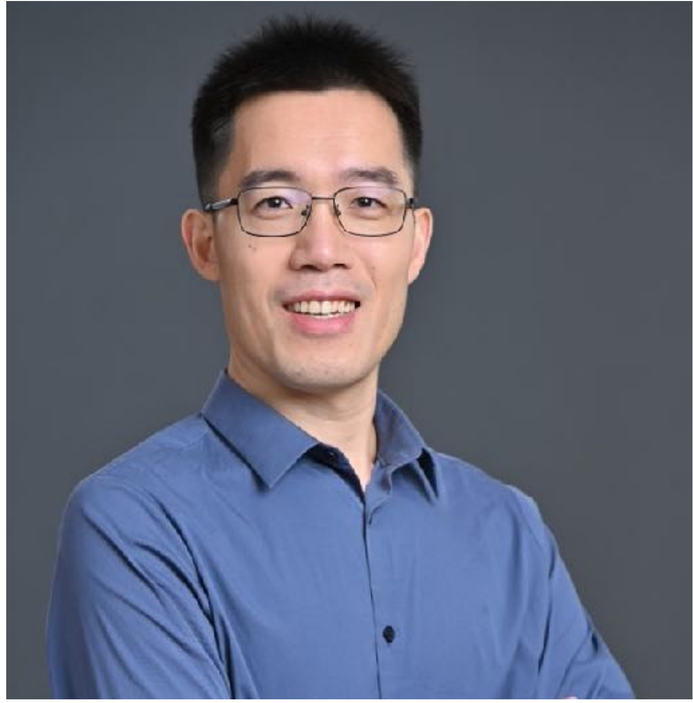
Fundamentals of Speech and Language Processing

CSC3160

Zhizheng Wu

Lecture 1: Course introduction

Instructor



- Associate professor joined in Aug 2022
  - https://drwuz.com/
  - Email: wuzhizheng@cuhk.edu.cn

- Ex-Meta, ex-Apple, ex-Microsoft, ex-JD.COM

- Associate editor of IEEE/ACM Transactions on Audio Speech and Language Processing
- General Chair: IEEE Spoken Language Technology 2024
- Member of the IEEE Speech and Language Processing Technical Committee
- Co-founder of ASVspoof challenge, voice conversion challenge
- Organizer of Blizzard challenge 2019
- Founder of Amphion

Amphion

- An Open-Source **A**udio, **M**usic and S**p**eec**h** Genera**tion** Toolkit
  - Educational purpose
  - Producible research and fair comparison
  - Targeting audiences/users
    - Undergraduate and postgraduate students
    - (Research) Engineers who want to work on audio/music/speech generation

Amphion: Recognitions

drwuz Paper author 11 days ago

The corresponding repo: https://github.com/open-mmlab/Amphion

Also we are working hard on HF demos, checkout this week: https://huggingface.co/amphion

13 +

julien-c 10 days ago

Hugging Face联合创始人、CTO

@drwuz looking forward 🔥

Jason Calacanis · Following

I invest in 100 new startups a year...
get a meeting with my team at launc...

Book an appointment

16h •

Sunny Madra just blew my mind on This Week in Startups with a new AI demo.

Watch as he demos Amphion's open-source Singing Voice Conversion Model on Hugging Face.

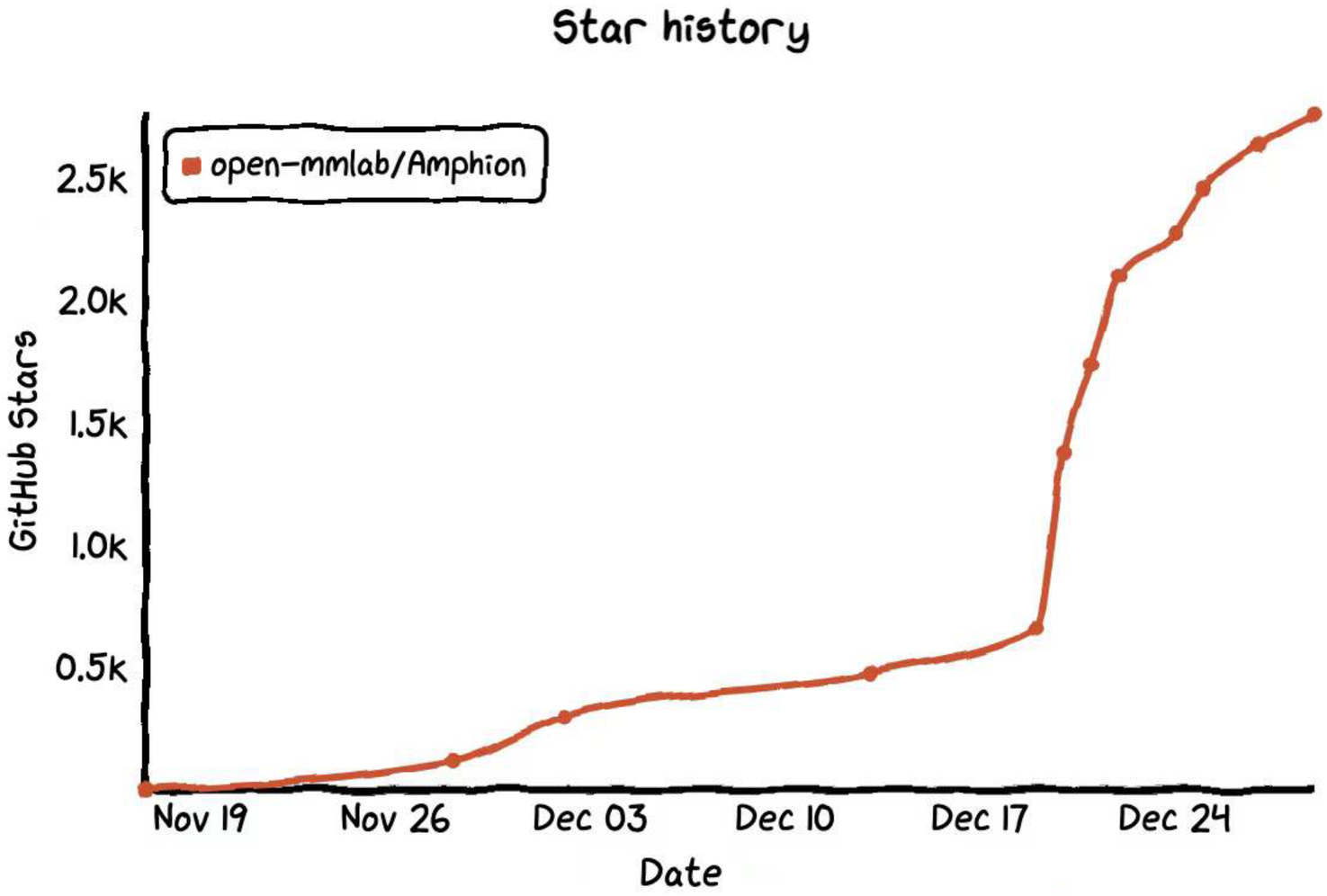
It perfectly converts Adele's 'Someone Like You' to John Mayer's vocals.

It's 70% there off the bat.

There is incredible revenue potential here.

Imagine if Spotify figures out how to integrate this into their product, and for artists to automatically generate more royalties from leasing out their voices.

A+! Bravo!



Amphion: Recognitions

Jason **Calacanis**

@jason

Sunny **Madra**

@sundeep

Course logistics

- Instructor: Zhizheng Wu
- TA: Li Wang

- Course website: https://drwuz.com/CSC3160/
  - Discord: https://discord.gg/8REEWxE7RH

- Lecture time and location
  - Monday/Wednesday 10:30AM - 11:50AM in TA107
- Tutorials: There is NO plan to have tutorials

- Office hours
  - Zhizheng Wu: **Wed 9:00-10:00 AM**. TXC715

What I am proud of previous CSC3160?

- Three ICASSP papers
  - Yicheng Gu, Jiaqi Li: Their first-author papers accepted by ICASSP 2024, supporting them to South Korea to attend the conference
  - Jiaqi Li, Yuhao Luo, Jiahao Zheng: They contributed to an ICASSP 2024 paper led by my PhD student
- Internship
  - Ting Wang received an internship offer from UBTech during the poster session
  - Yicheng Gu received invitation from Tencent but declined the opportunity
  - Jiaqi Li: Recommended to intern with Microsoft Research USA
- Open-source Amphion
  - Zihao Fang, Haopeng Chen partcipated in Amphion as core members

Communication and feedback

- We will send out two course feedback surveys during the semester (0.5% credit each)

- Feel free to send me or TAs any feedback regarding the course
  - Both the instructor and TAs can **possibly make mistakes**! Communication will help

- Email is the preferred way for communication. BB is encouraged.

> Hands down top five instructors I've ever had. Prof. Wu taught really clearly & simply and he's also very enthusiastic, thus keeping me motivated throughout the semester. He's also very open to questions and feedback unlike any other instructor I had before. Prof. Wu as an academic and professional has inspired me since the start of this semester. As for the course, I think this course is well–built. Workload is still okay, and the exams too. No suggestions from me for now.

Presuming prior knowledge

- **Solid background** of python programming
- Knowledge of **statistics** is a plus
- Self-motivated

Grading (details are available on course website)

- Assignment (40%)
- Midterm exam (25%)
- Final exam (30%)
- Participation (5%)
  - Guest lecture attendance
  - Course evaluation

- AIR 6063: No final exam, but needs to work on a project

Workload

- If you just want to get an A or earn fundamental knowledge of speech and language processing, the workload is NOT heavy. **No more project this semester.**

- If you want to have a career in speech and language processing area, please spend more time on your own.
  - If you want some guidance, I am happy to do that via Amphion or other projects.

Assignment (40%)

- Assignment 1 (10%)
  - Speech alignment and audio synthesis
- Assignment 2 (10%)
  - Text processing
- Assignment 3 (10%)
  - Word embedding and classification
- Assignment 4 (10%)
  - TBD

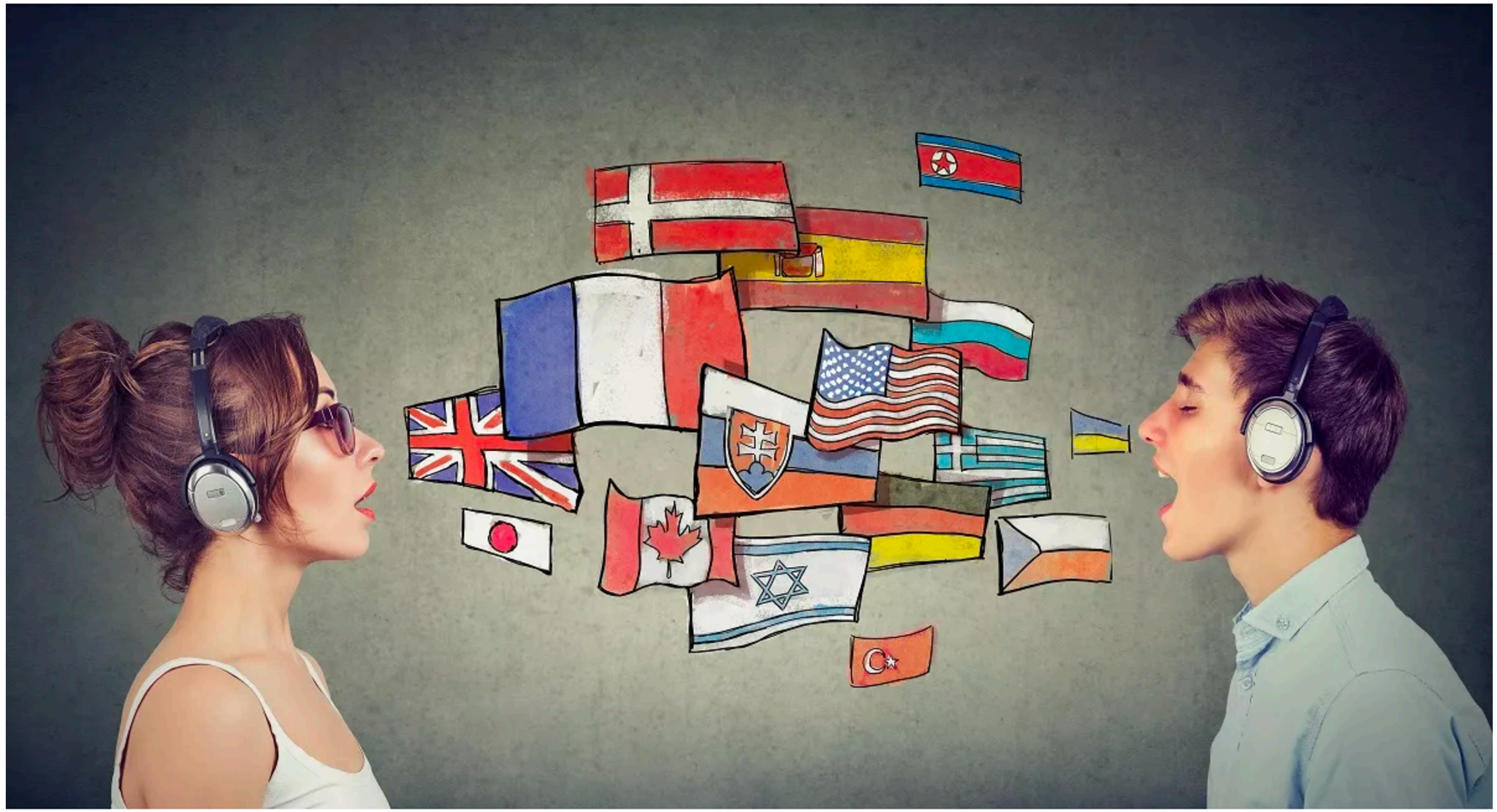
- **All assignments will be available by Jan 17th**

Honesty code

- Strict zero-tolerance policy for cheating or plagiarism
  - Discussions are encouraged, but not sharing code or copying code
- We will use software to detect plagiarism automatically

- Scenarios
  - A shares code with B, and B directly used the code for their assignment. **Both zero**
  - A and B directly copies from internet independently. **Both zero**

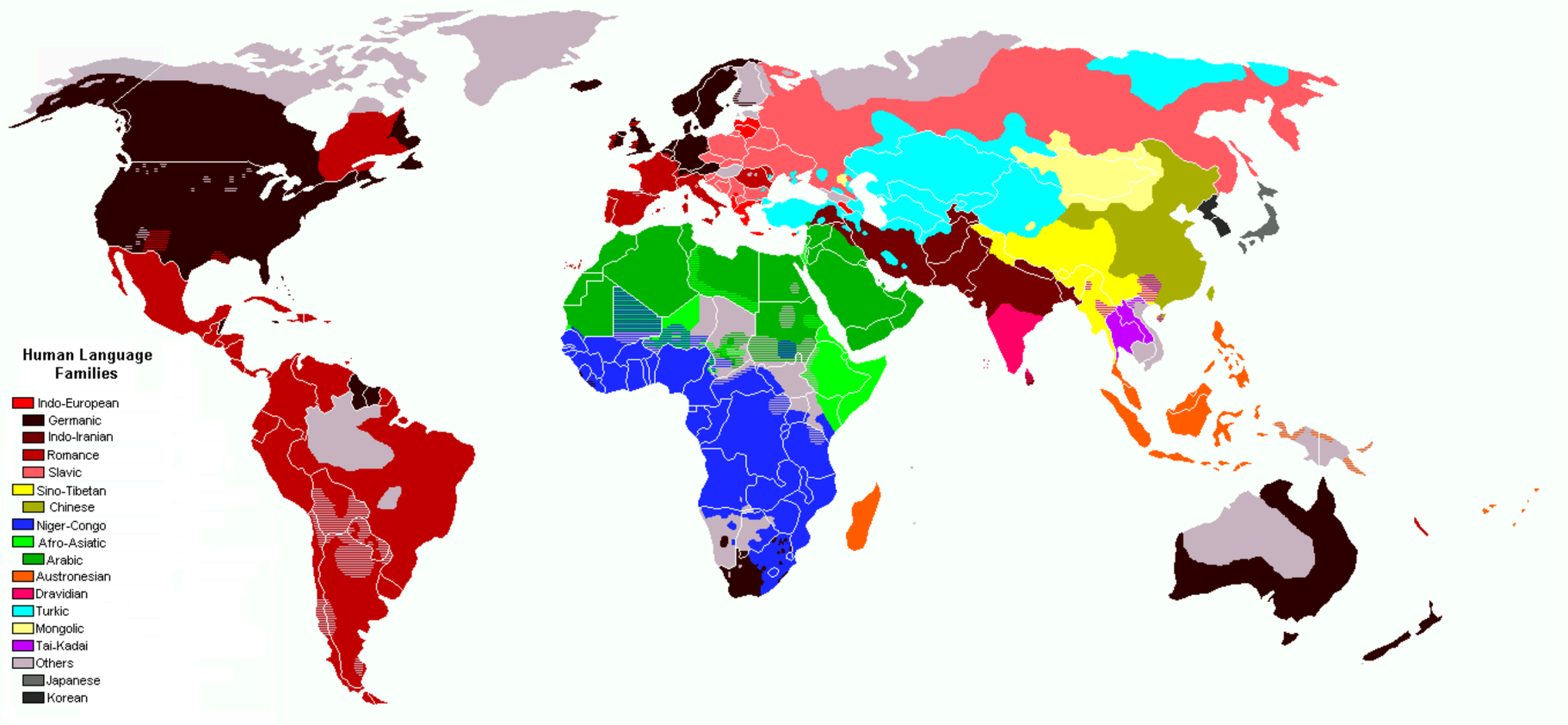
Human language



https://news.yahoo.com/did-humans-first-start-speak-152613750.html

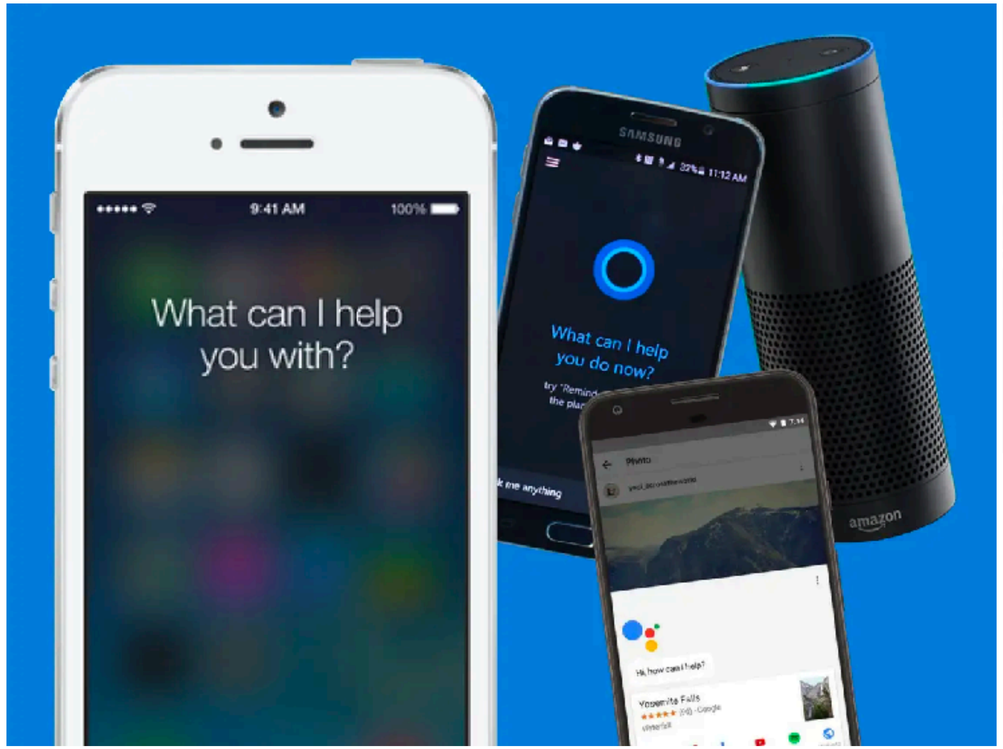
Languages in the world

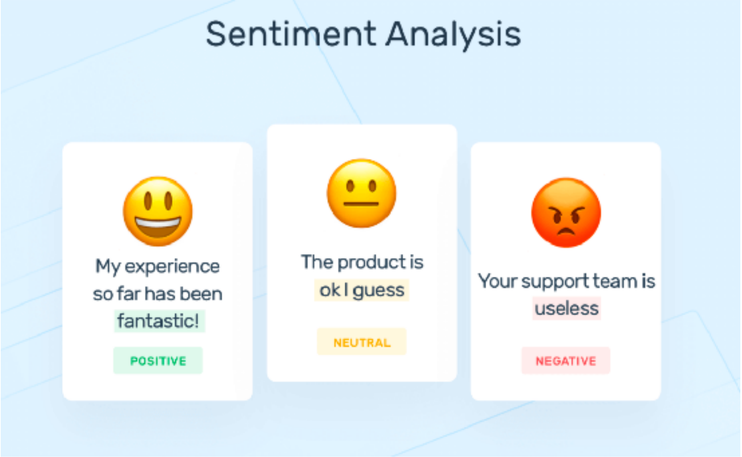
- About 7,000 languages spoken as of 2010. More than half of them have no written form



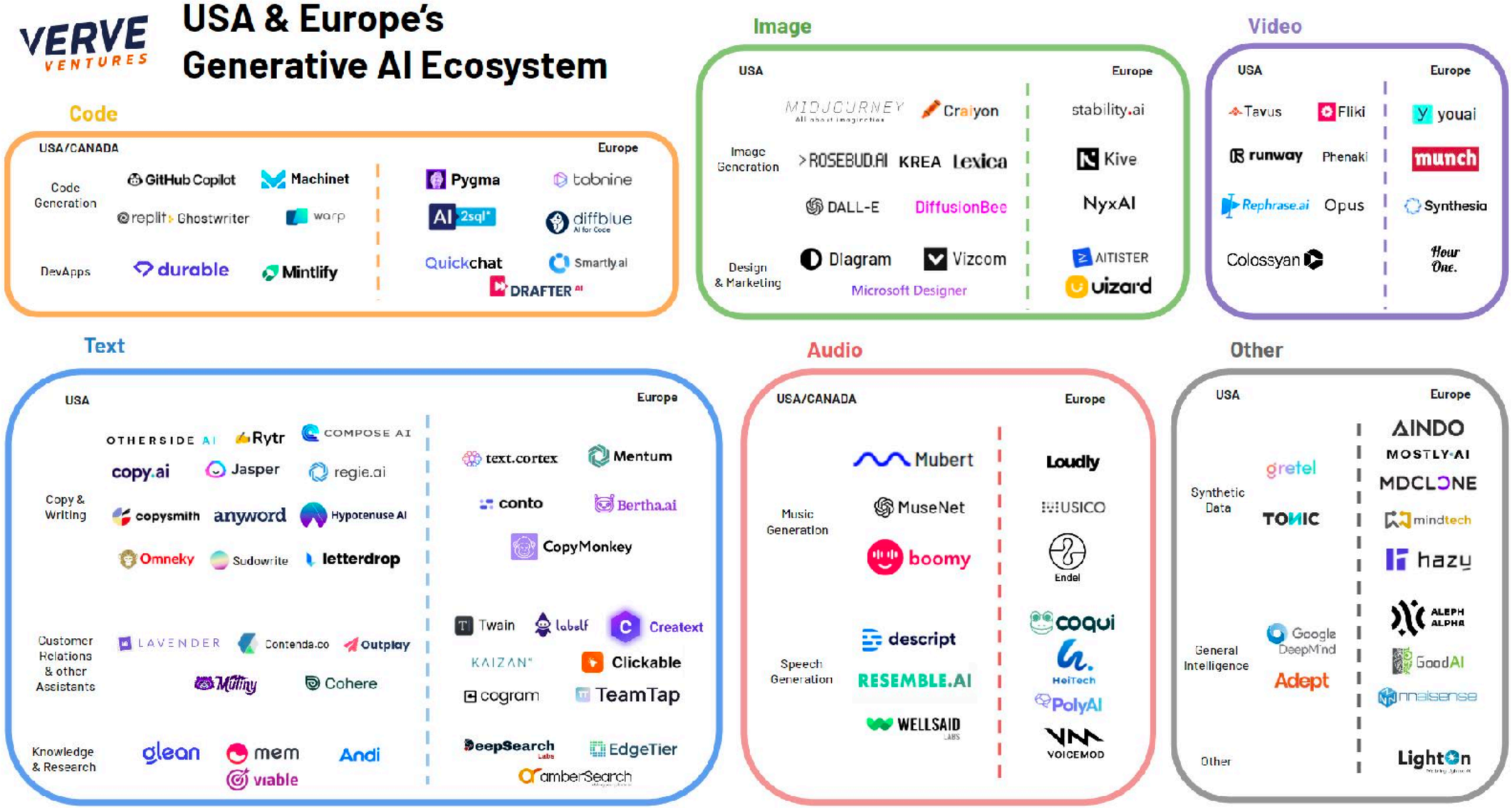
https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers

Applications of speech and language processing





Applications: Generative AI startups



What is this course about?

- Natural language can be ***speech*** or ***text,*** in other words, in ***spoken*** form or ***written*** form (NLP != Text processing)

- First half: Fundamentals knowledge of speech signals and language elements
  - Fundamentals of speech processing
    - Spectrogram, prosody, pronunciation, etc
  - Fundamentals of text processing
    - Language models, word embedding, syntax, tokenization, etc

- Second half: Applications of speech and language processing
  - Speech recognition, synthesis, question answering, chatbot, etc

One slide overview

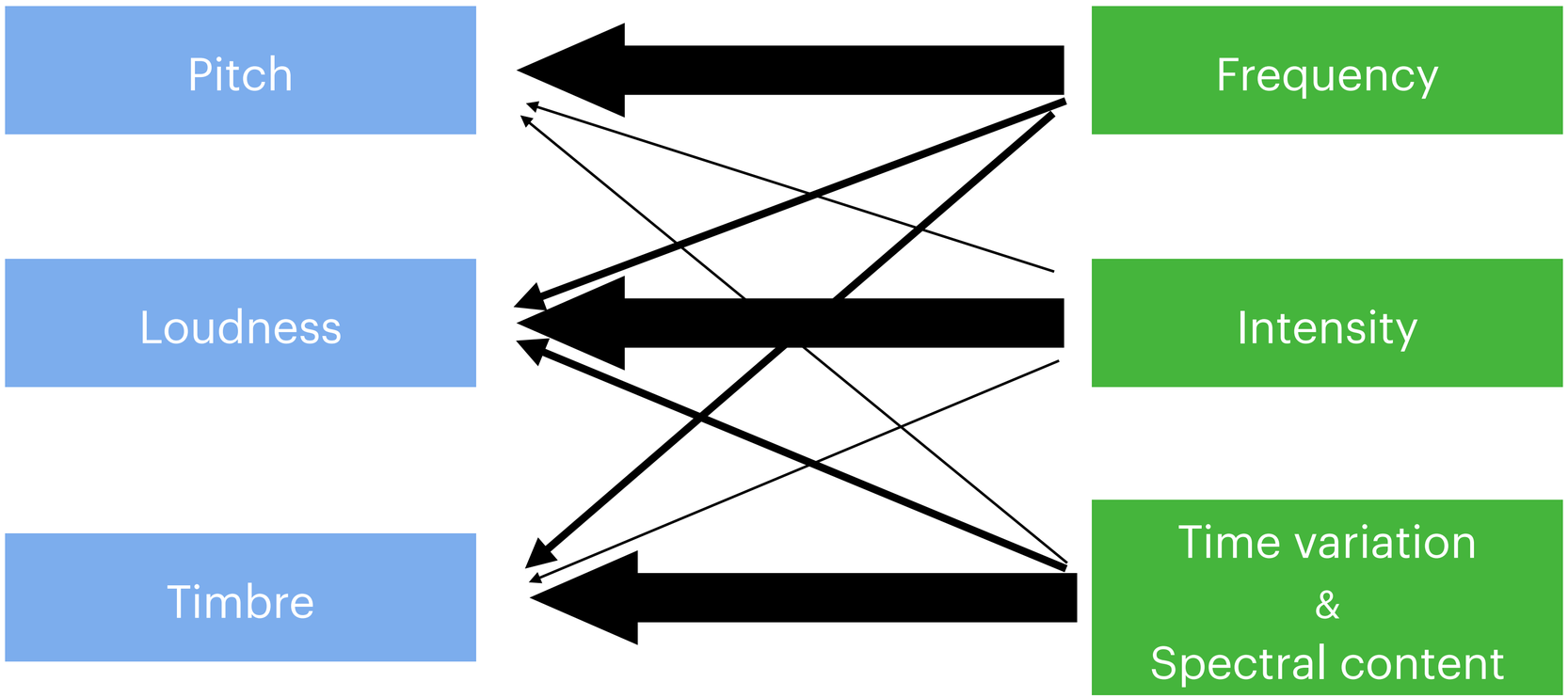
Applications

- Named entity recognition
- Speech recognition
- Machine translation
- Sentiment analysis
- Text-to-speech synthesis
- Question answering
- Text summarization
- Voice conversion
- Chatbot

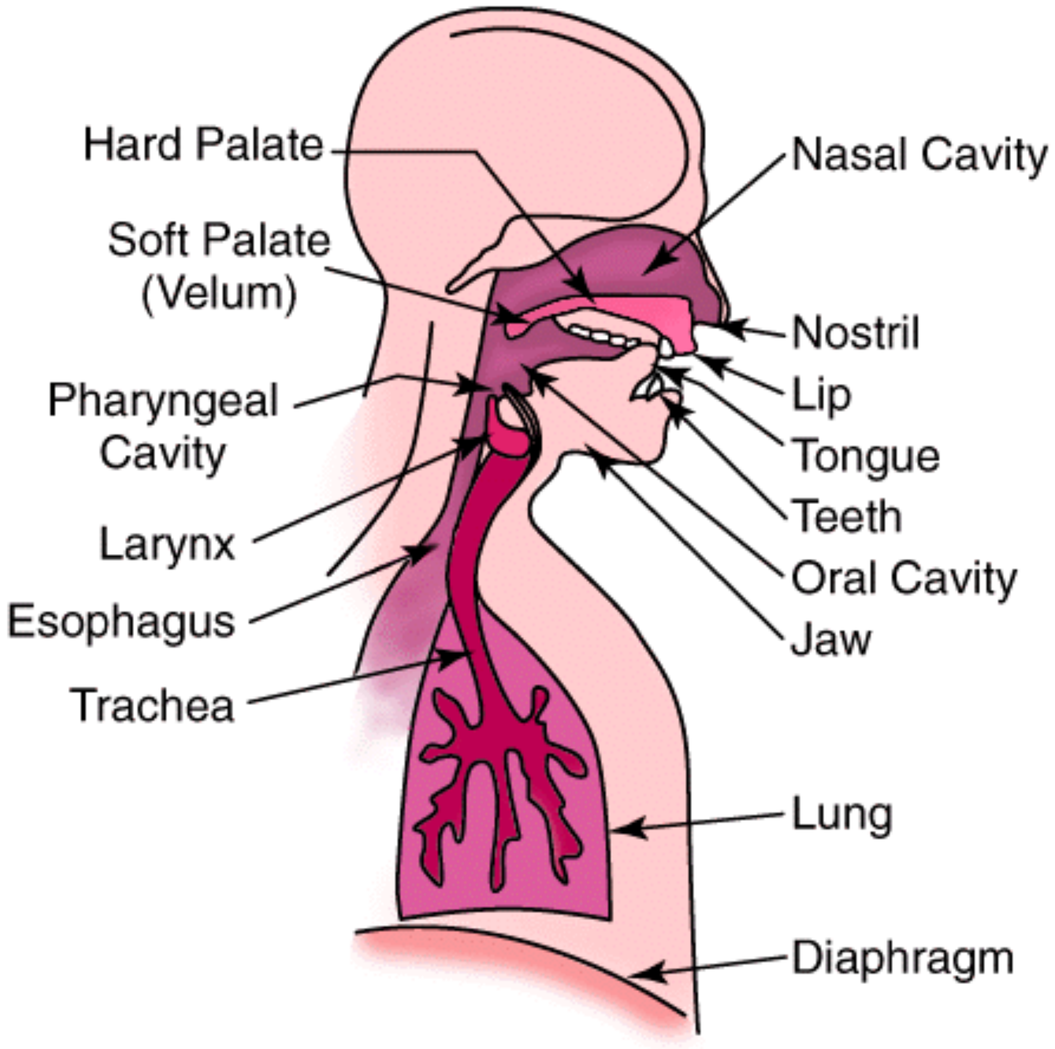
Fundamentals

- Basics of speech processing
- Basics of language
- Language models
- Phonetics
- Prosody and timbre
- Morphology
- Parts of speech
- Semantics and embedding
- Text normalization
- Syntax and parsing

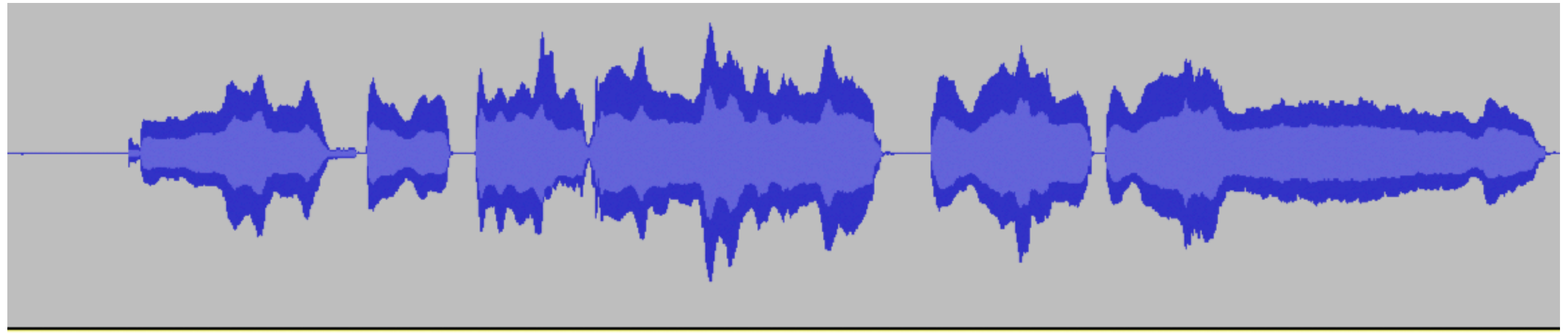
Lecture 2 - 3: Sounds and signal analysis

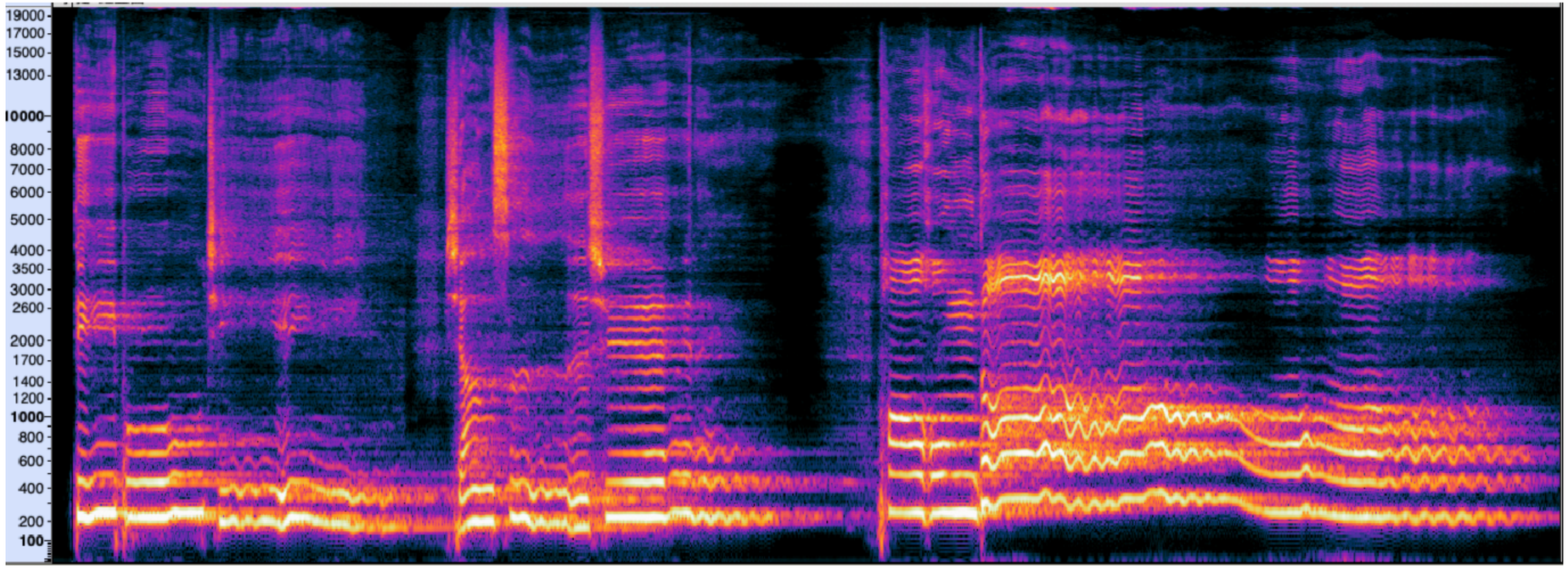


Lecture 4 Speech production

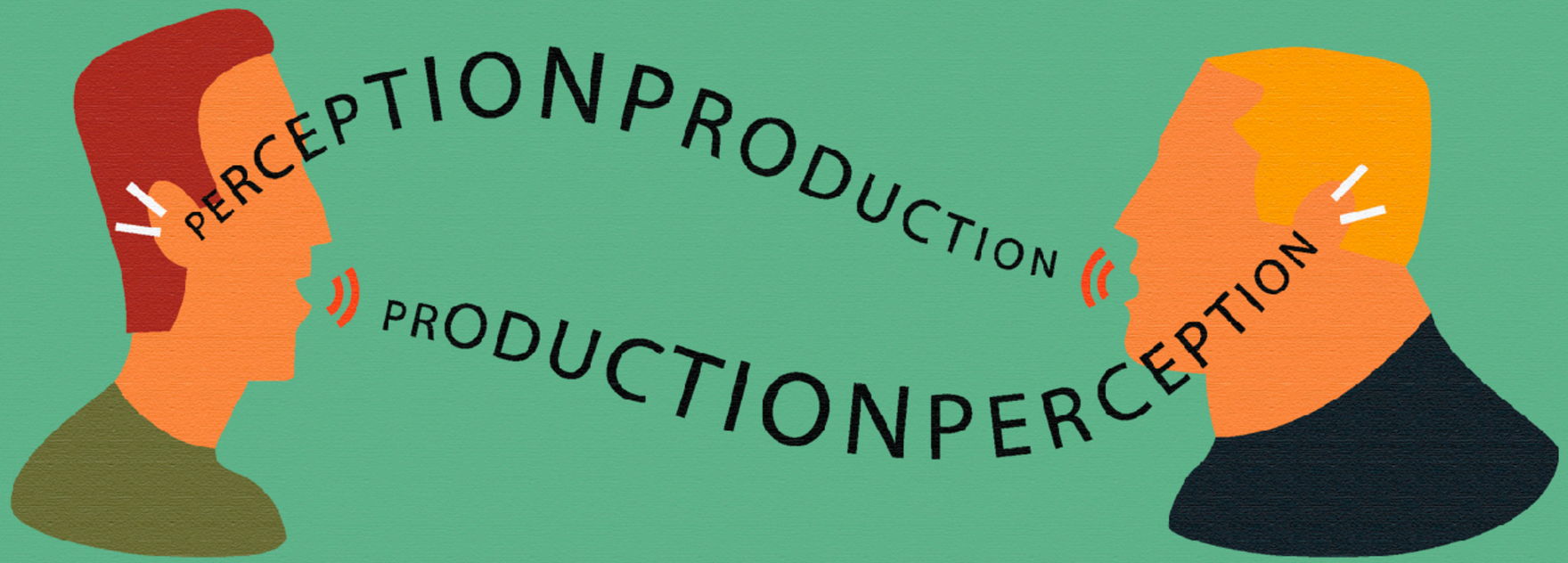


Lecture 5 Speech representation

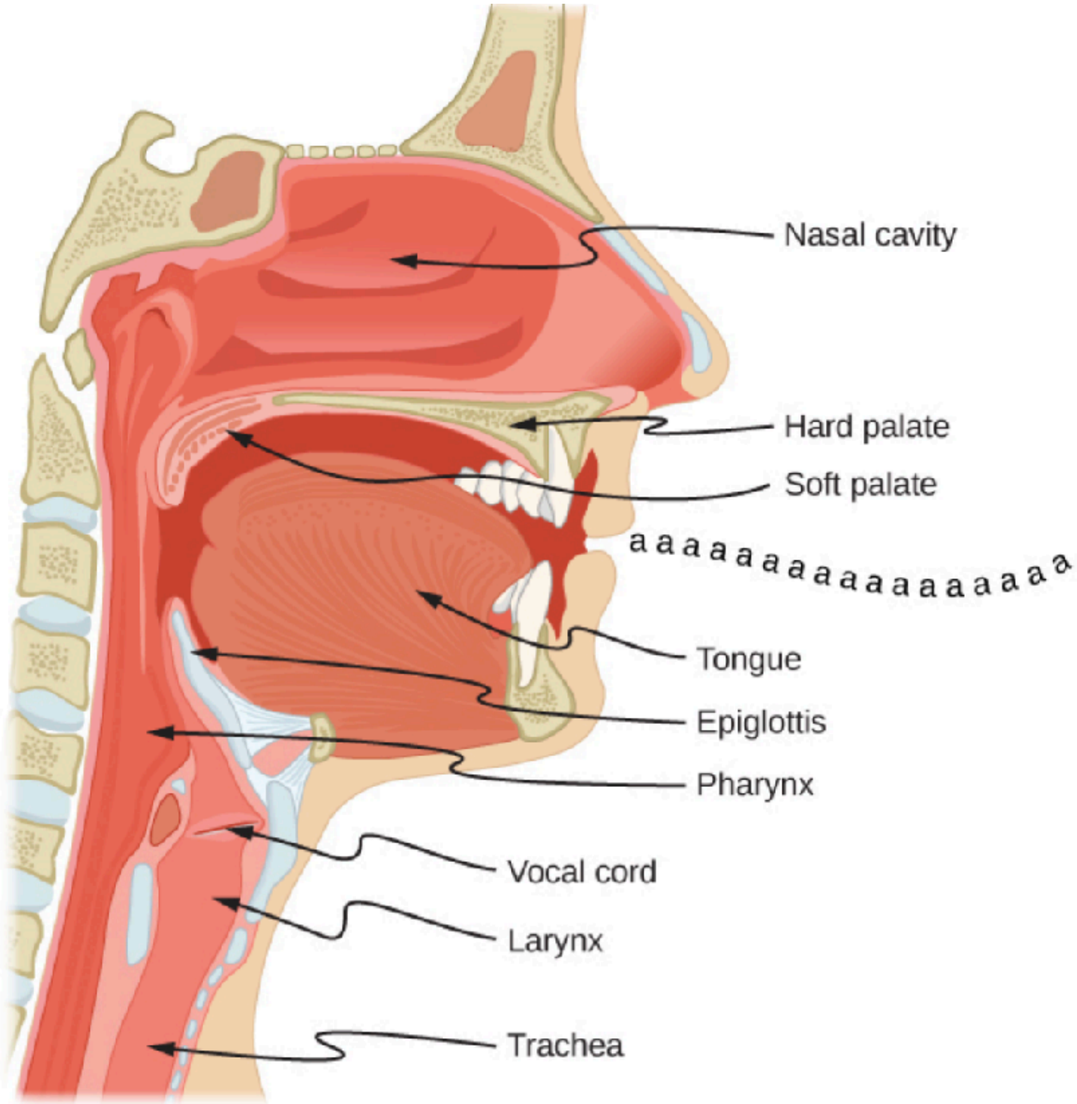


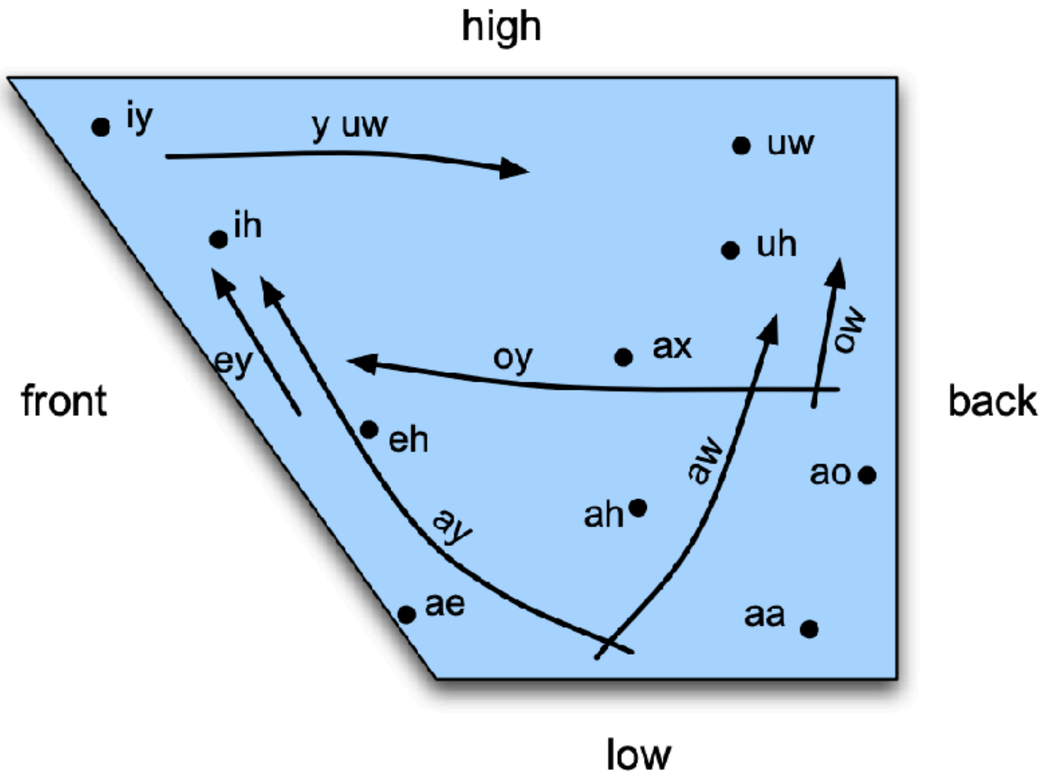


Lecture 6 Speech perception



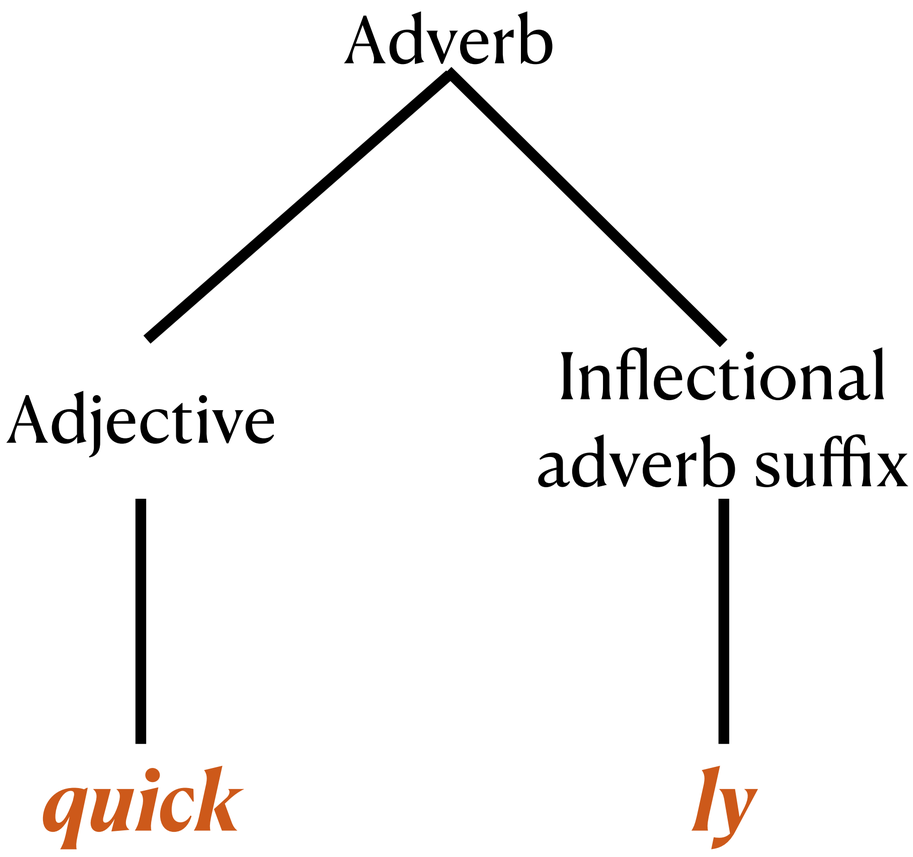
Lecture 7: Human sounds and their organization



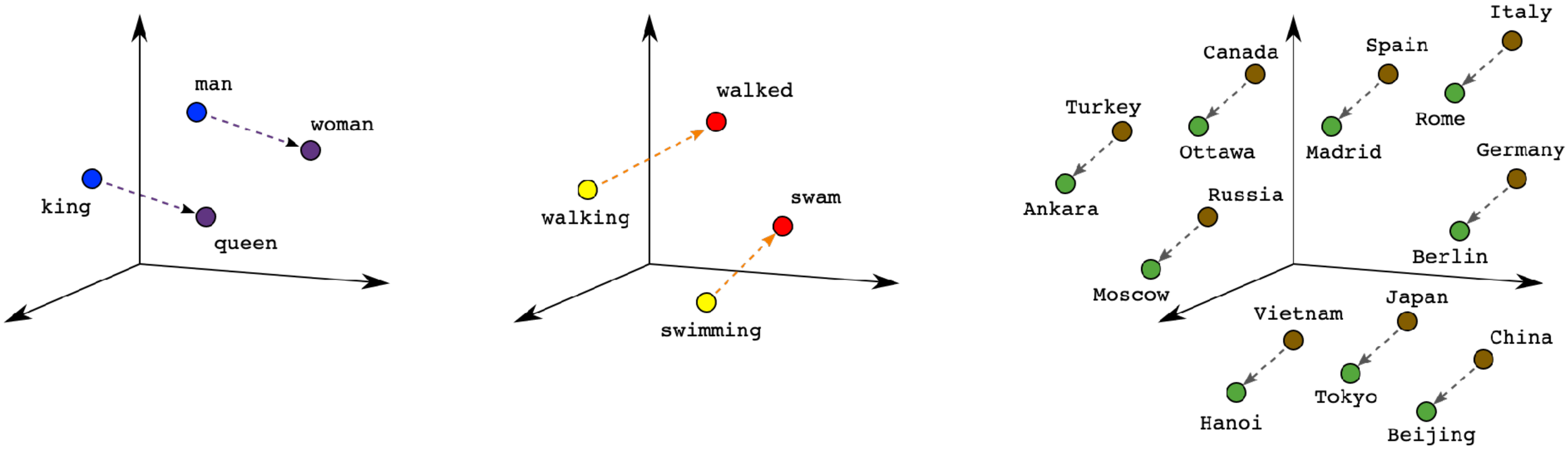


Lecture 8: Text processing and regular expressions

Lecture 9: Words, parts of speech and morphology



Lecture 10: Embedding



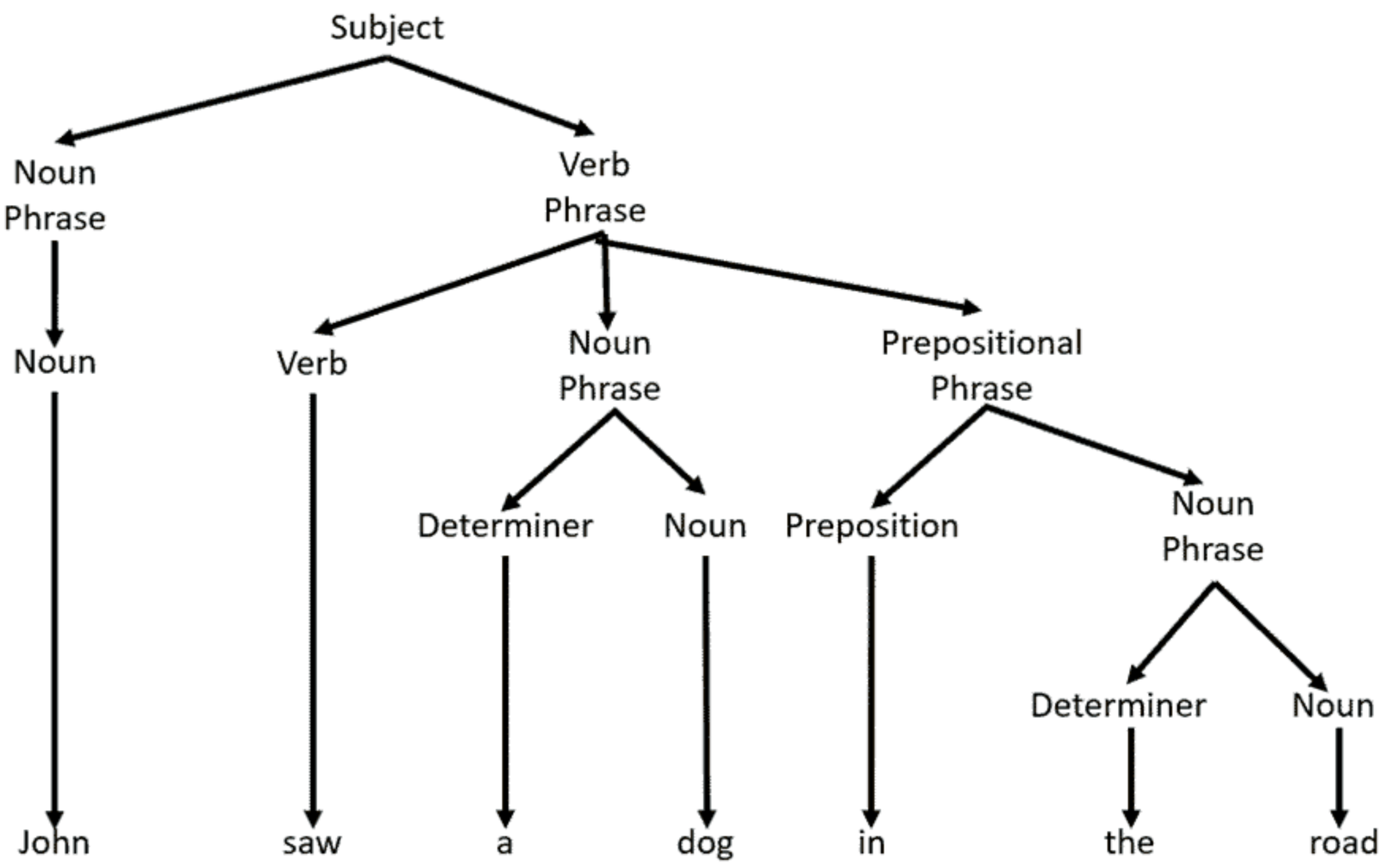
Male-Female

Verb Tense

Country-Capital

https://developers.google.com/machine-learning/crash-course/embeddings/translating-to-a-lower-dimensional-space

Lecture 11: Syntax - Structure of sentences



Lecture 12: Language model

S = I am tested positive

Previous words (Context)

Word being predicted

One slide overview

Applications

- Named entity recognition
- Speech recognition
- Machine translation
- Sentiment analysis
- Text-to-speech synthesis
- Question answering
- Text summarization
- Voice conversion
- Chatbot

Fundamentals

- Basics of speech processing
- Basics of language
- Language models
- Phonetics
- Prosody and timbre
- Morphology
- Parts of speech
- Semantics and embedding
- Text normalization
- Syntax and parsing

Lecture 13: Named entity recognition

In fact, the Chinese NORP market has the three CARDINAL most influential names of the retail and tech space – Alibaba GPE , Baidu ORG , and Tencent PERSON (collectively touted as BAT ORG), and is betting big in the global AI GPE in retail industry space . The three CARDINAL giants which are claimed to have a cut-throat competition with the U.S. GPE (in terms of resources and capital) are positioning themselves to become the 'future AI PERSON platforms'. The trio is also expanding in other Asian NORP countries and investing heavily in the U.S. GPE based AI GPE startups to leverage the power of AI GPE . Backed by such powerful initiatives and presence of these conglomerates, the market in APAC AI is forecast to be the fastest-growing one CARDINAL , with an anticipated CAGR PERSON of 45% PERCENT over 2018 - 2024 DATE .

To further elaborate on the geographical trends, North America LOC has procured more than 50% PERCENT of the global share in 2017 DATE and has been leading the regional landscape of AI GPE in the retail market. The U.S. GPE has a significant credit in the regional trends with over 65% PERCENT of investments (including M&As, private equity, and venture capital) in artificial intelligence technology. Additionally, the region is a huge hub for startups in tandem with the presence of tech titans, such as Google ORG , IBM ORG , and Microsoft ORG .

https://umagunturi789.medium.com/everything-you-need-to-know-about-named-entity-recognition-2a136f38c08f

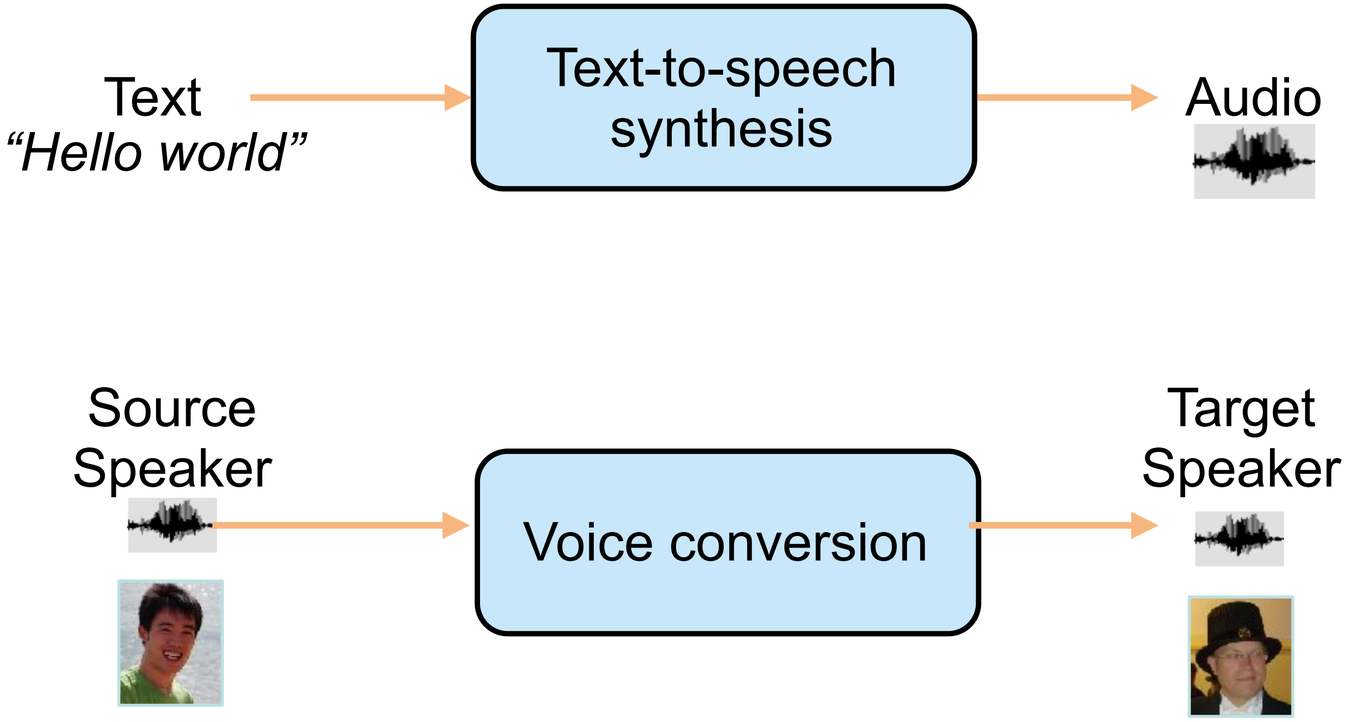
Lecture 14: SLP Application - Sentiment analysis

https://www.expressanalytics.com/blog/social-media-sentiment-analysis/

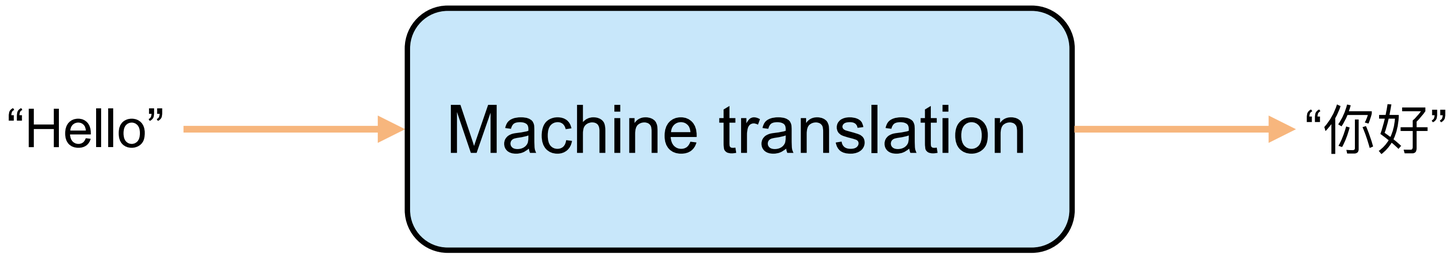
Lecture 15: Speech recognition

https://developer.nvidia.com/blog/solving-automatic-speech-recognition-deployment-challenges/

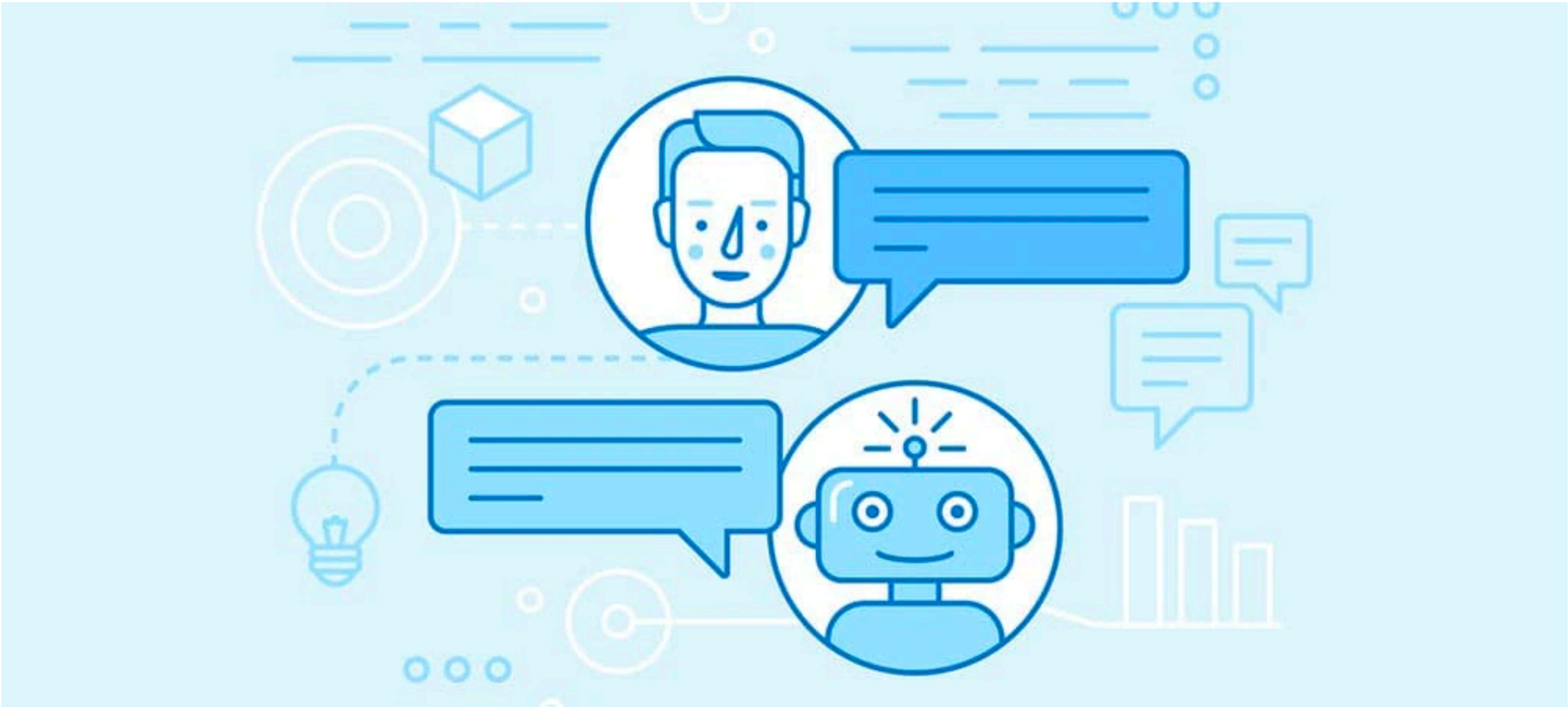
Lecture 16-17: Speech synthesis and voice conversion



Lecture 18: Machine translation



Lecture 19 - 20: Chatbot



https://www.hp.com/us-en/shop/tech-takes/what-is-a-chatbot

Guest lectures

- There is a two-day workshop on spoken language generative AI on Apr 20 and Apr 21.
  - Students from CSC3160/AIR6063 are free
  - Opportunities to meet high-profile researchers and industry experts

One slide overview

Applications

- Named entity recognition
- Speech recognition
- Machine translation
- Sentiment analysis
- Text-to-speech synthesis
- Question answering
- Text summarization
- Voice conversion
- Chatbot

Fundamentals

- Basics of speech processing
- Basics of language
- Language models
- Phonetics
- Prosody and timbre
- Morphology
- Parts of speech
- Semantics and embedding
- Text normalization
- Syntax and parsing

Thanks!

Zhizheng Wu
Associate professor
https://drwuz.com/

Course website: https://drwuz.com/CSC3160/