックとパスワード改善案の間で有意差（p<0.001）が見られた．詳細を以下に示す．

パスワード強度メータ．日本，英国，米国 [9] の参加者のうち 76%, 72%, 64%がパスワード強度メータは役立ったと回答した．視覚的なアシストはどの言語圏や文化圏においてもユーザビリティが高いと言える．表 4 からも，パスワード強度メータがユーザビリティに悪影響を与えていないことがわかる．また，パスワード強度メータがあった場合となかった場合で異なるパスワードになったと考えるユーザは日本，英国，米国 [9] でそれぞれ 72%，55%，43%で，パスワード強度メータが高スコアを示すことが重要だと回答しているのはそれぞれ 87%，43%，51%であった．英国，米国の参加者に比べ日本の参加者はパスワード強度メータをより活用できていると考えられる．

フィードバック．日本，英国，米国 [9] の参加者のうち 82%, 61%, 62%がフィードバックは役立ったと回答した．一方でそれぞれ 62%，73%，77%の参加者がフィードバックは分かりにくくなかったと回答した．日本の参加者にとってフィードバックは役立ったが分かりにくかったという結果になった．この要因として，本調査で用いたフィードバックは英語圏のユーザ向けの表現を用いた内容であった可能性があること，また日本の参加者にとってパスワードのフィードバックを受けることが珍しい経験であったため馴染みがなかったことが考えられる．表 4 を見ると，日本の参加者のうち NoFeedback 条件および None 条件の下でパスワード生成を行なった参加者は “イライラした” と回答している割合が低い．このことからも，フィードバックは日本においてユーザビリティを低下させる可能性があると考えられる．一方，英国においてはフィードバックがユーザビリティへ影響することは小さいと考えられる．

パスワード改善案．パスワード強度メータやフィードバックと比べると，パスワード改善案が役立ったと回答した参加者の割合は低く（日本 54%，英国 31%），分かりにくかったと回答した割合も他のアシスト技術に比べて日本および英国の双方で高かった．米国 [9] においても同様に，パスワード改善案が役に立ったと回答する参加者は多くなく（51%），さらに提示されたパスワード改善案を使用してパスワードを生成した参加者はわずか 7%であった．米国の参加者はパスワード改善案に対するネガティブな意見として “自己流の生成方法でないと覚えるのが困難”，“コンピュータが生成したパスワードは信用できない” などが挙がったとされている [9].

4.4 オフライン攻撃によるパスワード強度調査

Rockyou の漏洩パスワード，日本および英国の漏洩パスワードを PCFG で学習し，3 種類の 10^9 の推測パスワードを生成する．これらを用いて日本および英国の参加者が生成したパスワードをクラックする 4 パターン（Rockyou 漏洩パスワードから日本参加者のパスワード，Rockyou 漏洩パスワードから英国参加者のパスワード，日本の漏洩パスワードから日本参加者のパスワード，英国の漏洩パスワードから英国参加者のパスワード）の調査を行う．特定の推測数におけるクラックされた参加者の生成パスワードの割合を図 3，4，5，6 に示す．

表 4 各提示条件ごとのユーザビリティに関する回答結果

		日本		英国	
		思わない (%)	思う (%)	思わない (%)	思う (%)
Basic	イライラした	57	24	51	24
	楽しかった	41	18	44	16
	難しかった	**39**	**45**	**67**	**18**
NoMeter	イライラした	54	23	55	26
	楽しかった	37	23	55	11
	難しかった	49	40	62	19
NoFeedback	イライラした	77	6	62	12
	楽しかった	34	21	44	8
	難しかった	53	28	64	12
NoImprovement	イライラした	55	23	40	33
	楽しかった	45	14	47	11
	難しかった	41	45	53	24
None	イライラした	**68**	**9**	**40**	**33**
	楽しかった	43	16	42	12
	難しかった	59	23	65	19

表 5 各アシスタント技術のユーザビリティ関する回答結果

		日本		英国	
		思わない (%)	思う (%)	思わない (%)	思う (%)
メータ	役立った	6	76	8	72
	分かりにくかった	69	14	75	12
	異なるパスワードに	**13**	**72**	**31**	**55**
	高スコアは重要	**3**	**87**	**25**	**43**
	スコアは正しい	4	59	7	61
フィードバック	役立った	**5**	**82**	**18**	**61**
	分かりにくかった	**62**	**23**	**73**	**12**
	異なるパスワードに	**8**	**81**	**28**	**55**
	知見を得た	17	58	30	51
パスワード改善案	役立った	**20**	**54**	**44**	**31**
	分かりにくかった	**40**	**30**	**56**	**20**
	異なるパスワードに	**20**	**52**	**53**	**30**

Rockyou の漏洩パスワードを学習データとした場合（図 3，4），日本ではパスワード強度メータのみ提示されない NoMeter 条件，および英国では何もアシストがない None 条件が推測されやすかった．英国の参加者において，アシスト技術の種類を問わず，リアルタイムでパスワード強度を示すことで安全なパスワード生成を促すことができた．日本の参加者には，パスワード強度メータのような視覚的な評価を示すと効果的であった．一方，日本の None 条件で生成されたパスワードも推測されにくいことがわかる．日本の参加者は一般的な生成方法を指示せずとも十分に強度の高いパスワードを生成できる可能性がある．

各国の漏洩パスワードを使用して PCFG を学習させた場合，日本では NoMeter 条件の攻撃成功率が高いことが判明した（図 5）．これは図 3 の結果と類似しており，日本の参加者はパスワード強度メータを使用することで，一般的なパスワードベースの攻撃，対象国を絞った攻撃どちらに対しても安全なパスワード生成ができたことがわかる．また，フィードバックについては，日本の参加者に対しては

セキュリティの向上には寄与しなかった．ユーザビリティ評価アンケートにおいて，日本の参加者は 82%がフィードバックは役立ったと回答していた．しかしフィードバックが提示される Basic 条件とフィードバックが提示されない NoFeedback 条件との間で，参加者の生成パスワードの強度に差はあまり見られなかった．これは日本の参加者がフィードバックを有効に利用できていないことを意味する．

一方，英国の参加者が生成したパスワードにおいては，10^9 回の攻撃への耐性はどのアシスト技術を用いて生成されたパスワードにおいても同様にセキュリティが向上することが明らかになった（図 4）．Rockyou のような一般的なパスワードを元に攻撃を仕掛けてくる攻撃者に対しては 3 種類のアシスト技術すべて効果的であった．ただし，英国を対象とした攻撃に対しては残念ながらどのアシスト技術も効果がないことがわかる（図 6）．

5. 議論

5.1 制約と課題

本調査で使用した Ur ら [9] のパスワード生成画面は英語圏のユーザを対象として作られており，一般的な漏洩パスワードと英単語の辞書をもとにパスワード強度メータの示すスコアが算出されている．よって日本の参加者にとってパスワード強度メータの示すスコアは最適でなかった可能性がある．各アシスト技術自体を言語・文化ごとに最適化させることが今後の課題である．また，本研究で調査した日本と英国だけでなく，他の言語や文化を対象とした調査の拡張も重要であると考えられる．

5.2 研究倫理

本調査では参加者に擬似パスワードの生成を依頼し，提出された擬似パスワードを解析をした．参加者には調査をいつでも辞退できること，収集したデータは研究目的でのみ使用されること，調査結果が公表される場合でも参加者のプライバシーに関わる情報は守れられることを事前に説明し，参加への同意を得た．また，疑似パスワード生成に関して，決して実サービスで使用しているパスワードを使用しないよう注意喚起を行った．収集したデータは限られた研究者のみがアクセスできる環境で安全に保管した．具体的には，擬似パスワードは暗号化された状態で保管し，解析段階で復号されたパスワードを参加者と紐づけた状態で目視で確認することはしなかった．

5.3 サービス提供者への提唱

日本の参加者は，文字ベースのフィードバックがある場合にユーザビリティが低くなり，パスワード強度メータを使用した場合に安全なパスワードを生成した．よって日本のユーザに対しては文字ベースではなくパスワード強度メータのような視覚的な対策が適していると思われる．英国の参加者は，パスワード強度メータ，フィードバックのユーザビリティに対して肯定的であり，各アシスト技術は同程度に安全性に影響した．よってパスワード強度メータやフィードバックの使用が適していると思われる．

6. まとめ

本研究の狙いは，ユーザの言語および文化の違いが，パスワード生成アシスト技術の有効性にどのように影響を及ぼすかを明らかにすることにある．そのために，Ur ら [9] が作成したパスワード生成画面を用い，異なる言語圏・文化圏の参加者による追試実験を行った．特に，文献 [9] は米国在住の参加者を対象にしていたが，本研究では日本および英国の参加者を対象に調査を実施した．

調査の結果，日本の参加者は，パスワード強度メータがある場合の方が安全なパスワードを生成すること，およびフィードバックやパスワード改善案などの文字ベースのアシストがある場合に苛つきを感じる傾向があり，ユーザビリティが低下することが判明した．また，日本の参加者が生成するパスワードは，文字ベースのアシスト有無にかかわらず，英国の参加者が生成するパスワードよりもセキュリティ強度が高かった．英国の参加者は，パスワード強度メータやフィードバックが役に立つと感じる傾向があることから，これら手法が彼らにとってユーザビリティが高いことがうかがえる．また英国の参加者が生成するパスワードの強度に関して，どのアシスト技術も同程度の効果があることが判明した．以上の知見を総合すると，日本のユーザにはパスワード強度メータ，英国のユーザにはパスワード強度メータとフィードバックの使用が適していることが示唆される．言語や文化によるパスワード生成アシスト技術の影響の違いを理解することにより，サービスの利用者に対して適切なアシスト技術を提供することが可能となる．結果として，ユーザビリティの低下を最小限に抑えた上で，より安全なパスワード生成をユーザーに促すことができる．

参考文献

[1] JPCERT/CC: STOP! パスワード使い回し!キャンペーン 2018，https://www.jpcert.or.jp/pr/2018/stop-password2018.html.

[2] Lancers.inc: Lancers, https://www.lancers.jp/.

[3] Li, Z., Han, W. and Xu, W.: A Large-Scale Empirical Analysis of Chinese Web Passwords, *Proceedings of the 23rd USENIX Security Symposium*, pp. 559–574 (2014).

[4] Mori, K., Watanabe, T., Zhou, Y., Hasegawa, A. A., Akiyama, M. and Mori, T.: Comparative Analysis of Three Language Spheres: Are Linguistic and Cultural Differences Reflected in Password Selection Habits?, *2019 IEEE European Symposium on Security and Privacy Workshops*, pp. 159–171 (2019).

[5] NIST: NIST Special Publication 800-63B Digital Identity Guidelines, https://pages.nist.gov/800-63-3/sp800-63b.html.

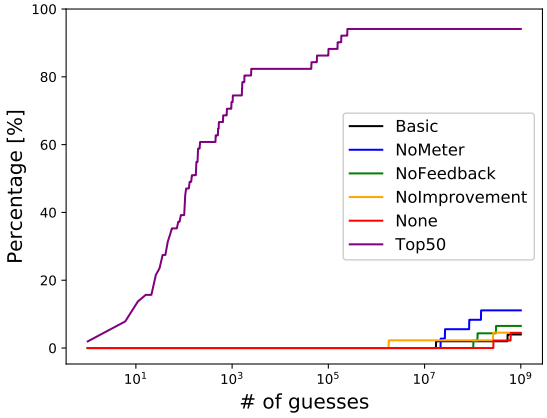

図 3 PCFG（Rockyou）から日本の参加者が生成したパスワードへの攻撃成功率

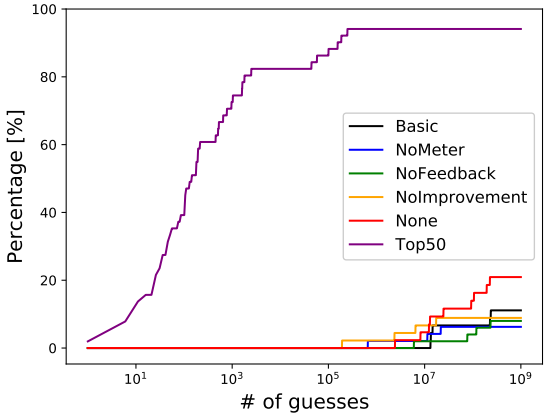

図 4 PCFG（Rockyou）から英国の参加者が生成したパスワードへの攻撃成功率

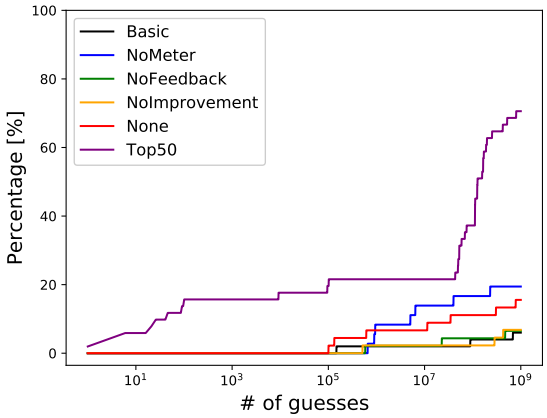

図 5 PCFG（日本）から日本の参加者が生成したパスワードへの攻撃成功率

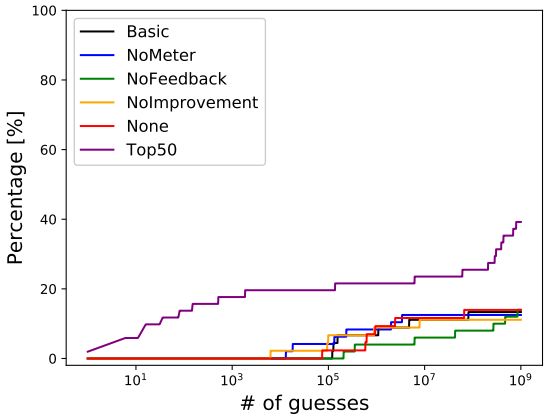

図 6 PCFG（英国）から英国の参加者が生成したパスワードへの攻撃成功率

[6] Prolific: Prolific, https://www.prolific.co.

[7] Segreti, S. M., Melicher, W., Komanduri, S., Melicher, D., Shay, R., Ur, B., Bauer, L., Christin, N., Cranor, L. F. and Mazurek, M. L.: Diversify to Survive: Making Passwords Stronger with Adaptive Policies, *Thirteenth Symposium on Usable Privacy and Security, SOUPS 2017*, pp. 1–12 (2017).

[8] Shay, R., Komanduri, S., Durity, A. L., Huh, P. S., Mazurek, M. L., Segreti, S. M., Ur, B., Bauer, L., Christin, N. and Cranor, L. F.: Designing Password Policies for Strength and Usability, *ACM Trans. Inf. Syst. Secur.*, Vol. 18, No. 4, pp. 13:1–13:34 (2016).

[9] Ur, B., Alfieri, F., Aung, M., Bauer, L., Christin, N., Colnago, J., Cranor, L. F., Dixon, H., Naeini, P. E., Habib, H., Johnson, N. and Melicher, W.: Design and Evaluation of a Data-Driven Password Meter, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3775–3786 (2017).

[10] Ur, B., Kelley, P. G., Komanduri, S., Lee, J., Maass, M., Mazurek, M. L., Passaro, T., Shay, R., Vidas, T., Bauer, L., Christin, N. and Cranor, L. F.: How Does Your Password Measure Up? The Effect of Strength Meters on Password Creation, *Proceedings of the 21th USENIX Security Symposium*, pp. 65–80 (2012).

[11] Wang, D., Wang, P., He, D. and Tian, Y.: Birthday, Name and Bifacial-security: Understanding Passwords of Chinese Web Users, *Proceedings of the 28th USENIX Security Symposium*, pp. 1537–1554 (2019).

[12] Weir, M., Aggarwal, S., de Medeiros, B. and Glodek, B.: Password Cracking Using Probabilistic Context-Free Grammars, *Proceedings of the 30th IEEE Symposium on Security and Privacy S&P 2009*, pp. 391–405 (2009).

[13] Zeng, J., Duan, J. and Wu, C.: Empirical study on lexical sentiment in passwords from Chinese websites, *Computers & Security*, Vol. 80 (2019).

[14] コーナン商事：コーナンPayサービス提供の一時停止について（第2報），https://www.hc-kohnan.com/wp-content/uploads/2019/08/392161d279255f019766ea43aa8fd837.pdf.

[15] セブン&アイ・ホールディングス：「7pay（セブンペイ）」サービス廃止のお知らせとこれまでの経緯、今後の対応に関する説明について，https://www.7andi.com/library/dbps_data/_template_/_res/news/2019/20190801_01.pdf.

[16] ヤマト運輸：クロネコメンバーズにおける不正ログインについて，http://www.kuronekoyamato.co.jp/ytc/info/info_190724.html.

[17] 総務省：国民のための情報セキュリティサイト，http://www.soumu.go.jp/main_sosiki/joho_tsusin/security/business/staff/01.html.