

# スタディサプリENGLISH における Google Speech-to-Text の英語学習への適用

.....  
富田恭平・松田大輝・高山智道  
株式会社リクルートマーケティングパートナーズ

# 登壇者紹介



富田 恭平

- Planner



松田 大輝

- PdM



高山 智道

- Data Scientist

# Agenda

- 1. スタディサプリENGLISHとは
- 2. 解決すべき課題と解決策
- 3. Speech-to-Text 選定プロセスと選定基準
- 4. 文字起こし結果の採点処理
- 5. アーキテクチャと工夫した点
- 6. 今後の展望



# スタディサプリENGLISHとは

# スタディサプリENGLISHとは

- 「1回3分！」のTVCMでお馴染みのスマホでできる英語学習サービス



TOEIC® TEST対策



新日常英会話コース



ビジネス英語コース

# スタディサプリENGLISH とは

- toC だけでなく企業や学校向けにもサービスを提供

toC

toB (企業)

toB (学校)



スタディサプリ

ビジネス英語コース



スタディサプリ

新日常英会話コース

本日お話しさせていただく対象コース



スタディサプリ

ENGLISH

TOEIC® TEST

TOEIC® L&R TEST 対策コース

ベーシックプラン

パーソナルコーチプラン



スタディサプリ

日常英会話コース



スタディサプリ

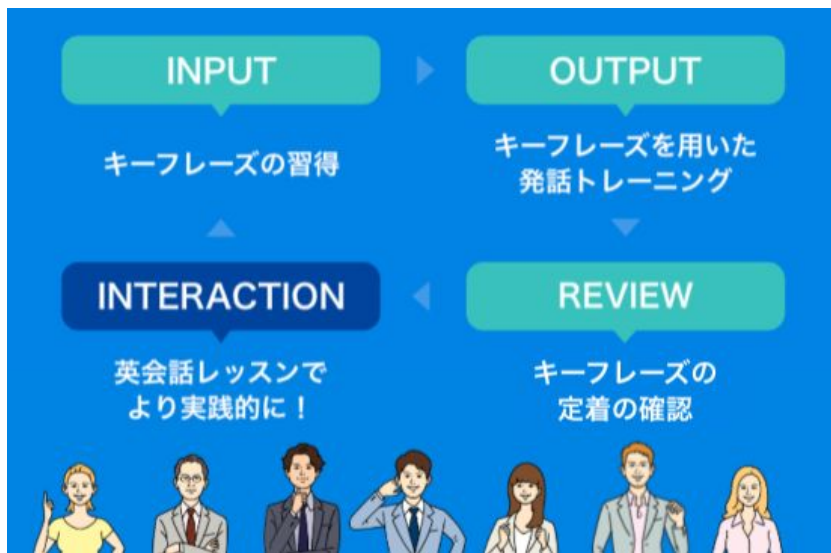
英語 4 技能コース

Google Cloud

# ビジネス英語コース・新日常英会話コースとは

- 本格的なトレーニングで、インプットだけでなくアウトプットまで可能

## 独自の学習メソッド



Google Cloud

## 本格トレーニング



# ビジネス英語コース・新日常英会話コースとは

- どちらもリリースから約1年。アプリストアでも高い評価



ビジネス英語コースの評価

Google Play : ★4.5

App Store : ★4.6



新日常英会話コースの評価

Google Play : ★4.3

App Store : ★4.6

## 勉強する習慣ができた。

毎日どれくらいの時間勉強できたかが記録され、一度もアプリを起動していない日にはリマインドの通知が来るなど、嫌でも毎日勉強せざるを得なくなった。机に向かって時間をとって勉強するのはとても苦手だが、スタディサプリは通勤時間や、ちょっと気が向いたときにやるのに向いているので勉強に対するハードルは下がった。勉強をする習慣がない人にこそお勧めしたい。

## ドラマ形式がインプットに良い。

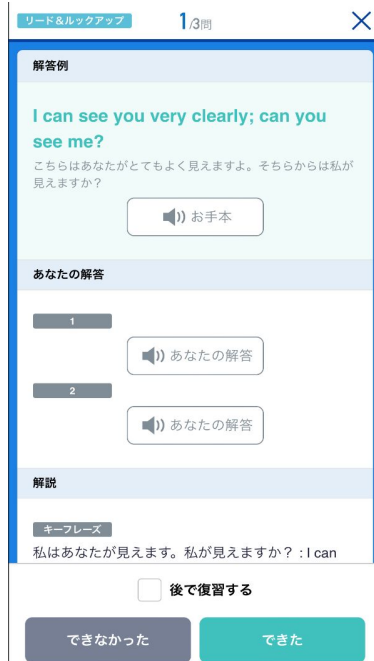
このアプリを使って、改めて何かを覚えていく際には『ストーリー性』が重要であると気づかされました。ドラマの次が気になるだけでなく、一人一人のキャラが立っているのも、ああ前にこの人がこんなシチュエーションでこのフレーズ言ってたなって感じで思い出しながら覚えていけます。全ての会話に感情がこもっているのも良いです。



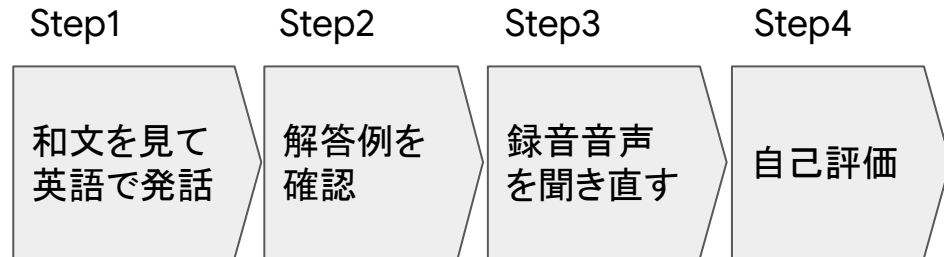
# 解決すべき課題と解決策

# 解決すべき課題

- スピーキング強化トレーニング(リード&ルックアップ)の2つの課題



## 元々のトレーニングの流れ(一部)

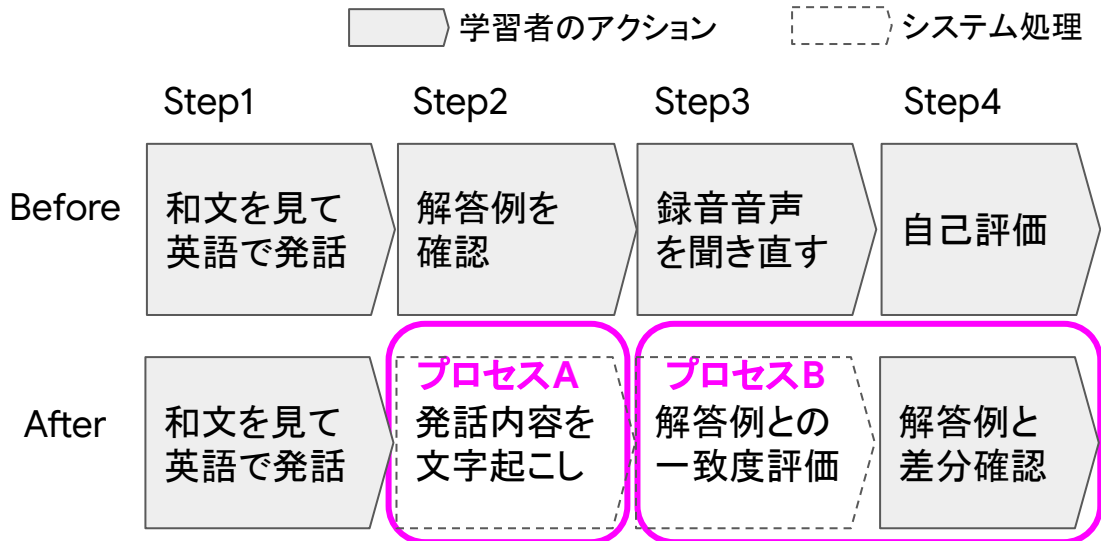


- 課題①: 発話内容の自己評価が面倒
- 課題②: 発音が適切か分からない

# 解決策

- 解決策: 発話内容を文字起こしし、それを自動で評価してくれる機能

## トレーニングの流れの Before / After イメージ



Google Cloud

### 課題①: 発話内容の自己評価が面倒

- プロセス B で解答例との一致度を自動評価することで、学習者の負担を軽減
- 発話内容と解答例の差分をテキストで可視化することで誤った点を直感的な理解を可能に

### 課題②: 発音が適切か分からない

- プロセス A でネイティブが認識可能な発音のみ文字起こしをすることで、どの単語の発音は適切かを学習者が把握可能に

# 解決策

- “After”の各プロセスの具体的なUIイメージは以下

## 英文を見て英語で発話

🔍 和文を見て先ほどの英文を発話しましょう

\_\_\_\_\_ ;  
\_\_\_\_\_ ?

こちらはあなたがとてもよく見えますよ。そちらからは私が見えますか？

🔍 ヒント



パス

## 発話内容を文字起こし

🔍 和文を見て先ほどの英文を発話しましょう

\_\_\_\_\_ ;  
\_\_\_\_\_ ?

こちらはあなたがとてもよく見えますよ。そちらからは私が見えますか？

🔍 ヒント

あなたの解答

I can see you're very clear can you see



もう一度発話する

解答する

## 解答例との一致度評価

キーフレーズ

I can see you; can you see me?  
私はあなたが見えます。私が見えますか？

解答例

🔊 I can see you **very** clearly; can you see me?  
こちらはあなたがとてもよく見えますよ。そちらからは私が見えますか？

あなたの解答

🔊 I can see **you're** very **clear** can you see me

**Great!**

後で復習する

再チャレンジ 次へ

再チャレンジ

次へ

## 解答例と差分確認

キーフレーズ

I can see you; can you see me?  
私はあなたが見えます。私が見えますか？

解答例

🔊 I can see you **very** clearly; can you see me?  
こちらはあなたがとてもよく見えますよ。そちらからは私が見えますか？

あなたの解答

🔊 I can see **you're** very **clear** can you see me

後で復習する

再チャレンジ 次へ

再チャレンジ

次へ

# Speech-to-Text 選定プロセスと 選定基準

# 課題解決する上での技術選定プロセス

- 内製 or 外部 API
  - 文字起こし (Speech-to-Text) はスピード優先で外部 API
  - 採点部分は内製
- Speech-to-Text の選定要件
  - 日本人学習者の英語でも適切に文字起こしされること、が最重要
  - レスポンスやその他の要件は横並び
- 代表的な Speech-to-Text を調査し、最終的に Google Speech-to-Text とカスタマイズ性に優れるライバル製品の 2 つが残った

# なぜ Google Speech-to-Text を選んだか

- 決め手は圧倒的な精度
  - 特にプレミアムモデル
- 定量観点、定性観点どちらも Google Speech-to-Text が勝っていた
  - 定量的な文単位での精度評価
  - 定性的な誤り傾向評価



# なぜ Google Speech-to-Text を選んだか (定量観点 1/2)

- 実際のトレーニング英文を使って精度評価を実施
  - 初級者から上級者までの発話を収集、文字起こしし正解文と比較
  - 文単位での正解率を検証
  - 伝わることを重視し軽微なミスは正解と見なす
- Google Speech-to-Text のプレミアムモデルがライバル製品より 20% 以上高い精度

## 検証結果

Google Speech-to-Text(Premium)	72.6%
Google Speech-to-Text(Default)	53.2%
ライバル製品	45.2%



# なぜ Google Speech-to-Text を選んだか (定量観点 2/2)

- 正しくない発音でも文字起こしされるのではという懸念
  - ネイティブの英語講師の採点結果と比較することで、大きな乖離がないことを確認し、現時点では問題がないと判断
- 今後この問題が顕在化したら対処する。例えば下記
  - 発音チェックを別で設ける
  - 音響モデル/言語モデルを独自に設定する

# なぜ Google Speech-to-Text を選んだか (定性観点)

- 前後の文脈から大きく離れるような間違え方が少なかった

## Google Speech-to-Text

正解	Could you <b>please send</b> those pamphlets directly to our office?
Speech-to-Text	Could you <b>present</b> those pamphlets directory to our office?

## ライバル製品

正解	Do you have <b>anything without alcohol</b> ?
Speech-to-Text	Do you have <b>any sink results are core</b> ?

# 文字起こし結果の採点処理

# 文字起こし結果の採点処理

- 要件: 発話内容の大まかな評価と、間違えた箇所を可視化したい
- 懸念: 語順の入れ替え/削除/挿入が複雑に絡むと計算量が増える  
→ 現実的な処理時間で可視化を実現する独自ロジック実装(特許出願中)
- 今後: より多角的な観点で採点しフィードバックしていく想定

## 間違えた箇所の可視化

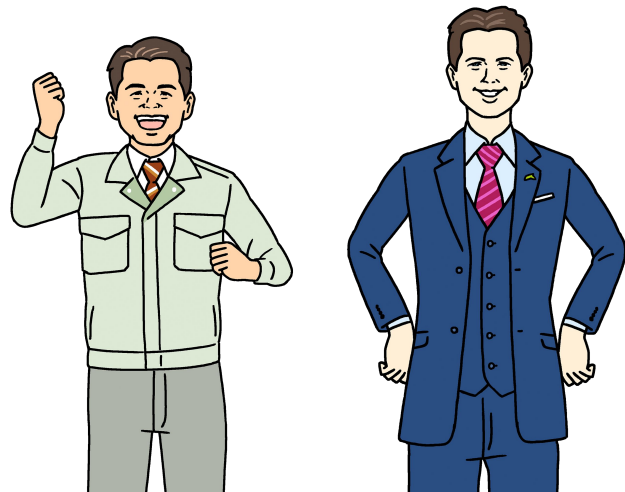
解答例

)) So why don't you know anything about Mr. Foster?

では、なぜフォスター様のことを何も知らないのですか？

あなたの解答

)) So why you don't know something about Mr. Foster



# 文字起こし結果の採点処理

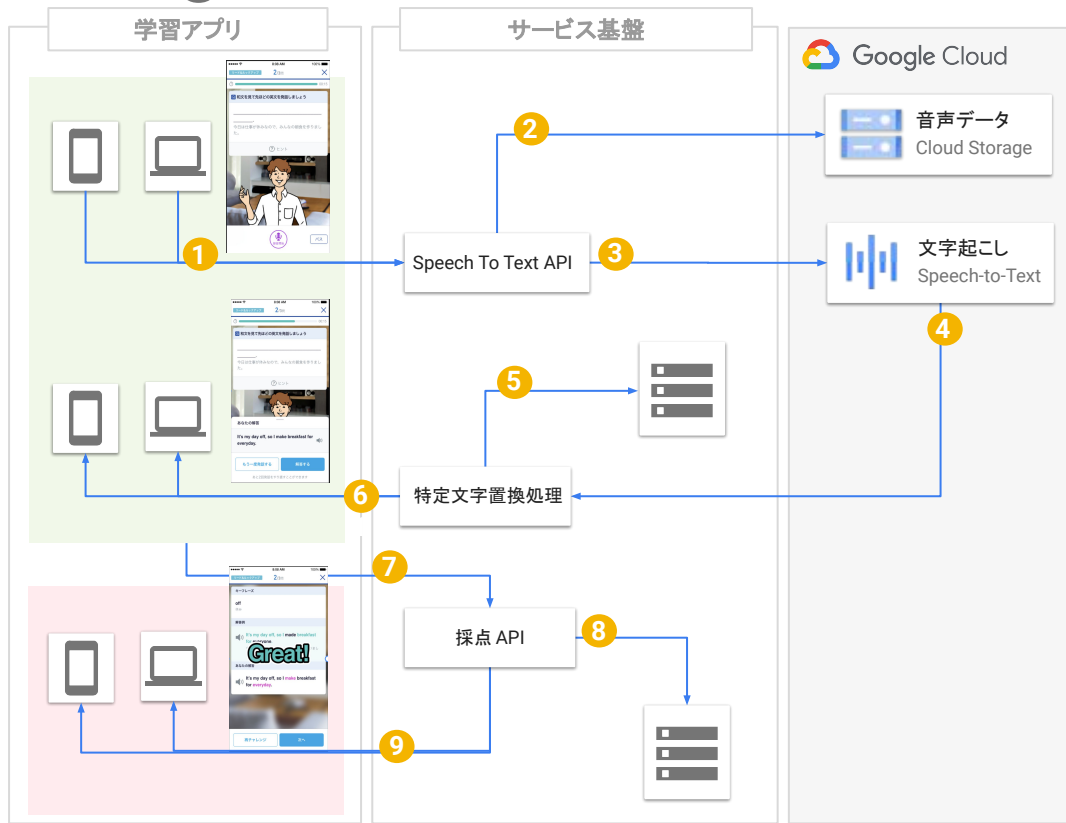
- 採点の前処理について
  - 採点前に出来るだけ表記を揃えておく

処理前	is it <b>okay</b> if I bring my girlfriend <b>to</b>
処理後	is it <b>ok</b> if I bring my girlfriend <b>too</b>

- 記号類の削除や名寄せ
- 単語の名寄せ
  - 短縮系や数字列、同音異表記(ok, okay)など
- 同音異義語の修正
  - too/to や we'll/will など
- 固有名詞への対応

# アーキテクチャと工夫した点

# Google Cloud 利用、構成図



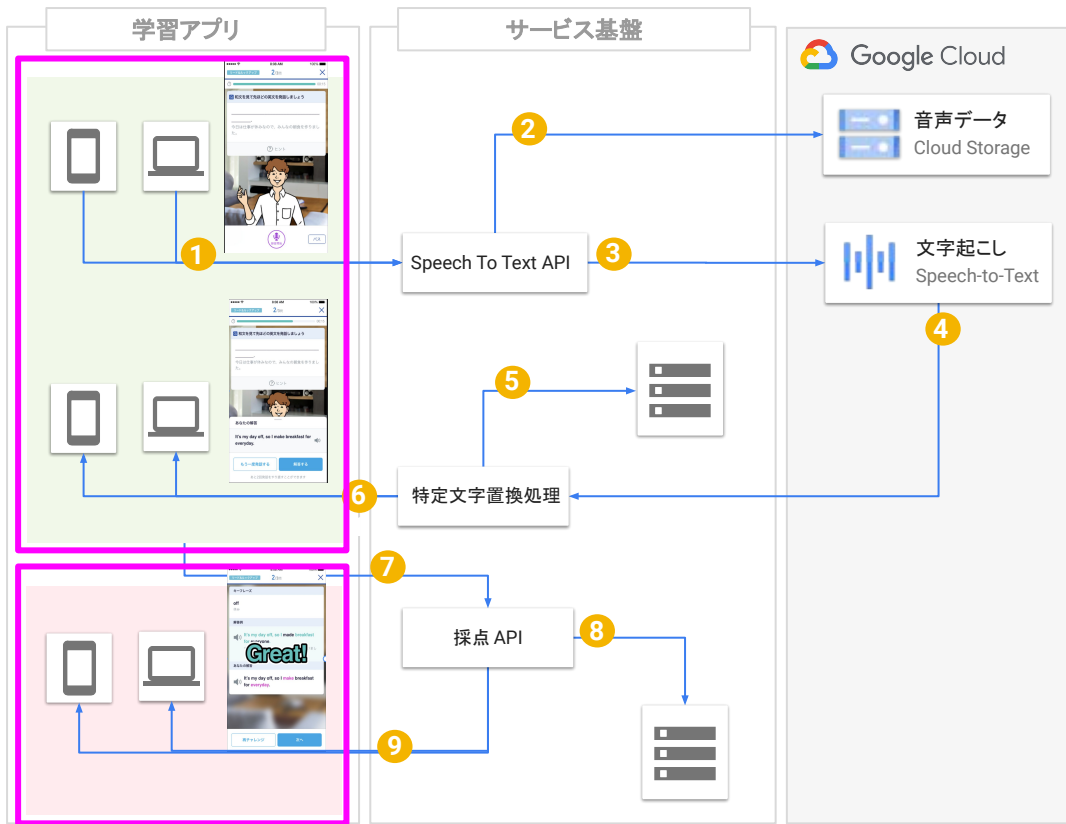
- 1 発話音声の送付
- 2 発話音声データの格納
- 3 音声データを Speech-to-Text に送付
- 4 文字起こし結果を送付
- 5 文字起こし結果を格納
- 6 文字起こし結果を送付
- 7 文字起こし結果と模範解答を送付
- 8 採点結果を格納
- 9 採点結果を送付

Google Cloud

# 工夫した点①(文字起こしと採点の分離)

文字起こし

採点



(理由 1)

点数に意識が行ってしまい、「点数を上げること」が目的になってしまう



まずは、「自分の発話が伝わっているか」= 文字起こしされるのかを意識してもらい、学習効果を高めることが目的

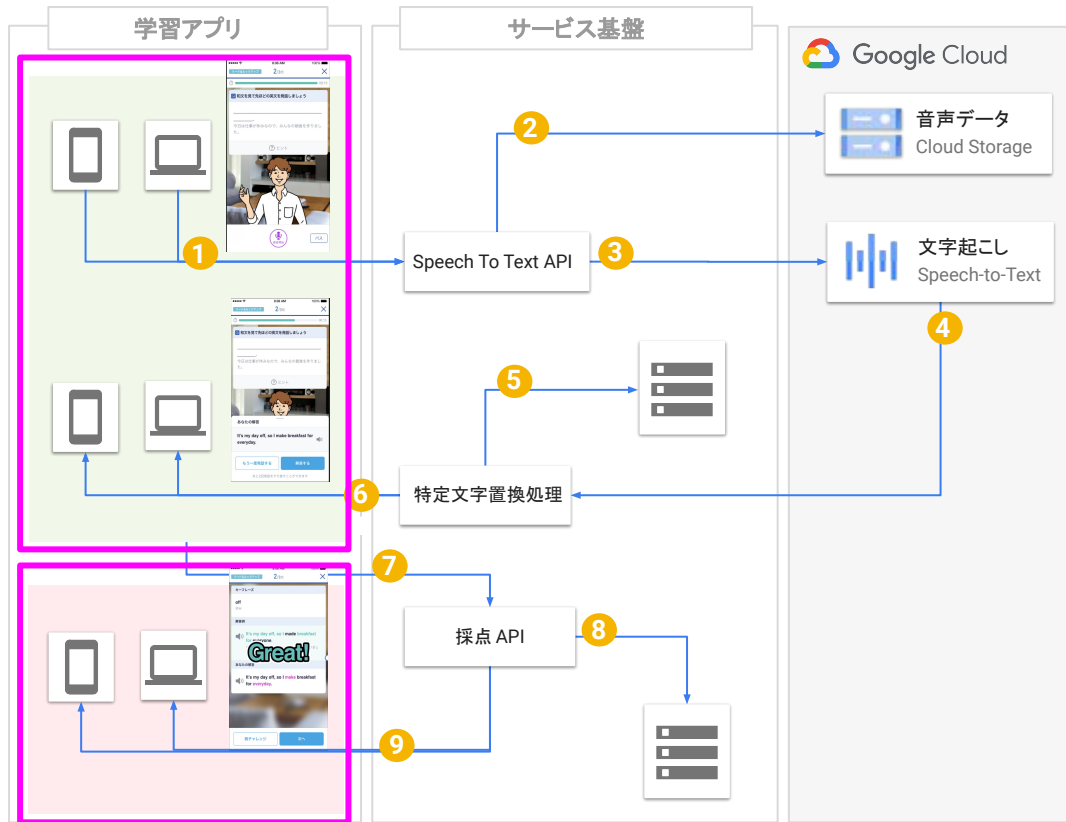
Google Cloud



# 工夫した点①(文字起こしと採点の分離)

文字起こし

採点



(理由 2)

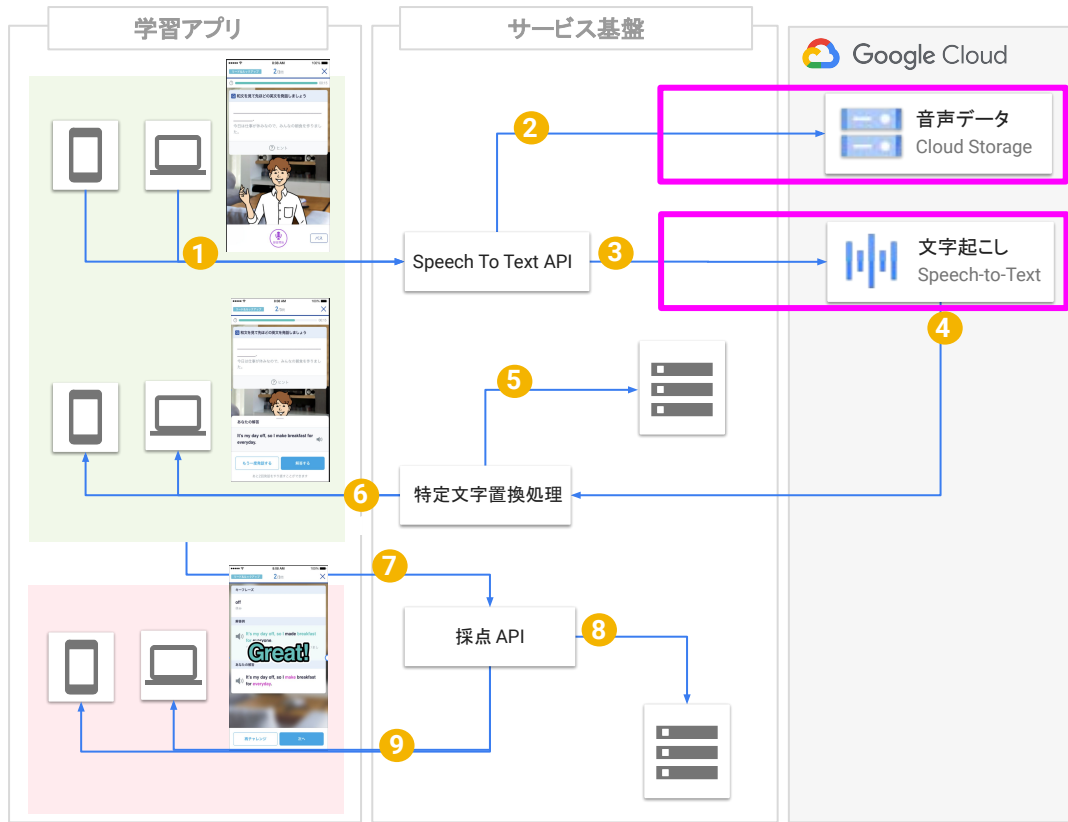
UI/UX に関する速度的な懸念  
「文字起こし」+「採点」  
を同一処理で実施すると、結果描  
画に時間がかかる懸念あり



処理を分離することで、ユーザー  
の待ち時間軽減を実現

Google Cloud

# 工夫した点②(文字起こしと格納の非同期処理)



「文字起こし処理」

と

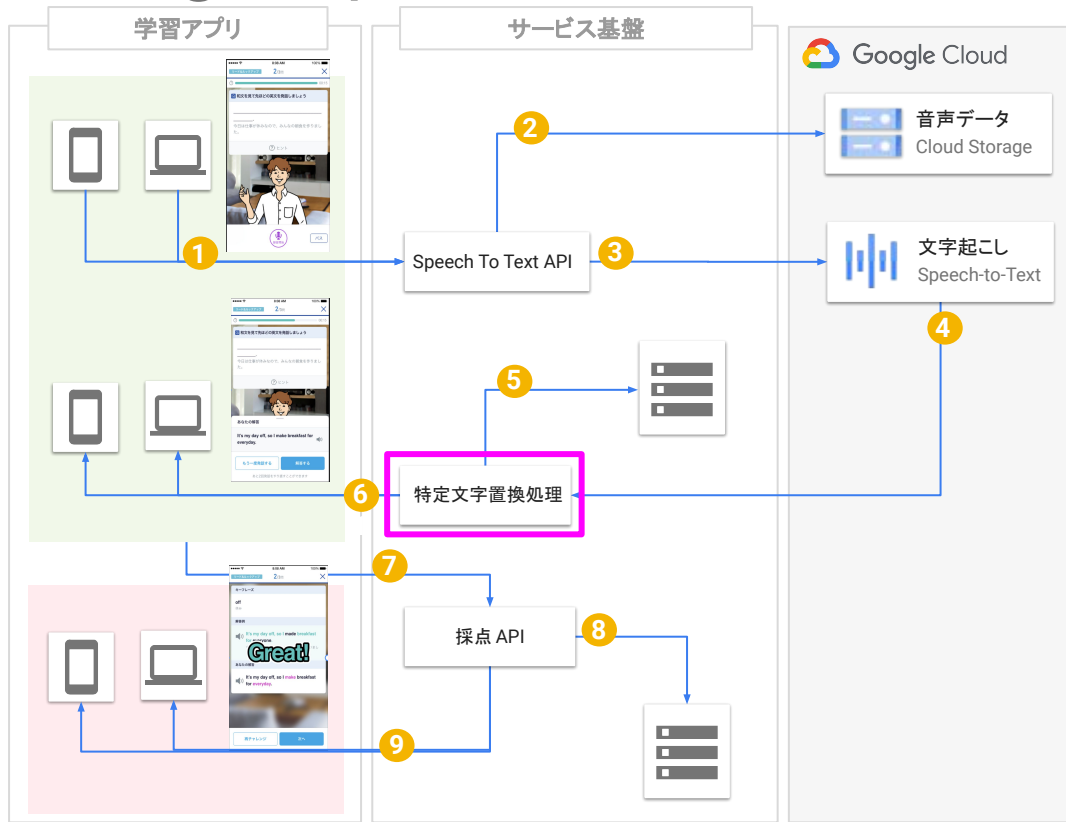
「音声データ格納処理」

を非同期で実施

「学習体験を損ねない速度」を  
実現しつつ、振り返りを可能と  
するための「学習データの格  
納」も実現

Google Cloud

# Google Speech-to-Textへのリクエスト



文字起こしできない部分も存在

- 日本の地名/人名
- 同音異義語 (“too”, “two”)
- などは、まだまだ文字起こしできない

→サービス内で置換辞書を準備  
→文字起こし結果に対して適用  
→今後よりよくなっていく見込み

今後、是非日本の地名や人名、同音異義語も適切に文字起こしできるように対応いただきたい

Google Cloud

# 今後の展望

# 最後に：今後の展望

- リード&ルックアップに限らず、他トレーニングへの Speech-to-Text 適用も検討
- 将来的にはユーザーが自由に発話した内容をより多面的に評価できる仕組みを構築
- この機会に是非、スタディサプリENGLISH をお試しください！



# Thank you

