

LESSON 14

중급 초6·중2 AI 교실

AI는 내 목소리를 어떻게 알아들을까? 음성 인식

목소리의 떨림을 숫자로 바꾸고, 소리 조각을 맞히고, 문장으로 다듬어요.

🎯 학습 목표 이번 시간이 끝나면...

- ✓ 음성 인식의 4단계(소리를 숫자로 → 짧게 자르기 → 소리 단위 찾기 → 문장으로 다듬기)를 설명할 수 있어요.
- ✓ 파형과 음소의 뜻을 말할 수 있어요.
- ✓ 소음, 말 빠르기, 학습 데이터가 정확도에 주는 영향을 알 수 있어요.

• 중급 14

AI는 내 목소리를 어떻게 알아들을까? 음성 인식

목소리도 숫자로 바뀌어요.

소리 조각을 글자로 맞추고 앞뒤 말로 다듬어요.



“시끄러워도
앞뒤 말로 알아들었어!”

음성 인식도 앞뒤 말로 글자를 짐작해요.



📖 꼭 알아야 할 단어

파형 = 소리의 떨림을 그린 모양

음소 = 말소리를 이루는 가장 작은 소리 단위

Q. 음성 인식은 주변이 시끄러워도
항상 똑같이 정확하다?

☐ O / ☒ X 정답: X

그림 1 한눈에 보는 14강 — 소리 그래프(파형)가 글자가 되기까지를 따라가 보세요.



쉽게 풀어 보기

내 목소리가 화면 속 글자가 되는 과정을 알아봐요.



목소리도 숫자가 돼요

말을 하면 공기가 떨려요. 마이크는 이 떨림을 받아 **파형(waveform)**, 곧 소리의 떨림을 그린 모양으로 바꾸고, 다시 숫자로 저장해요.

음성 인식 AI는 이 숫자를 1초보다 훨씬 짧은 조각으로 잘라요. 그리고 조각마다 어떤 소리에 가까운지 맞혀요.

이때 맞는 소리 단위가 **음소(phoneme)**예요. 음소는 말소리를 이루는 가장 작은 소리 단위예요. 예를 들어 '밥'은 **ㅃ**, **ㅏ**, **ㅂ**으로 이루어져요.



시끄러운 운동장에서 알아듣기로 이해해요

쉬는 시간 운동장은 시끄러워서 친구 말이 띄엄띄엄 들려요. 그래도 "...이따 ...구 하자!"라고 들리면 "이따 축구 하자!"라고 알아듣지요. **들린 소리 조각에 앞뒤 말과 상황을 더해 빈칸을 채운** 거예요.

음성 인식 AI도 소리 조각을 맞힌 다음, 앞뒤 말로 가장 그럴듯한 문장을 골라요.

그래서 소리가 너무 작거나 시끄러우면 **영똥한 문장을 고르기도** 해요.



한 줄 요약

마이크는 목소리의 떨림을 숫자로 바꿔요. 음성 인식 AI는 이 숫자를 아주 짧은 조각으로 잘라 어떤 소리인지 맞히고, 앞뒤 말을 살피어 알맞은 문장으로 다듬어요.

생활 속 음성 인식

키보드의 마이크 버튼으로 말을 글자로 받아 적기, 영상의 자동 자막, AI 비서, 자동차의 음성 명령이 모두 음성 인식이에요.

손이 바쁘거나 글자를 치기 어려운 사람에게 큰 도움이 돼요. 회의나 수업 내용을 글로 남기는 데도 쓰여요.

하지만 조용하고 또렷할수록 정확하고, **소음, 여러 사람이 동시에 말하기, 아주 빠른 말**에서는 틀리기 쉬워요. 받아 적힌 글은 꼭 다시 읽고 고쳐요.

미리 보기 뒤에 나오는 '직접 해보기 미션'에서 같은 문장을 세 가지 상황에서 말해 비교해 봐요.

미리 생각하기 음성 인식이 틀리기 쉬운 상황을 하나 더 떠올려 적어 보세요.

예: 바람 부는 바깥, 마스크를 쓰고 말할 때

123 숫자로 보는 14강

사실

1,000분의 1초

1밀리초(ms)의 길이. 음성 인식은 소리를 수십 밀리초 조각으로 잘라요.

예시

ㅂ·ㅈ·ㅂ

'밥'을 이루는 음소 3개. AI는 조각마다 이런 소리 단위를 맞혀요.

예시

3가지

미션에서 비교할 상황: 조용히
또박또박, TV 소리 켜고, 아주
빠르게

사실 = 실제로 있었던 일 · **예시** = 이해를 도와려고 만든 숫자



표로 정리해요

목소리가 문장이 되는 흐름을 정리해요.



음성 인식의 4단계

단계	하는 일	“밥 먹자” 예시
① 소리를 숫자로	마이크가 공기의 떨림을 파형(소리 그래프)으로 바꾸고, 이를 숫자로 저장해요.	“밥 먹자” 소리가 파형이 돼요.
② 짧게 자르기	소리를 1초보다 훨씬 짧은 조각으로 잘라요.	아주 짧은 조각 여러 개
③ 소리 단위 찾기	조각마다 ‘ㅂ’, ‘ㅌ’ 같은 소리 단위 중 무엇에 가까운지 맞혀요.	ㅂ, ㅌ, ㅂ, ㅌ, ㅌ...
④ 문장으로 다듬기	앞뒤 말을 살피어 “밥 먹자”처럼 알맞은 낱말과 문장으로 바꿔요.	“밥 먹자” (“밤 먹자”가 아니라)



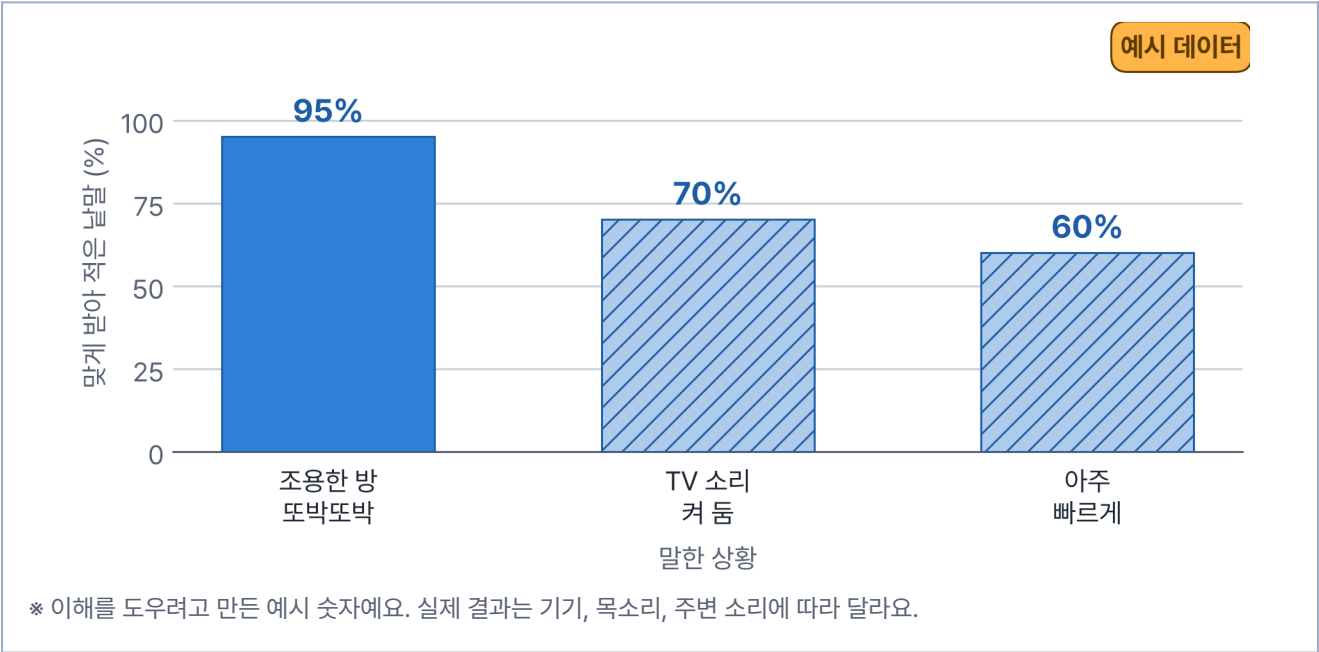
잘 알아들을 때와 틀리기 쉬울 때

살펴봐요	잘 알아들어요	틀리기 쉬워요
주변 소리	조용한 곳	소음이 큰 곳, TV 소리
말하는 방법	또박또박, 알맞은 빠르기	아주 빠르게, 웅얼웅얼
말하는 사람	한 사람씩	여러 사람이 동시에
말의 종류	학습 데이터에 많은 말	데이터가 적은 사투리, 새 낱말

생각해 보기 영상 자동 자막이 틀렸을 때, 보는 사람에게 어떤 문제가 생길 수 있을까요?

자막에 기대는 사람을 떠올려 봐요.

 상황에 따라 맞게 받아 적은 낱말 비율 (예시)



그래프 읽기 어떤 상황에서 가장 정확했나요? 정확도를 높이려면 어떻게 말하면 좋을까요?

이 그래프는 예시 데이터예요. 미션 결과와 비교해 봐요.

가 꼭 알아야 할 단어

파형

waveform

소리의 떨림을 그린 모양

음소

phoneme

말소리를 이루는 가장 작은 소리 단위 (예: ‘밥’은 ㅂ, ㅏ, ㅍ)

**직접 해보기 미션**

손과 눈으로 직접 확인해 봐요.

세 가지 상황에서 음성 입력하기

보호자와 함께 스마트폰 키보드의 음성 입력(마이크 버튼)을 켜고, 같은 문장 **“내일 아침에 운동장에서 만나자”**를 세 번 말해 보세요. 받아 적힌 글자가 어떻게 달라지는지 비교해요.

- 1 조용한 방에서 또박또박 말해요.
- 2 텔레비전 소리를 켜 둔 채로 말해요.
- 3 아주 빠르게 말해요.

※ 새로 가입할 필요는 없고, 이름이나 주소 같은 개인 정보는 말하지 않아요.

기록표

	받아 적힌 글자	틀린 낱말 수
① 조용히 또박또박		
② TV 소리 켜고		
③ 아주 빠르게		

- 가장 많이 틀린 상황은 무엇이었나요? 왜 그랬을까요?

잘못 알아들은 낱말과 비슷한 소리

음성 입력을 쓸 때 나의 약속

🔑 정리 음성 인식은 소리 조각과 앞뒤 말로 문장을 짐작해요. 그래서 상황에 따라 정확도가 달라져요.

**확인 퀴즈**

스스로 풀고, 정답은 마지막 쪽에서 확인해요.

1. **O/X** 음성 인식은 주변이 시끄러워도 항상 똑같이 정확하다.☐

O

☐

X

2. **고르기** 음성 인식이 틀리기 쉬운 경우가 아닌 것은?☐

① 주변 소음이 클 때

☐

② 여러 사람이 동시에 말할 때

☐

③ 조용한 곳에서 또박또박 말할 때

3. **빈칸** 말소리를 이루는 가장 작은 소리 단위를 _____(이)라고 해요.4. **쓰기** 사투리나 어린이 목소리에서 음성 인식이 덜 정확할 수 있는 까닭을 써 보세요.

**중2 한 걸음 더**

스펙트로그램과 목소리 데이터

음성 인식 AI는 소리를 보통 수십 밀리초(1밀리초는 1000분의 1초) 길이의 조각으로 자른 뒤, 높고 낮은 소리가 얼마나 섞였는지를 그림처럼 나타낸 '**스펙트로그램(spectrogram)**'으로 바꿔 분석해요. 요즘은 소리에서 글자까지 한 번에 바꾸는 방식도 많이 쓰여요.

사투리나 어린이 목소리처럼 학습 데이터가 적은 말은 정확도가 낮을 수 있어서, 다양한 목소리 데이터가 필요해요. **목소리도 나를 알아볼 수 있는 개인 정보**이니 함부로 녹음해 올리지 않아요.

- 생각해 봐요: 다양한 목소리 데이터를 모을 때 지켜야 할 약속은 무엇일까요?



용어 정리

이번 시간에 나온 어려운 말을 모았어요.

용어	영어	쉽게 말하면	예시
음성 인식	speech recognition	목소리를 글자로 바꾸는 기술	음성 입력, 자동 자막
파형	waveform	소리의 떨림을 그린 모양	오르내리는 소리 그래프
음소	phoneme	말소리를 이루는 가장 작은 소리 단위	ㅂ, ㅌ, ㅍ
마이크	microphone	소리의 떨림을 전기 신호로 바꾸는 장치	스마트폰 마이크
소음	noise	듣고 싶은 소리를 가리는 다른 소리	TV 소리, 바람 소리
밀리초	millisecond	1,000분의 1초	20밀리초 조각
스펙트로그램	spectrogram	높고 낮은 소리가 섞인 정도를 그림으로 나타낸 것	소리의 지문 같은 그림
맥락	context	앞뒤 말과 상황	“이따 ...구 하자” → 축구
자동 자막	auto caption	영상 속 말을 자동으로 글자로 보여 주기	영상 앱 자막
사투리	dialect	지역마다 다른 말씨	경상도, 전라도 말씨



정답과 풀이

퀴즈를 다 풀 다음에 확인해요

1. X — 소음이 크면 소리 조각을 잘못 맞춰 틀리기 쉬워요.
2. ㉓ — 조용하고 또렷할수록 정확해져요.
3. 음소(phoneme) — ‘밥’은 ㅂ, ㅌ, ㅍ 세 음소예요.
4. (예시 답) 학습 데이터에 그런 목소리가 적어서, AI가 그 소리의 특징을 충분히 배우지 못했기 때문이에요.