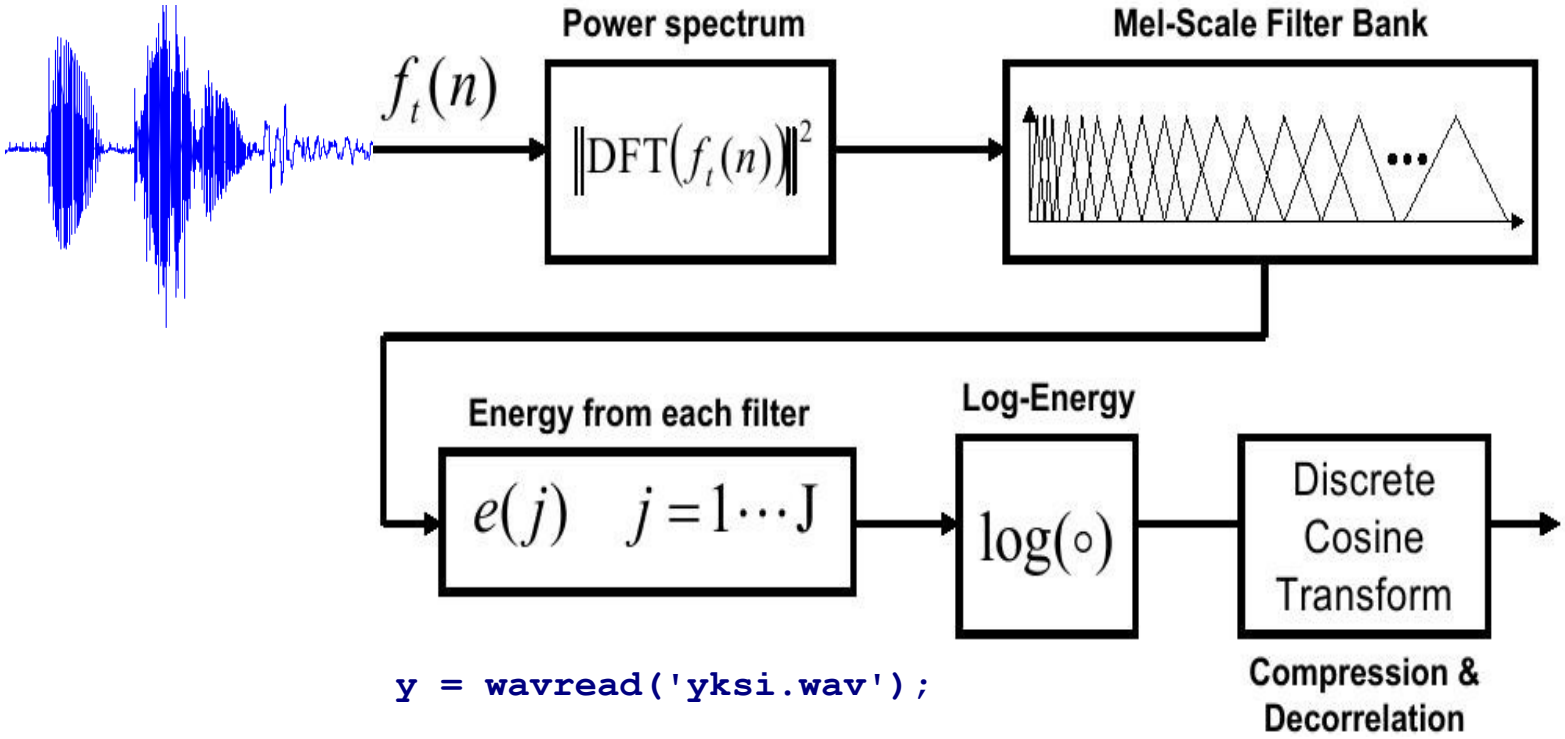
Timeline in the course

	Meetings Wednesdays	Thursdays or Fridays	Home exercises	Project work status
Week1	Speech features	Classification	Feature classifier	Literature study
	entry test			Meet tutors Wed
Week2	Phoneme modeling	Recognition	Word recognizer	Work plan
				Meet tutors Wed
Week3	Lexicon and language	Language model	Text predictor	Analysis
				Meet tutors Wed
Week4	Continuous speech	LVCSR	Speech recognizer	Experimentation
	advanced search			Meet tutors Wed
Week5	End-to-end ASR	End-to-end	End-to-end recognizer	Preparing reports
				Meet tutors Wed
Week6	Projects1	Projects2		Presentations
Week7	Projects3	Projects4		Report submissio
		Conclusion		

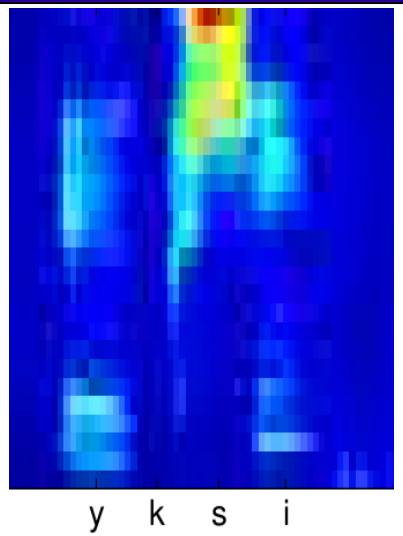
Learning goals for this week

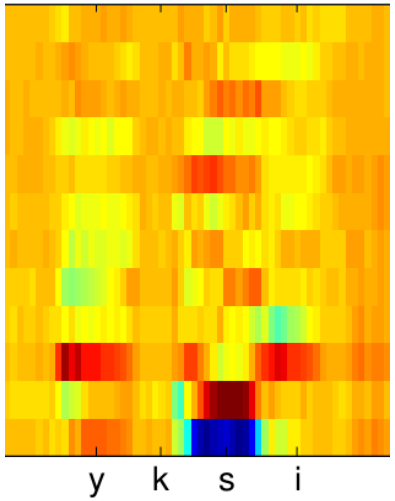
1. **Preprocessing, features, GMM**
   - remind of last week
2. Phonemes
   - know different units of speech
3. HMM
   - learn to build a temporal model of speech units
4. Home exercise 2: Build a GMM-HMM system to recognize spoken words

Review: computation of MFCC

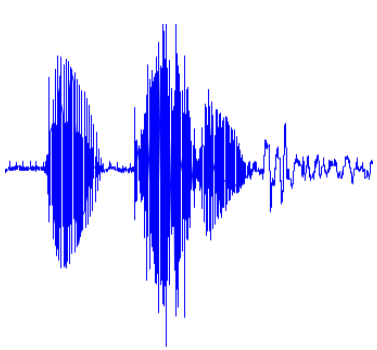


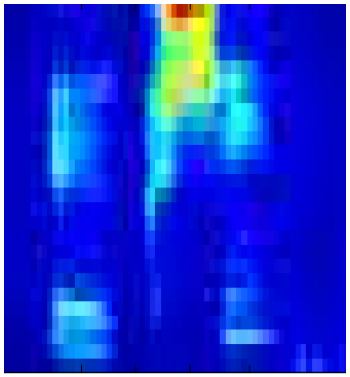

```
y = wavread('yksi.wav');
s = spectrogram(y,hamming(400),240);
spec = sqrt(abs(s));
mfcc = D*log(1+M*spec);
```



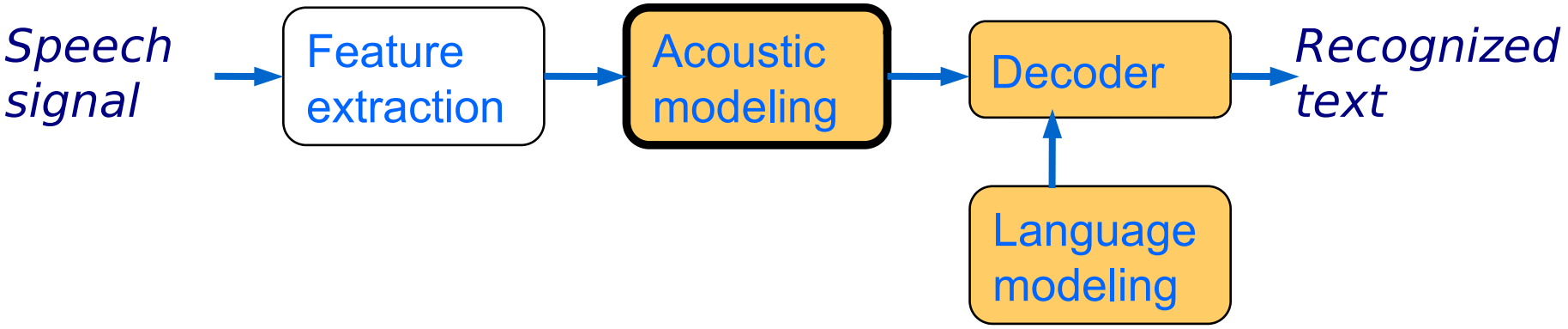


Review: speech recognition -from beginning to end





Today:



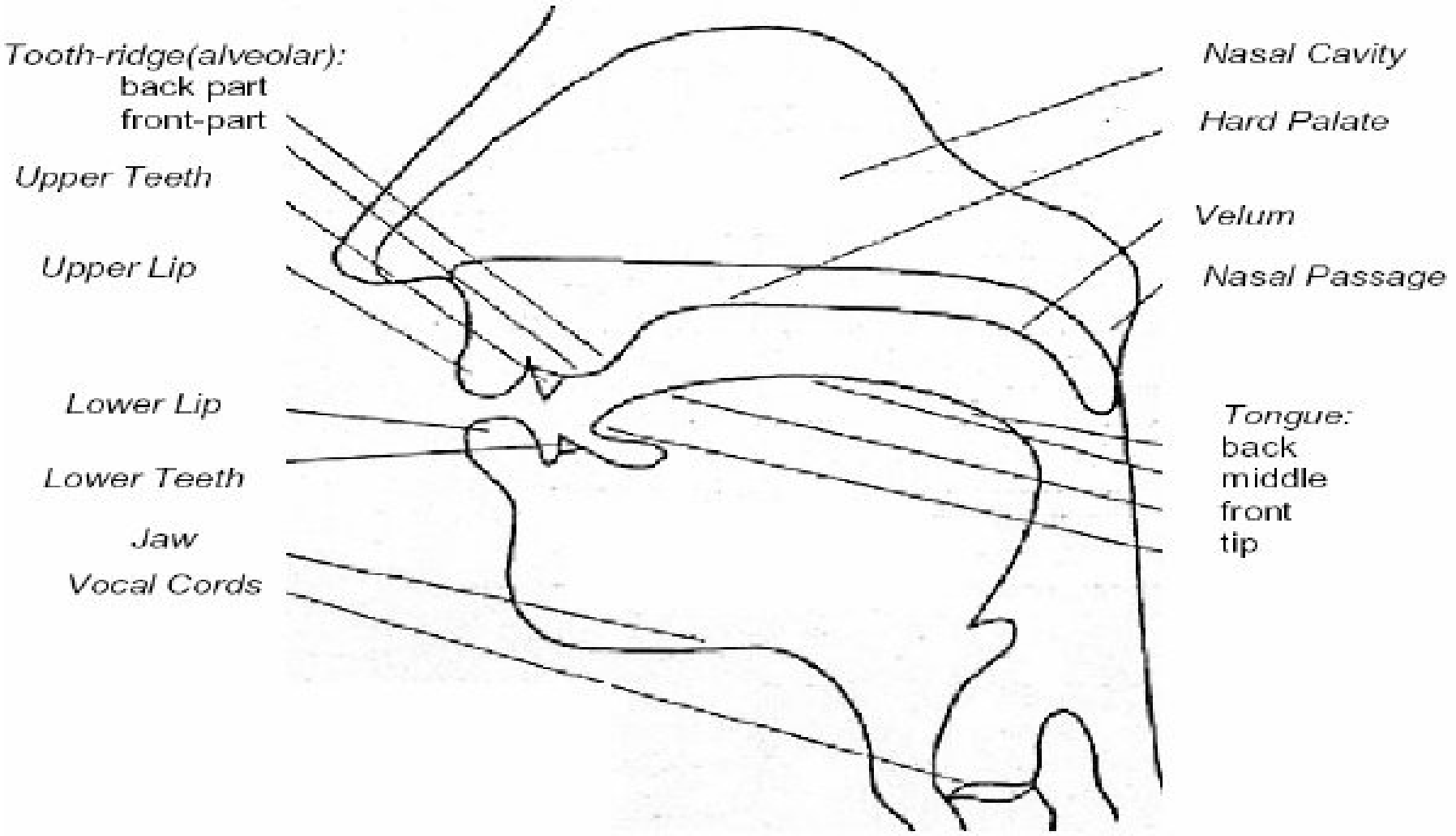
Content today

1. Preprocessing, features, GMM
2. **Phonemes**
3. HMM
4. Home exercise 2: Build a GMM-HMM system to recognize spoken words

Description of speech sounds

- **Speech can be written down** using abstract units called phonemes
- **Phonemes** describe the sounds **by the way they are produced by human**
- Main classes:
  - vowels: air flow is not obstructed
  - consonants: air flow is partially or totally obstructed
- There are different writing systems, e.g. IPA (International Phonetic Alphabet)
- The phoneme sets differ depending on language

Production of speech sounds



Picture from Huang's text book (2001)

IPA symbols for US English

PHONEME	EXAMPLE	PHONEME	EXAMPLE	PHONEME	EXAMPLE
/iʸ/	beat	/s/	see	/w/	wet
/ɪ/	bit	/š/	she	/r/	red
/eʸ/	bait	/f/	fee	/l/	let
/ɛ/	bet	/θ/	thief	/y/	yet
/æ/	bat	/z/	z	/m/	meet
/ɑ/	Bob	/ž/	Gigi	/n/	neat
/ɔ/	bought	/v/	v	/ŋ/	sing
/ʌ/	but	/ð/	thee	/č/	church
/oʷ/	boat	/p/	pea	/j/	judge
/ʊ/	book	/t/	tea	/h/	heat
/uʷ/	boot	/k/	key		
/ɝ/	Burt	/b/	bee		
/ɑʸ/	bite	/d/	Dee		
/ɔʸ/	Boyd	/g/	geese		
/ɑʷ/	bout				
/ə/	about				

Table from MIT Course Notes: 6.345 Automatic Speech Recognition, Spring 2003

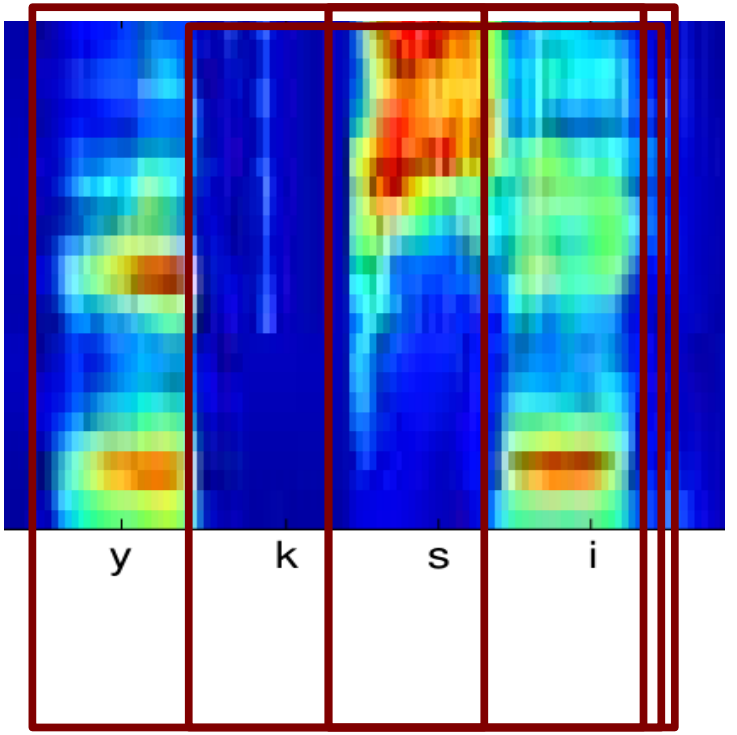
CMU Sphinx ASR system symbols

Phone	Example	Phone	Example	Phone	Example
AA	odd	EY	ate	P	pee
AE	at	F	fee	PD	lip
AH	hut	G	green	R	read
AO	ought	GD	bag	S	sea
AW	cow	HH	he	SH	she
AX	abide	IH	it	T	tea
AXR	user	IX	acid	TD	lit
AY	hide	IY	eat	TH	theta
B	be	JH	gee	TS	bits
BD	Dub	K	key	UH	hood
CH	cheese	KD	lick	UW	two
D	dee	L	lee	V	vee
DD	dud	M	me	W	we
DH	thee	N	note	Y	yield
DX	matter	NG	ping	Z	zee
EH	ed	OW	oat	ZH	seizure
ER	hurt	OY	toy	SIL	(silence)

Acoustic model of speech

- Discussion: What speech units would suit for ASR?
- (how long, how many, language-dependence)
- (is the linguistic phoneme definition optimal?)

Why these discussions?
Learning happens, when:
+ brains are active and alert
+ new knowledge contradicts your old believes



In ASR: Context-dependent phonemes

- **Context independent model,** Monophone /X/
  - Example: three => th + r + iy
  - does a phoneme sound the same in all contexts ?
- **Context dependent model,** Triphone /Left-X+Right/
  - Example: three => sil-**th**+r + th-**r**+iy + r-**iy**+sil
  - 25 phonemes => 25*25*25 = 15 625 triphones
  - do all the contexts exist ?
  - do all the contexts sound different ?
  - can we share parts of the model between some contexts, e.g. beginning, center, middle part?

Content today

1. Preprocessing and features, GMM
2. Phonemes
3. **Hidden Markov Model**
4. Home exercise 2: Build a GMM-HMM system to recognize spoken words

Test what you remember from week 1

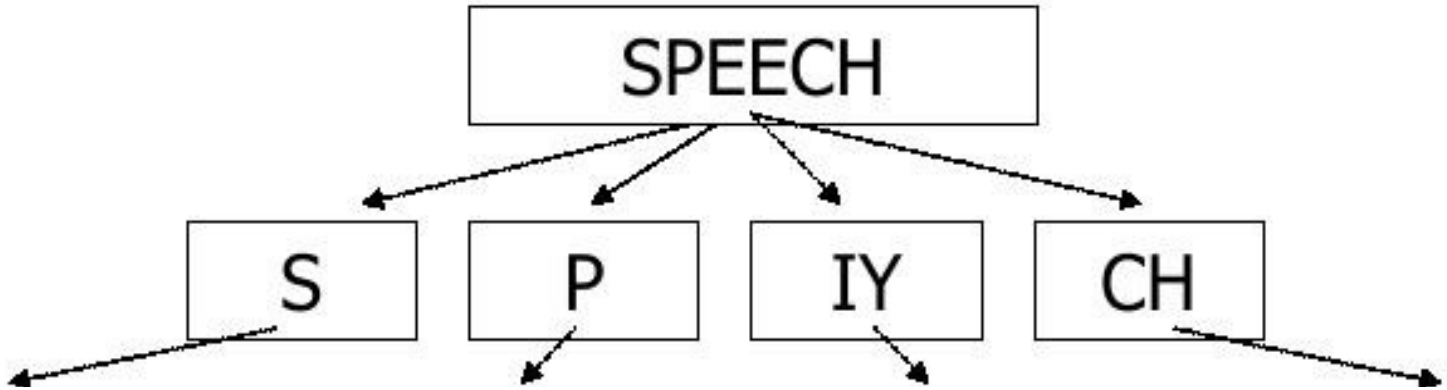
Individual test for everyone, now:

1. Go to ***https://kahoot.it*** with your phone/laptop
2. Type in the ID number you see on the screen (also in chat)
3. Give your **REAL surname**
4. Answer the questions by selecting **only one** of the options
   - There may be several right (or wrong) answers, but just pick one
   - About 1 min time per question
5. 1 activity points for everyone + 0.2 per correct answer in time
   - Kahoot time/score is just for fun, only the answers matter

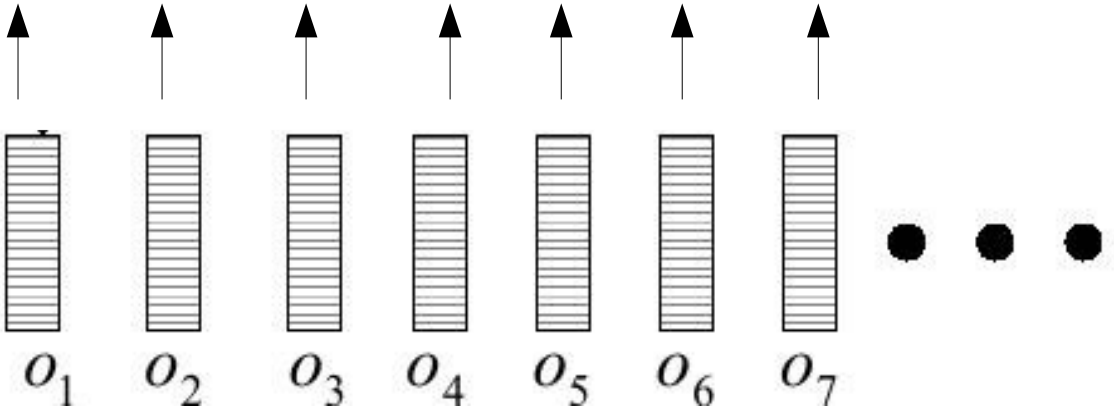
Results of GMM classification?

- This is not yet speech recognition, not even phoneme recognition!
- How to utilize this in phoneme recognition?

How to model a sequence of frames or phonemes?



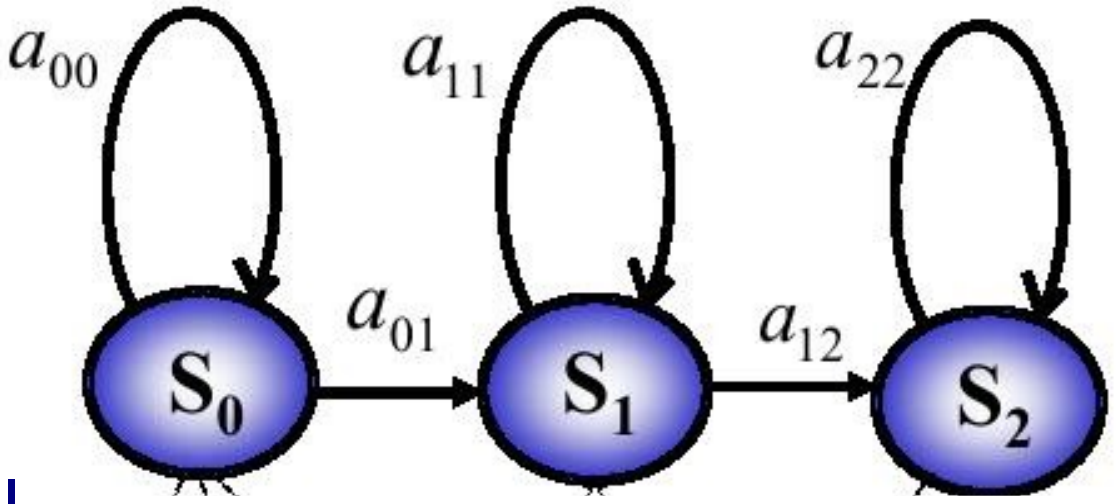
S S S H P P T E IY IY IY E S H CH



Picture by B.Pellom

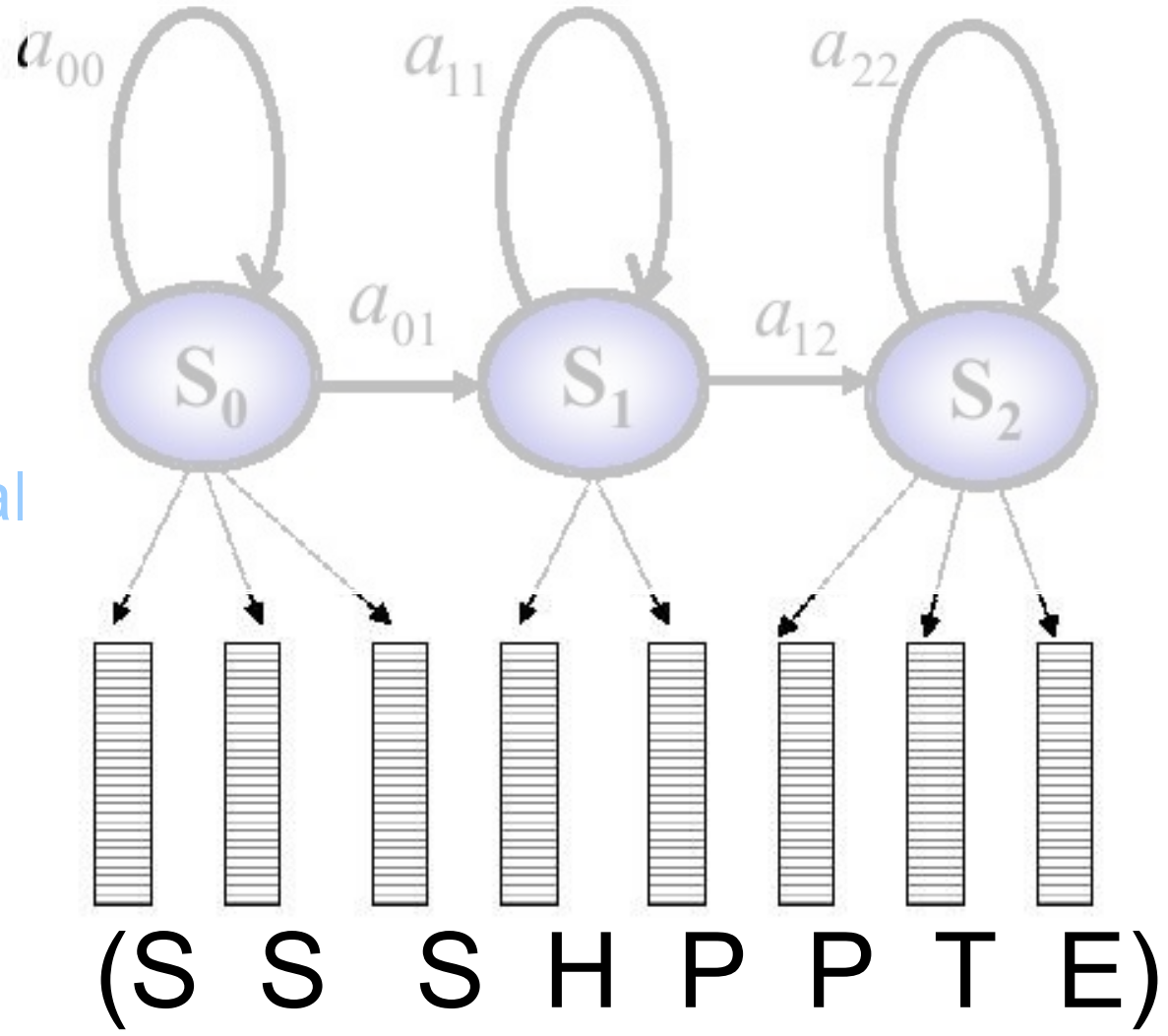
Hidden Markov model

1. HMM is a system that has a set of operational states
2. From state *i* it moves to state *j* by probability *a(ij)*
3. Each state emits a characteristic sound signal
4. Signals are measured by feature vectors
5. The system's internal **state is hidden**, only the **feature vectors** are **measured**



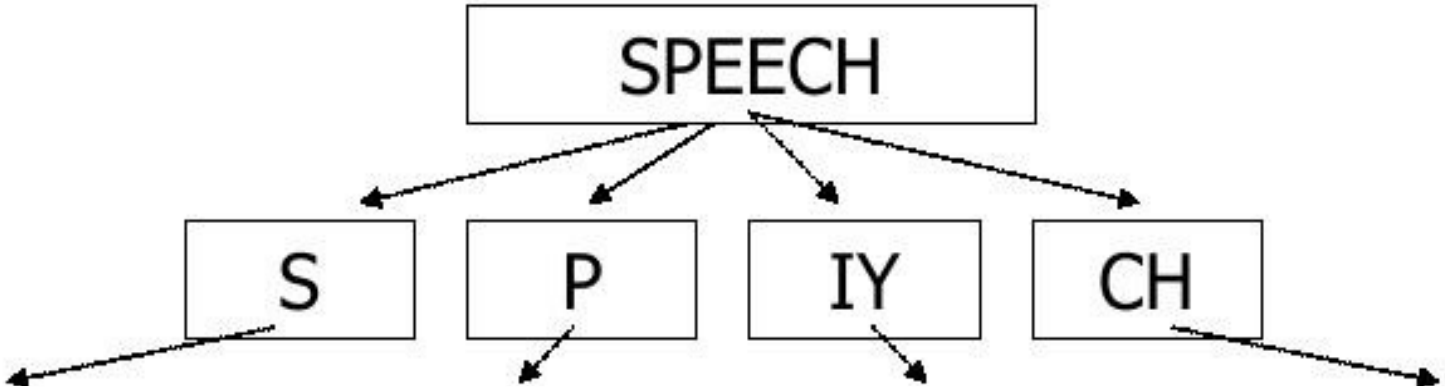
Hidden Markov model

1. HMM is a system that has a set of operational states
2. From state *i* it moves to state *j* by probability *a(ij)*
3. Each state emits a characteristic sound signal
4. Signals are measured by feature vectors
5. The system's internal **state is hidden**, only the **feature vectors** are **measured**

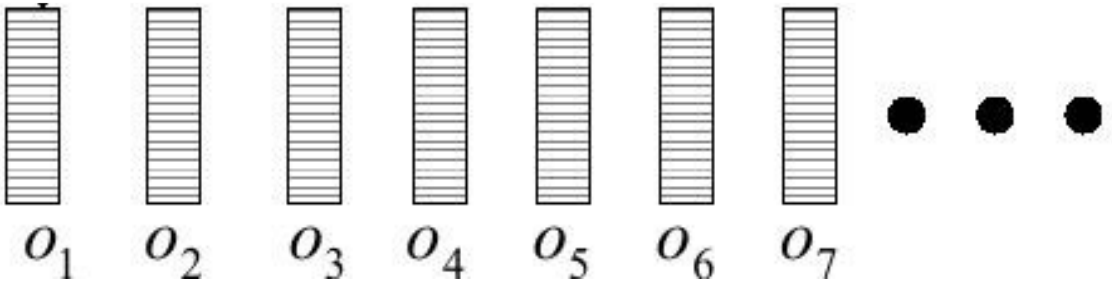


Picture by B.Pellom

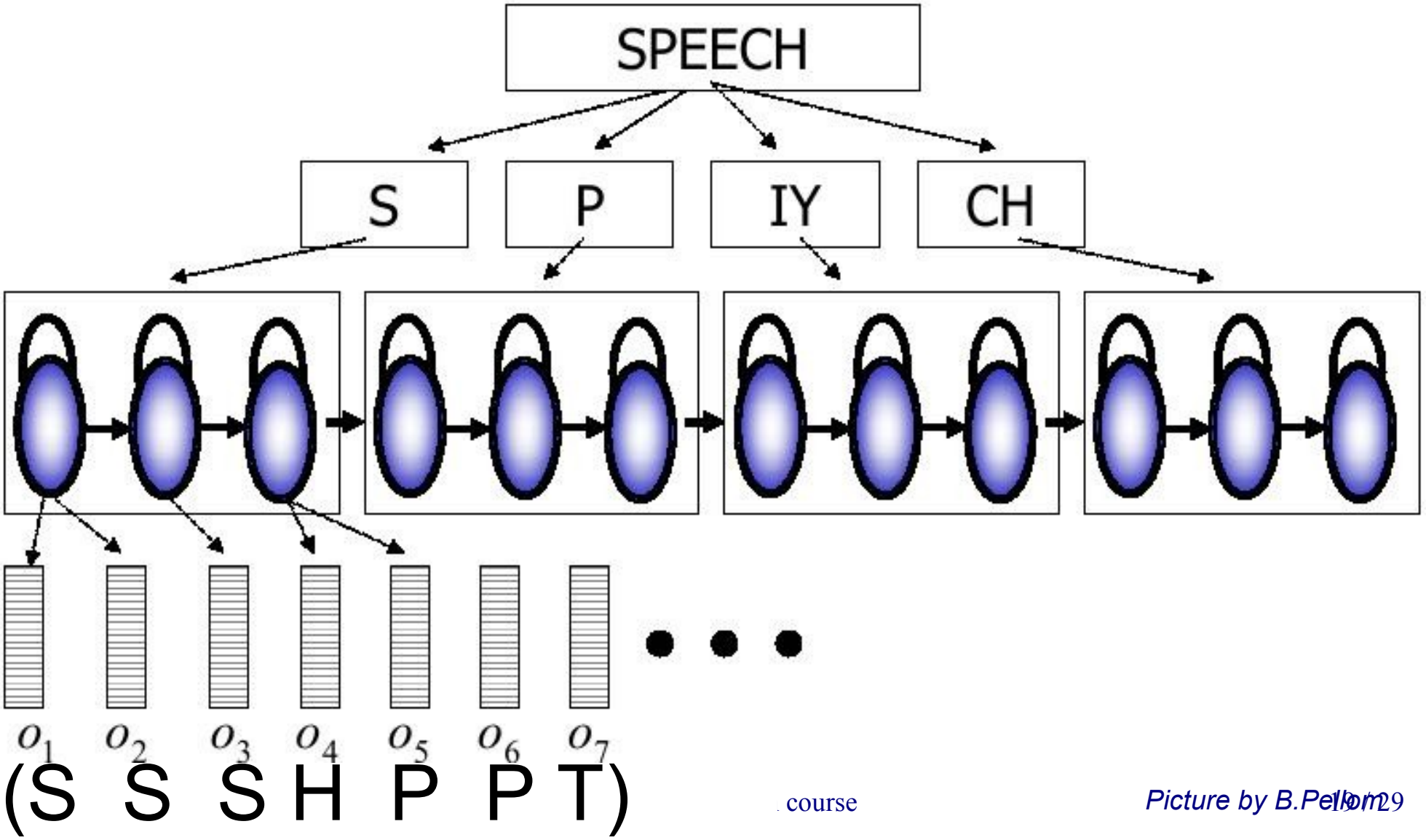
How to model a sequence of frames or phonemes?



?

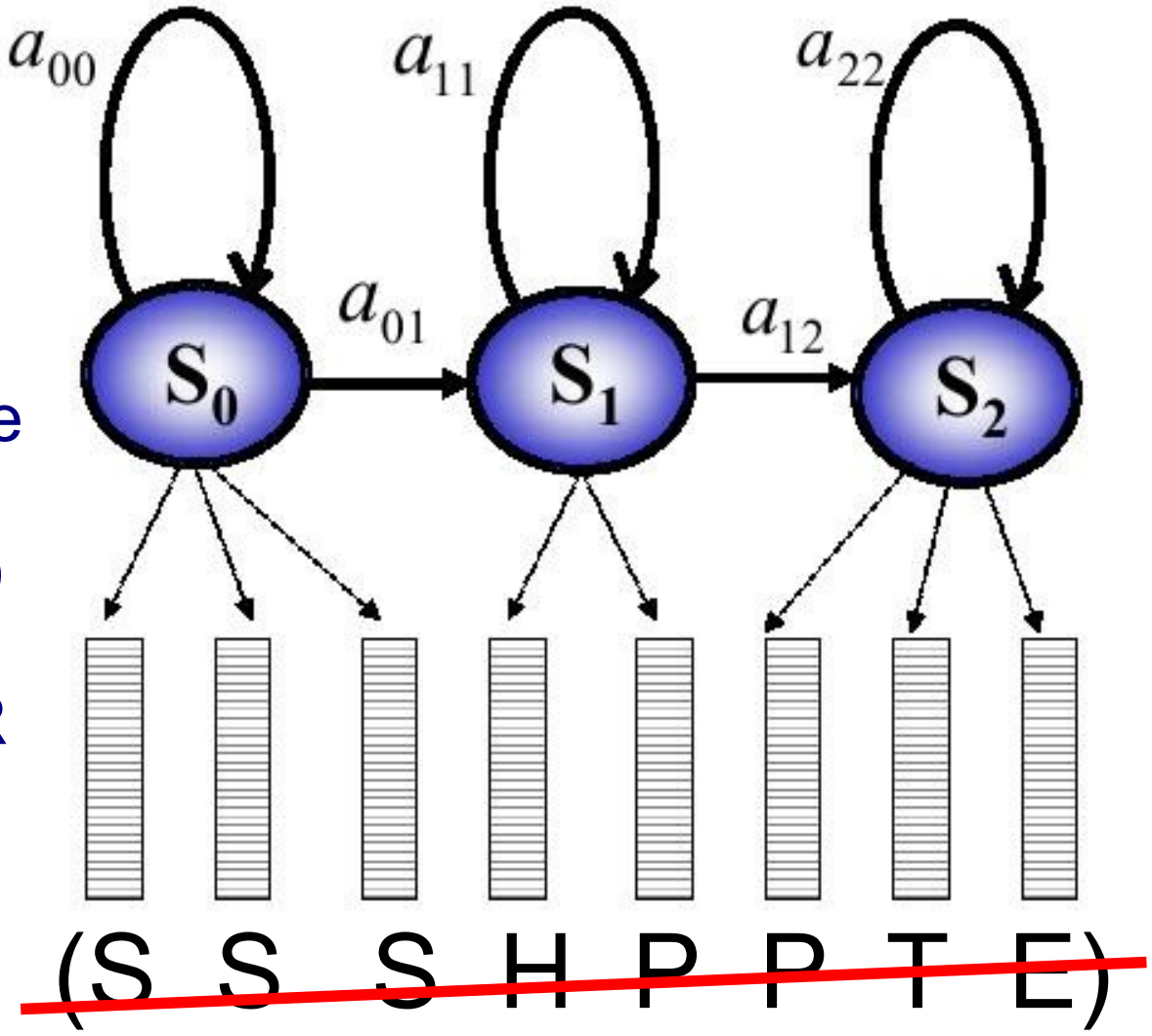


HMM as a phoneme model



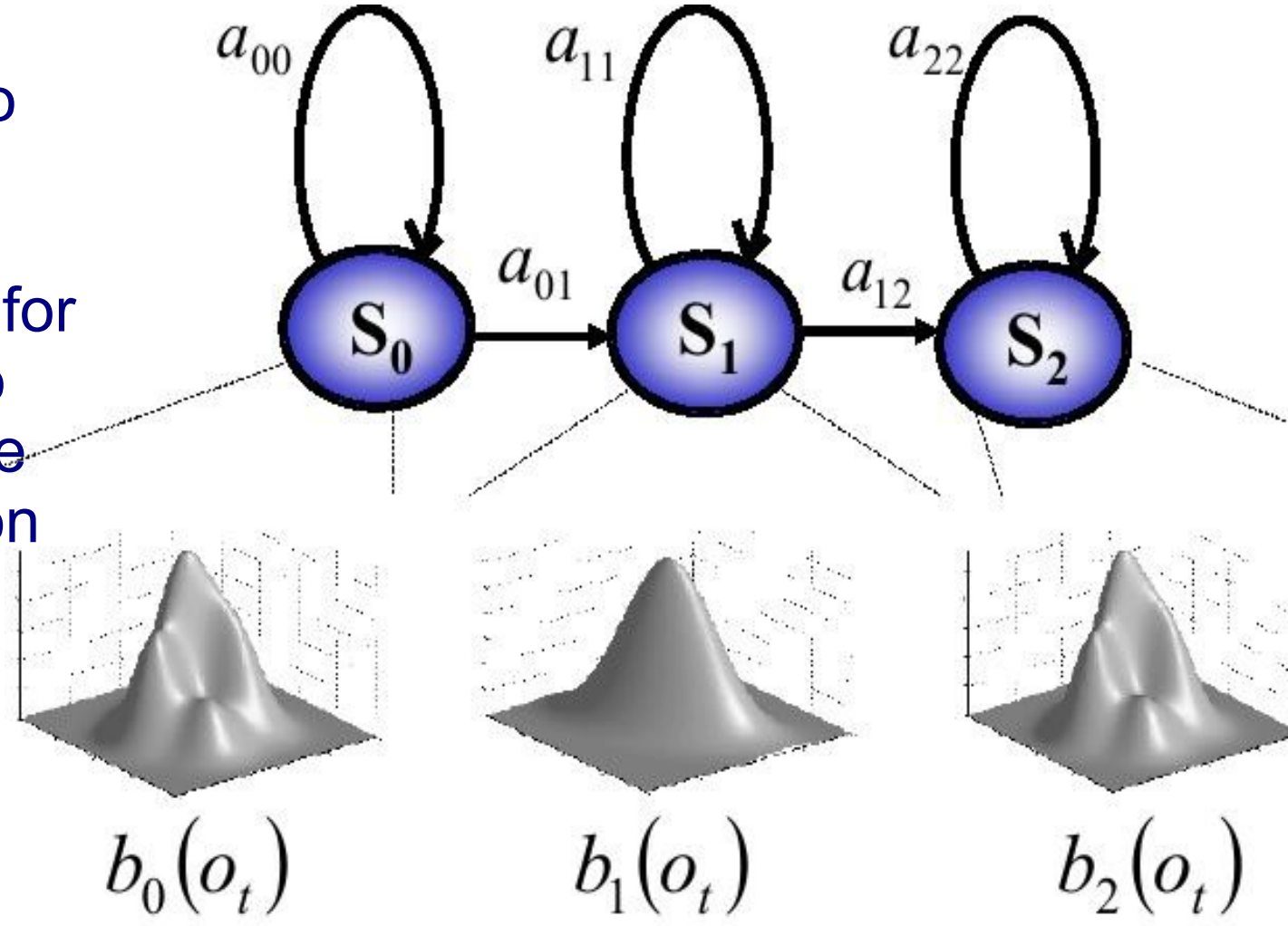
HMM as a phoneme model

- After **segmenting** each word sample into sounds, we find the set of feature vectors that represent a certain state
- These feature vectors are used to model the outputs in the state (by GMM e.g.)
- After modeling the states the HMM is ready for ASR



GMM-HMM system

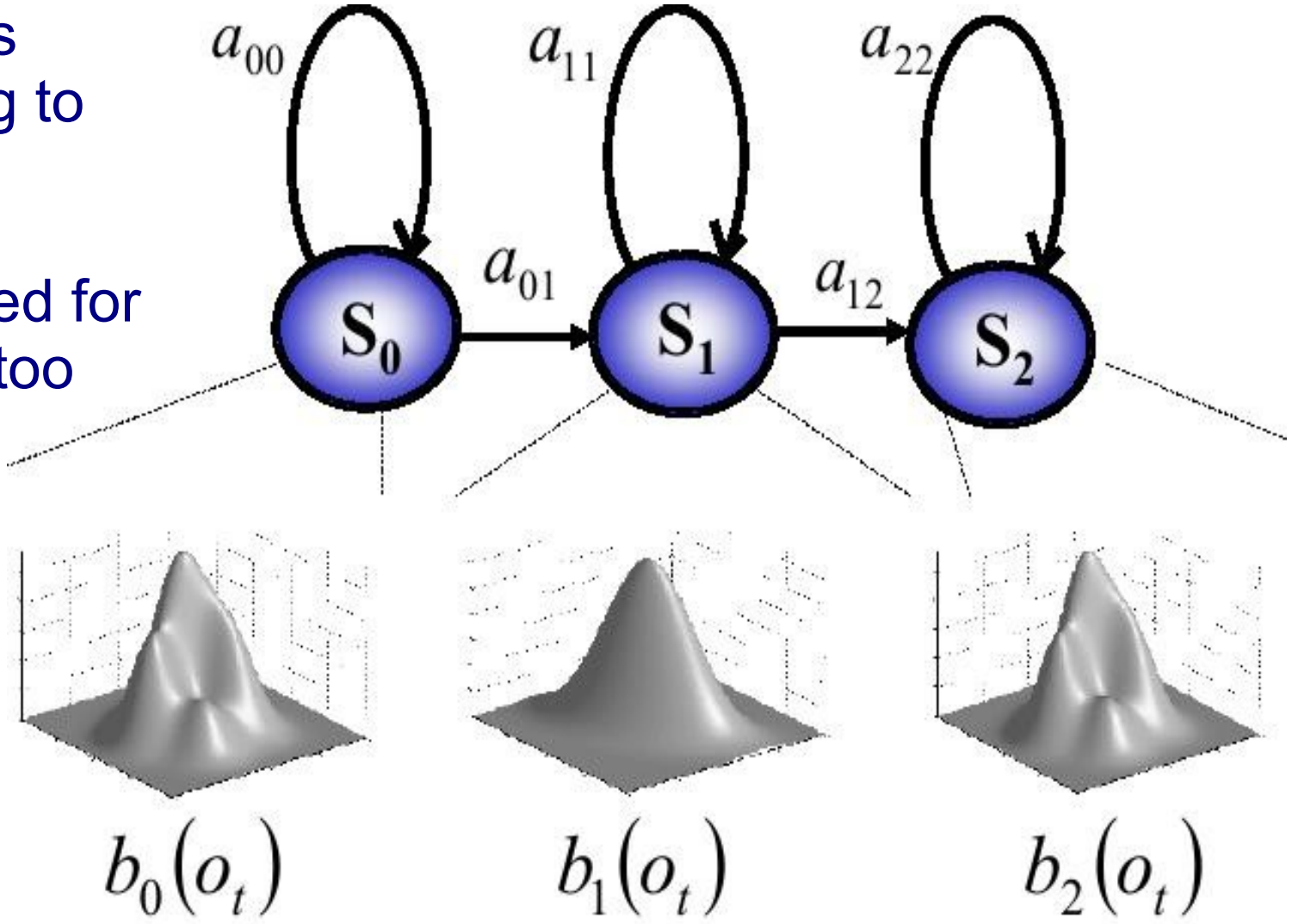
- Each state emits sounds according to its GMM model
- This generative model can be used for **text-to-speech**, too
- The higher *a(ii)*, the longer is the duration



GMM-HMM system

- Each state emits sounds according to its GMM model
- This generative model can be used for **text-to-speech**, too

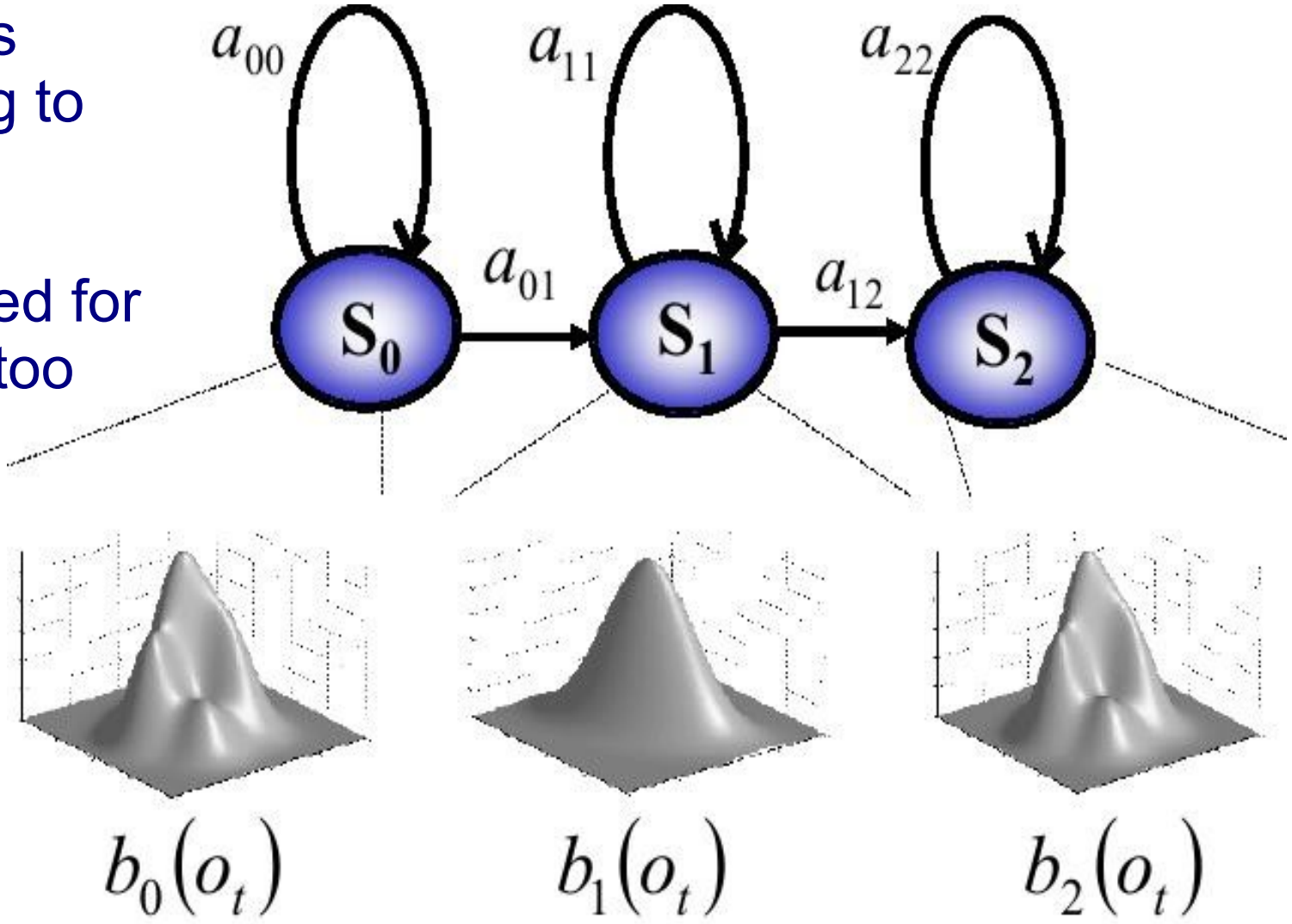
- Sample 1
- Sample 2
- Sample 3



GMM-HMM system

- Each state emits sounds according to its GMM model
- This generative model can be used for **text-to-speech**, too

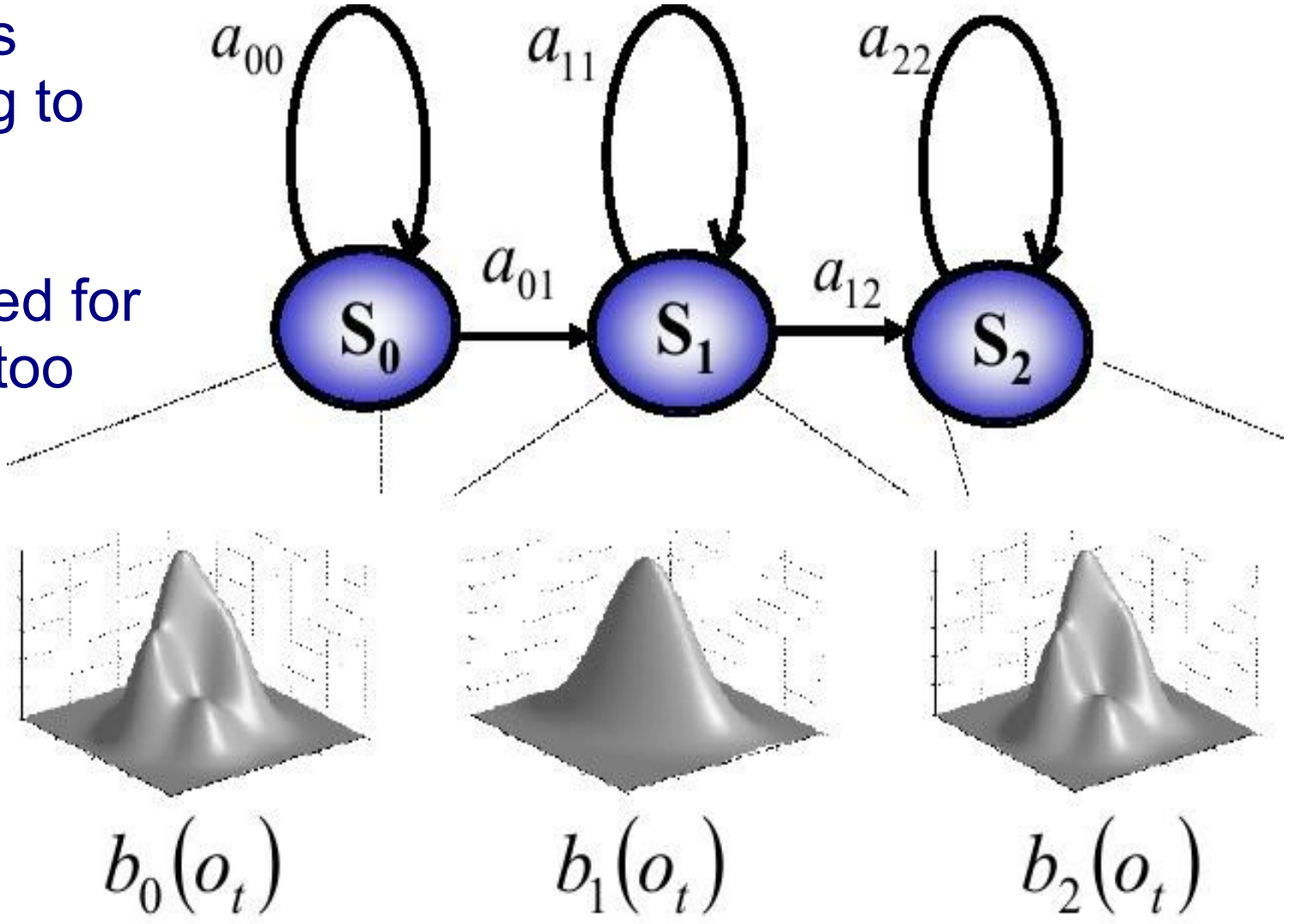
- Sample 1
- Sample 2
- Sample 3



GMM-HMM system

- Each state emits sounds according to its GMM model
- This generative model can be used for **text-to-speech**, too

- Sample 1
- Sample 2
- Sample 3



Basic operations with HMMs

1. **Scoring**: - How to compute the probability of the observation sequence for a model?

2. **Decoding**: - How to compute the best state sequence for the observations?

3. **Training**: - How to set the model parameters to maximize the probability of the training samples?

Article: Rabiner (1989), *Tutorial on hidden Markov models and selected applications*

Feedback

Now: Go to **MyCourses > Lectures > Lecture 2 feedback** and fill in a feedback form to get an activity point.

Some of the feedback from the previous week:

+ interactive, which keeps me awake

+ many ways to get lecture points, could add even more

+ discussions in groups

- add references for self-study

- more focus on intuition behind formulas

- it was difficult to ask questions from physical to zoom

- possibility to attend remotely or have lectures recorded

Thanks for all the valuable feedback!

Summary of today

- Phonemes
- GMM and HMM
- **Next meeting:** Thu 10.15 – 12 or Fri 14.15 – 16: Speech recognition by HTK toolkit
  - check *http://htk.eng.cam.ac.uk/docs/docs.shtml*
  - This exercises is useful for most project works!
- **Next week:** Language models and lexicon

Project work receipt

1. Form a group (3 persons)
2. Get a topic
3. Get reading material from M*ycourses* or your group tutor
4. 1st meeting: Specify the topic, start literature study
5. **2nd meeting: Write a work plan**
6. 3rd - 5th meetings: Perform analysis, experiments, and write a report
7. Book your presentation time for weeks 6 - 7
8. Prepare and keep your 20 min presentation
9. Return the report

This week

Check **MyCourses > Projects** to see your group, topic and tutor

Final project report

1. Abstract: (your working plan)
2. Introduction: (your literature review)
   - Remember to cite every article you read
3. Experiments: Describe what you did
4. Results: Describe the results you got
5. Conclusion: Your conclusion of the work
6. References: (list of articles that you read)