



음성 인식(Speech Recognition)의 기본

음성 인식(Speech Recognition)이 무엇인가요?

음성 인식의 기술은 학생들의 음성 입력을 원어민 스피커 모델과 비교하여 점수를 매깁니다. 입력이 목표 모델 문장과 충분히 비슷하면 인식이 됩니다. 유사성 레벨이 0.2 이상이면 문장이 인식됩니다. DynEd 는 시스템이 너무 까다로움으로 인해 학습자들이 힘들어하지 않도록 최소 레벨을 정했습니다. 이는 억양의 변화도 허용합니다. SR 의 강점은 학생들의 말하기 연습을 권하고 증가된 연습을 통해 능통함과 발음을 향상시켜 줍니다.

인식기는 어떤 음성 입력에 맞춰져 있나요?

DynEd는 원어민처럼 빨리 말하지 못하고, 명확하게 말해야 하는 영어 학습자를 위해 모델을 설정했습니다. 아주 빨리 말하고 명확하게 말하지 않는 원어민은 인식되지 않을 수도 있습니다. 그들은 목표 사용자가 아닙니다.

어떻게 해서 인식기(recognizer)를 속일 수 있는 건가요?

인식기는 음성 입력의 변동을 허용합니다. 또한 학습자들이 모델을 따라 할 때, 답답하지 않도록 충분히 낮은 레벨로 설정되어있습니다. 만약 인식기가 의도적이거나 랜덤한 변동을 찾게 설정된다면, 시스템은 영어 학습자들에게 너무 까다로워집니다. 참고로 SR레벨은 코스와 레슨마다 달라집니다. 고급 레벨 학생을 위해서 코스 디자이너들은 시스템을 더 까다롭게 할 수 있습니다.

음성 측정기(speech gauge)는 무엇을 표시하나요?

음성 측정기는 입력 음성이 얼마나 목표 모델과 일치하는지 표시합니다.

일치성이 떨어지면 빨간 막대가 1개를 표시하고, 가장 낮은 점수를 받은 문장의 위치를 표시합니다. 이것은 학생의 문장의 해당 부분에 집중할 수 있게 도와줍니다. 문장 일치 점수가 0.3에서 0.4사이엔 측정기가 노랑색을 표시합니다.

(아래 그림 참고)



이 예에서는 측정기가 노랑색을 표시하며 학생이 ' for work?' 을 더 명확하게 말하라고 합니다. 참고로 이 초점 정보는 측정기가 녹색 막대를 보일 때는 나타나지 않습니다.

일치성 점수(confidence score)는 무엇인가요?



음성 측정기는 인식기가 입력 음성과 목표가 얼마나 일치하는지 표시합니다. 위에서처럼 녹색 4개가 표시되면 일치성 점수는 0.5 이상입니다. 녹색이 3개 표시되면, 일치성 점수는 0.4에서 0.5사이 입니다. 표시와 일치성 점수의 범위는 아래와 같습니다.

4 Greens	0.5 – 1.0
3 Greens	0.4 – 0.5
2 Yellows with focus hint	0.3 – 0.4
1 Red with focus hint	0.2 – 0.3
??? or no recognition	0.0 – 0.2

일치성 점수는 문장 전체에 대한 점수입니다. 참고로 단어 한 개의 점수가 0점이어도 문장 점수는 합격 점수를 받을 수 있습니다. 이것은 발음이 좋은 사람이 단어를 잘못 바꿨을 때 일어날 수 있습니다. 한 단어가 0점을 받아도 다른 단어들을 명확하게 발음하면 전체적으로 합격점수를 받을 수 있습니다. 원어민이라도 단어 하나를 잘못 말했을 때, 다른 것들을 이해했다면 듣는 사람은 의도했던 의미를 알 수 있습니다.

학생이 어떻게 다시 시도할 수 있나요?

만약 학생이 같은 문장을 다시 녹음하고 다시 인식 받기를 원한다면 음성 측정기의 마이크 버튼을 클릭하면 됩니다. 이는 아주 좋은 습관이며 권장되어야 합니다. 측정기가 빨강이나 노랑색을 표시하면 포커스정보가 학생으로 하여금 어느 부분을 더 명확하게 말해야 하는지 알 수 있게 도와줍니다. 학생들은 또한 자신이 녹음한 것을 듣고 원어민 모델과 비교하는 것이 좋습니다. 매번 똑같이 반복만은 하지 마세요.

데이터는 무엇을 보여주나요?

수년간 수개국에서 수집된 데이터는 다음을 보여줍니다.

1. 학생들은 음성인식 훈련의 사용을 즐기고 그런 레슨을 선호합니다.
2. 학생들은 음성 인식을 사용함으로써 발음과 구두 능통함에 큰 향상이 있습니다.
3. 3개월동안 자주 연습을 하자, 언어 과제가 더 어려워지는데도 많은 학생들은 평균 인식 점수를 50% 이하에서 65% 이상으로 향상시켰습니다.

음성인식에 관한 질문

1. 인식기는 왜 어떤 단어들을 더 까다롭게 처리하나요?

인식기는 레슨 형태에 따라서 핵심 단어나 구조에 집중합니다. 예를 들어 어떤 레슨에서는 인식기는 live와 lives 또는 do와 don't 의 차이에 집중합니다.

다른 레슨에서는 개별 단어가 아니라 어순에 집중합니다. 만약 핵심단어가 틀렸거나 인식이 안되었다면, 문장내의 다른 단어들을 잘 발음하더라도 인식기는 문장을 인식하지 않습니다.

2. 왜 인식기는 맞는 단어를 말해도 문장의 끝에 물음표(?)를 표시하나요?

어떤 사람들은 문장의 끝에서 목소리가 작아지게 합니다. 많은 경우에 인식기는 문장이 완전히 끝났는지 확인하고자 합니다. 그래서 마지막 단어가 명확히 들리지 않는다면 명확히 말을 해도 문장을 인식하지 않습니다.

3. 어떨 때 인식기는 모든 것을 명확하게 말했는데도 녹색 3개만 표시합니다. 왜 4개가 안보이나요?

음성 인식 기술을 아직 한계가 있습니다. 어떤 발음 단위는 다른 것들보다 덜 인식이 됩니다. 학생이 녹색 3개를 받으면 보통 충분히 잘한 것입니다. DynEd는 음성 평가의 신뢰도를 향상시키기 위해 지속적으로 인식기를 조절하고 있습니다. 어떤 목소리는 다른 목소리에 비해서 인식이 더 잘되기도 합니다.

어린이보다는 성인이 인식이 더 잘됩니다. 이는 원어민 모델이 일반적으로 성인 말하기에 기반하고 있기 때문입니다. 인식기가 녹색 3개나 4개를 표시하면 포커스 정보는 제공되지 않습니다.

4. 인식기가 음성의 높낮이나 강약도 측정하나요?

이러한 패턴의 변화가 크고 감지하기가 어렵습니다. 현재로서는 인식기는 이러한 특성을 감지하지 않습니다. 그러나 너무 강하거나 약한 발음은 단어를 빠뜨리거나 왜곡할 수 있으며 이는 불량한 인식으로 이어집니다.

5. 인식기는 왜 특정한 경우나 특정한 컴퓨터에서는 작동이 잘 안되나요?

인식기는 좋은 마이크를 필요로 하고 적은 배경 소음을 원합니다. 학생은 또한 입바람이 왜곡을 하기 때문에 마이크를 입에서 떨어뜨려야 합니다. 마이크를 입 앞에서 떨어뜨리고 약간 옆으로 위치시킵니다. 컴퓨터의 마이크 설정도 문제입니다. 어떨 때는 마이크 출력 지원이 필요하며 기본 컴퓨터 설정을 바꿔줘야 합니다.