

---

# Speech Signal Processing

## Exercise 7

### Decoding

1. The output of a statistical speech recognition system is often given in the form of a lattice. Use Viterbi-Decoding to compute the most probable path through the given lattice (see attachment).

Start node	End mode	Sum of acoustic model, LM and the accumulated predecessors		accumulated probability
0	1	-1432,27-0	=	-1432,27
0	2	-1500,93-0	=	-1500,93
0	3	-3759,32-0	=	-3759,32
0	4	-3829,60-0	=	-3829,60
1	5	(-2434,05-87,29)-1432,27	=	-3953,61
2	5	(-2431,55-87,29)-1500,93	=	

2. Phonology

- (a) What is a phoneme?
- (b) What is the difference between monophones and triphones?
- (c) In theory, how many triphones do exist? Why is the number of triphones used in practice much lower?

3. Review

- Sketch the single components of a statistical speech recognizer. Describe and explain the components.
- Write down Bayes' formula. Name the different probabilities.
- What is the cepstrum?
- What is a bigram and where is it used?
- What is the Viterbi-algorithm? Where do we need it within the speech recognition field?
- What is a Hidden Markov Model? Why do we refer to it as a production model?
- What is a phoneme lexicon?
- For which purpose do we use the Linear Discriminant Analysis?
- Describe the Baum-Welch algorithm.
- Explain *tying*.
- What is the vocal tract tube model?

- 
- Which stochastic distribution is typically used when phonemes are modelled?
  - What is a *feature vector*? Explain the difference between static and dynamical feature vectors.
  - Explain *dynamic time warping*.

---

# Attachment

## Lattice mit N=24 L=39

Node definitions:

I=0	t=0.00	I=6	t=0.71	I=12	t=0.72	I=18	t=0.81
I=1	t=0.25	I=7	t=0.72	I=13	t=0.73	I=19	t=0.81
I=2	t=0.26	I=8	t=0.72	I=14	t=0.78	I=20	t=1.33
I=3	t=0.61	I=9	t=0.72	I=15	t=0.78	I=21	t=2.09
I=4	t=0.62	I=10	t=0.72	I=16	t=0.80	I=22	t=2.09
I=5	t=0.62	I=11	t=0.72	I=17	t=0.80	I=23	t=2.85

Link definitions:

J=0	S=0	E=1	W=!ENTER	v=0	a=-1432.27	l=0.00
J=1	S=0	E=2	W=!ENTER	v=0	a=-1500.93	l=0.00
J=2	S=0	E=3	W=!ENTER	v=0	a=-3759.32	l=0.00
J=3	S=0	E=4	W=!ENTER	v=0	a=-3829.60	l=0.00
J=4	S=1	E=5	W=TO	v=3	a=-2434.05	l=-87.29
J=5	S=2	E=5	W=TO	v=1	a=-2431.55	l=-87.29
J=6	S=4	E=6	W=AND	v=3	a=-798.30	l=-69.71
J=7	S=4	E=7	W=IT	v=0	a=-791.79	l=-62.05
J=8	S=4	E=8	W=AND	v=2	a=-836.88	l=-69.71
J=9	S=3	E=9	W=BUT	v=0	a=-965.47	l=-51.14
J=10	S=4	E=10	W=A.	v=0	a=-783.36	l=-105.95
J=11	S=4	E=11	W=IN	v=0	a=-835.98	l=-49.01
J=12	S=4	E=12	W=A	v=0	a=-783.36	l=-59.66
J=13	S=4	E=13	W=AT	v=0	a=-923.59	l=-77.95
J=14	S=4	E=14	W=THE	v=0	a=-1326.40	l=-27.96
J=15	S=4	E=15	W=E.	v=0	a=-1321.67	l=-121.96
J=16	S=4	E=16	W=A	v=2	a=-1451.38	l=-59.66
J=17	S=4	E=17	W=THE	v=2	a=-1490.78	l=-27.96
J=18	S=4	E=18	W=IT	v=0	a=-1450.07	l=-62.05
J=19	S=5	E=18	W=IT	v=0	a=-1450.07	l=-110.42
J=20	S=6	E=18	W=IT	v=0	a=-775.76	l=-85.12
J=21	S=7	E=18	W=IT	v=0	a=-687.68	l=-125.32
J=22	S=8	E=18	W=IT	v=0	a=-687.68	l=-85.12
J=23	S=9	E=18	W=IT	v=0	a=-687.68	l=-50.28
J=24	S=10	E=18	W=IT	v=0	a=-689.67	l=-108.91
J=25	S=11	E=18	W=IT	v=0	a=-706.89	l=-113.78
J=26	S=12	E=18	W=IT	v=0	a=-689.67	l=-194.91
J=27	S=13	E=18	W=IT	v=0	a=-619.20	l=-100.24
J=28	S=4	E=19	W=IT	v=1	a=-1567.49	l=-62.05
J=29	S=14	E=20	W=DIDN'T	v=0	a=-4452.87	l=-195.48
J=30	S=15	E=20	W=DIDN'T	v=0	a=-4452.87	l=-118.62
J=31	S=16	E=20	W=DIDN'T	v=0	a=-4303.97	l=-189.88
J=32	S=17	E=20	W=DIDN'T	v=0	a=-4303.97	l=-195.48
J=33	S=18	E=20	W=DIDN'T	v=0	a=-4222.70	l=-78.74
J=34	S=19	E=20	W=DIDN'T	v=0	a=-4235.65	l=-78.74
J=35	S=20	E=21	W=ELABORATE	v=2	a=-5847.54	l=-62.72
J=36	S=20	E=22	W=ELABORATE	v=0	a=-5859.59	l=-62.72
J=37	S=21	E=23	W=!EXIT	v=0	a=-4651.00	l=-13.83
J=38	S=22	E=23	W=!EXIT	v=0	a=-4651.00	l=-13.83

---

## Phonetic notation in SAMPA

### Vocals

a	<u>Land</u>
6	<u>Mutter</u>
@	<u>bitte</u>
E	<u>Bett</u>
e	<u>Leben</u>
2	<u>schön</u>
9	<u>Götter</u>
o	<u>Hose</u>
O	<u>Post</u>
i	<u>Schiene</u>
I	<u>Kind</u>
y	<u>Lüge</u>
Y	<u>füllen</u>
u	<u>Mut</u>
U	<u>Kuss</u>

### Plosives

p	<u>Bap</u>
b	<u>Bube</u>
t	<u>Tag</u>
d	<u>Gnade</u>
k	<u>Kamm</u>
g	<u>gegen</u>

### Nasals

m	<u>Mama</u>
n	<u>Nenner</u>
N	<u>Gong</u>

### Fricatives

s	<u>Tasse</u>
z	<u>Hose</u>
S	<u>Lasche</u>
Z	<u>Garage</u>
tS	<u>Quatsch</u>
dZ	<u>Jeans</u>
pf	<u>Pfirsich</u>
f	<u>Fisch</u>
v	<u>Wagen</u>
C	<u>Licht</u>
x	<u>acht</u>
R	<u>Trier</u>
h	<u>Hase</u>

### Diphthongs

aI	<u>Rhein</u>
aU	<u>Krause</u>
OY	<u>teuer</u>

### more consonants

r	(vorne) rollendes „R“
R	(hinten) rollendes „R“
l	<u>Laut</u>
j	<u>Jacke</u>

### Prolongation

:	<u>Bahn</u>
---	-------------