SRed 기술서 - 딥러닝을 이용한 언어장애 음성 변환기

정상적인 발음이 불가능한 **언어장애인**도 **AI 스피커** 등 발전한 음성인식 기술의 혜택을 누릴 수 있도록 **딥러닝**을 이용하여 **언어장애용 음성 변환기 (Speech Interpreter)** 제작을 목표로 한다. 음성 변환기는 **사용자의 음성 데이터**를 **학습**하며 이에 해당하는 문장(**올바른 문장**)을 소리로 출력한다.

(딥러닝을 이용하기 위해서는 특정 레이블에 대한 학습 데이터에 항상 공통된 특징이 존재해야 한다. 가령 ‘거실 불 켜’ 문장에 대한 학습 데이터가 입력될 때 마다 너무 다른 특징을 지닐 경우 학습이 불가능하다. 따라서 사용자가 표현하고자 하는 문장에 따라 규칙적인 소리를 말할 수 있어야 한다.)

**음성 변환 시스템**은 **AI 스피커에서 사용되는 특정 문장**에 해당하는 **사용자의 음성**을 **딥러닝**으로 **학습**시킨 **모델**을 이용한다. 사용자는 **스마트폰** **앱**을 통해 시스템 이용이 가능하며 **앱**에서 **사용자의 음성**을 **음성인식 모델**이 있는 **서버**(미정)로 전송한다. **앱**에서 **전송된 음성 데이터**를 가지고 **음성인식 모델**을 **학습** 시키거나 **올바른 문장**으로 변환하는 과정은 **서버**에서 이루어진다.

**음성 변환기**는 표준발음이 불가능한 언어장애인이 사용하는데 언어장애인마다 같은 문장을 다르게 발음하기 때문에 **음성인식 모델**이 **사용자** 개인에 맞추어 학습 및 적용되어야 한다. 이를 위해 **회원제**를 이용하며 **베이스 모델**(사용자 음성 데이터 학습 전 상태의 음성인식 모델)을 기반으로 사용자마다 **개인화된 음성인식 모델**을 구축하고 그 정보를 **데이터베이스**에 저장한다. (회원정보는 데이터베이스에, 음성인식 모델은?)

**앱** 최초 실행 시(비로그인 상태) ‘**로그인/회원가입 화면**’이 출력되며 **사용자**는 회원가입 후 로그인하여 **앱**을 이용할 수 있다. **앱**에 로그인하면 ‘**학습 화면**’이 출력된다. **학습 화면**에는 **AI 스피커에서 사용되는 문장**이 출력되며 사용자는 문장에 대한 자신의 **음성**을 일정 횟수 반복적으로 입력한다. (문장 개수 미정, 조사 필요, 50개 내외 예상 / 음성 입력 횟수 미정, 5~10회 정도 예상, One-shot learning / 사용자가 원하는 문장을 입력하고 이를 학습시켜 사용할 수 있는 기능 추가도 생각해볼 필요) 한 **문장**에 대한 입력이 끝나면 순차적으로 다음 문장이 출력되며 **음성**을 입력하는 과정을 반복한다. 모든 문장에 대해 입력이 끝나면 이를 학습한 **개인화 모델**이 **서버**(혹은 DB?)에 저장되며 **음성 변환 시스템** 이용이 가능해진다.

이후(로그인 상태, 학습된 개인화 모델 존재) **앱**의 메인 화면은 ‘**사용 화면**’이 된다. **사용 화면**에서 **음성**을 입력하면 **개인화된 모델**을 통해 **변환된 문장**이 화면에 **텍스트**로 표시된다. 또한, **TTS**를 통해 이를 **소리**로 출력하게 되며 이로 인해 **AI 스피커** 이용이 가능하다. (추가 학습 방법은 미정) 사용 화면에서 메뉴를 통해 **추가 학습**을 진행할 수 있으며 초기 학습과 유사한 방법으로 수행된다. 추가 학습이 이루어짐으로써 변환 시스템의 **정확도** 향상을 기대할 수 있다.