

Automatic Speech Recognition (ASR)

Prof. Brian Mak

Department of Computer Science & Engineering

Elements: Basic Units in Chemistry

Periodic Table of the Elements

1 1A 1A																	18 VIIIA 8A						
1 H Hydrogen 1.008																	2 He Helium 4.003						
3 Li Lithium 6.941	4 Be Beryllium 9.012																	5 B Boron 10.811	6 C Carbon 12.011	7 N Nitrogen 14.007	8 O Oxygen 15.999	9 F Fluorine 18.998	10 Ne Neon 20.180
11 Na Sodium 22.990	12 Mg Magnesium 24.305	3 IIIB 3B	4 IVB 4B	5 VB 5B	6 VIB 6B	7 VIIB 7B	8 VIII 8	9 VIII 8	10 VIII 8	11 IB 1B	12 IIB 2B	13 Al Aluminum 26.982	14 Si Silicon 28.086	15 P Phosphorus 30.974	16 S Sulfur 32.066	17 Cl Chlorine 35.453	18 Ar Argon 39.948						
19 K Potassium 39.098	20 Ca Calcium 40.078	21 Sc Scandium 44.956	22 Ti Titanium 47.867	23 V Vanadium 50.942	24 Cr Chromium 51.996	25 Mn Manganese 54.938	26 Fe Iron 55.845	27 Co Cobalt 58.933	28 Ni Nickel 58.693	29 Cu Copper 63.546	30 Zn Zinc 65.38	31 Ga Gallium 69.723	32 Ge Germanium 72.631	33 As Arsenic 74.922	34 Se Selenium 78.972	35 Br Bromine 79.904	36 Kr Krypton 84.798						
37 Rb Rubidium 85.468	38 Sr Strontium 87.62	39 Y Yttrium 88.906	40 Zr Zirconium 91.224	41 Nb Niobium 92.906	42 Mo Molybdenum 95.95	43 Tc Technetium 98.907	44 Ru Ruthenium 101.07	45 Rh Rhodium 102.906	46 Pd Palladium 106.42	47 Ag Silver 107.868	48 Cd Cadmium 112.411	49 In Indium 114.818	50 Sn Tin 118.711	51 Sb Antimony 121.760	52 Te Tellurium 127.6	53 I Iodine 126.904	54 Xe Xenon 131.294						
55 Cs Cesium 132.905	56 Ba Barium 137.328	57-71	72 Hf Hafnium 178.49	73 Ta Tantalum 180.948	74 W Tungsten 183.84	75 Re Rhenium 186.207	76 Os Osmium 190.23	77 Ir Iridium 192.217	78 Pt Platinum 195.085	79 Au Gold 196.967	80 Hg Mercury 200.592	81 Tl Thallium 204.383	82 Pb Lead 207.2	83 Bi Bismuth 208.980	84 Po Polonium [208.982]	85 At Astatine 209.987	86 Rn Radon 222.018						
87 Fr Francium 223.020	88 Ra Radium 226.025	89-103	104 Rf Rutherfordium [261]	105 Db Dubnium [262]	106 Sg Seaborgium [266]	107 Bh Bohrium [264]	108 Hs Hassium [269]	109 Mt Meitnerium [268]	110 Ds Darmstadtium [269]	111 Rg Roentgenium [272]	112 Cn Copernicium [277]	113 Uut Ununtrium unknown	114 Fl Flerovium [289]	115 Uup Ununpentium unknown	116 Lv Livermorium [293]	117 Uus Ununseptium unknown	118 Uuo Ununoctium unknown						

Lanthanide Series	57 La Lanthanum 138.905	58 Ce Cerium 140.116	59 Pr Praseodymium 140.908	60 Nd Neodymium 144.242	61 Pm Promethium 144.913	62 Sm Samarium 150.36	63 Eu Europium 151.964	64 Gd Gadolinium 157.25	65 Tb Terbium 158.925	66 Dy Dysprosium 162.500	67 Ho Holmium 164.930	68 Er Erbium 167.259	69 Tm Thulium 168.934	70 Yb Ytterbium 173.055	71 Lu Lutetium 174.967
Actinide Series	89 Ac Actinium 227.028	90 Th Thorium 232.038	91 Pa Protactinium 231.036	92 U Uranium 238.029	93 Np Neptunium 237.048	94 Pu Plutonium 244.064	95 Am Americium 243.061	96 Cm Curium 247.070	97 Bk Berkelium 247.070	98 Cf Californium 251.080	99 Es Einsteinium [254]	100 Fm Fermium 257.095	101 Md Mendelevium 258.1	102 No Nobelium 259.101	103 Lr Lawrencium [262]

Alkali Metal	Alkaline Earth	Transition Metal	Basic Metal	Semimetal	Nonmetal	Halogen	Noble Gas	Lanthanide	Actinide
--------------	----------------	------------------	-------------	-----------	----------	---------	-----------	------------	----------

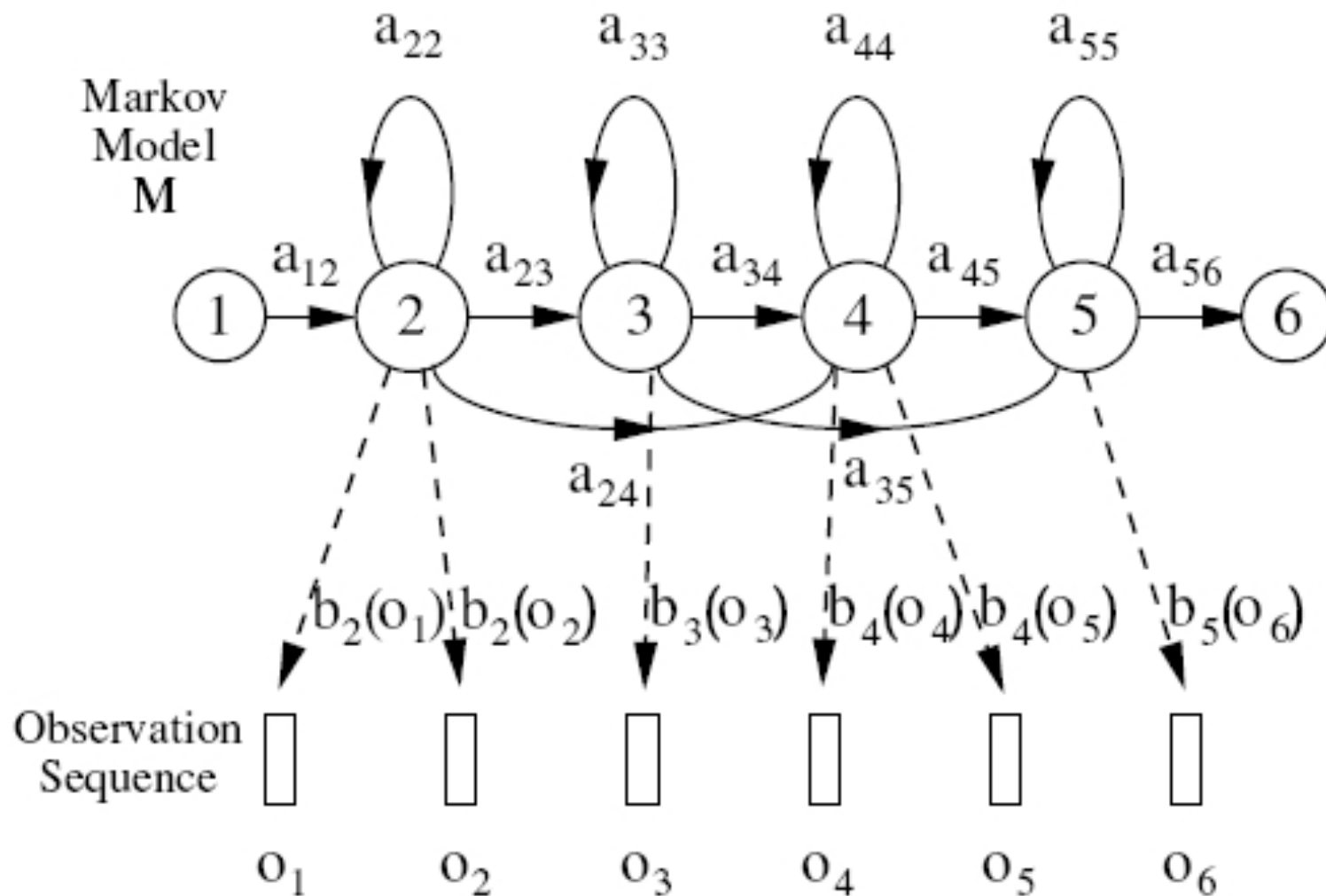
Phonemes: Basic Units in Speech

VOWELS	monophthongs				diphthongs		Phonemic Chart voiced unvoiced	
	i:	ɪ	ʊ	u:	ɪə	eɪ		
	sheep	ship	good	shoot	here	wait		
	e	ə	ɜ:	ɔ:	ʊə	ɔɪ	əʊ	
	bed	teacher	bird	door	tourist	boy	show	
	æ	ʌ	ɑ:	ɒ	eə	aɪ	aʊ	
	cat	up	far	on	hair	my	cow	
CONSONANTS	p	b	t	d	tʃ	dʒ	k	g
	pea	boat	tea	dog	cheese	June	car	go
	f	v	θ	ð	s	z	ʃ	ʒ
	fly	video	think	this	see	zoo	shall	television
	m	n	ŋ	h	l	r	w	j
	man	now	sing	hat	love	red	wet	yes

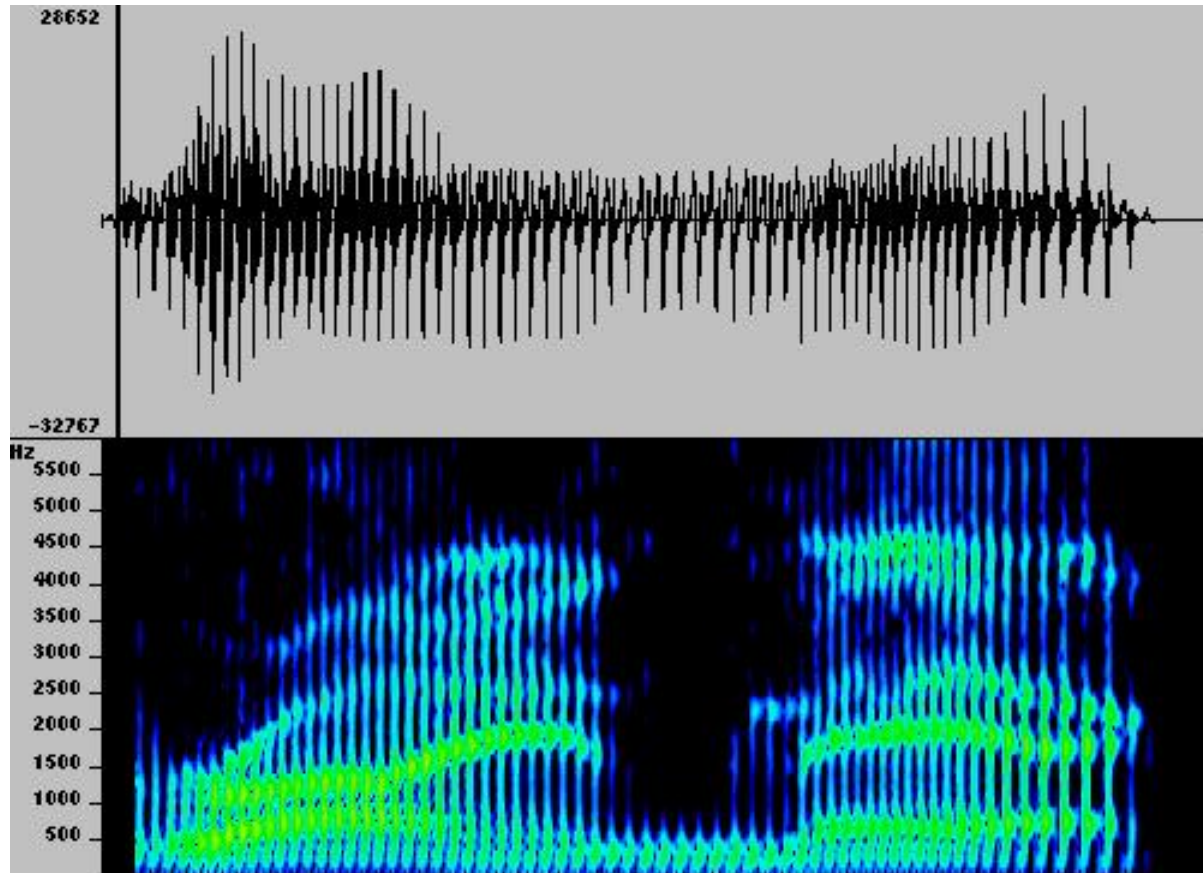
Phonemes

- English: 40 – 60 **phonemes**
- A spoken word is a **sequence** of phonemes:
 - cat = /k/ /ae/ /t/
- ASR: to recognize an **utterance** which is a sequence of words
 - => recognize a **sequence** of **phonemes**
 - “a pretty cat” = /ah/ /p/ /r/ /ih/ /t/ /iy/ /k/ /ae/ /t/

Acoustic Modeling: Hidden Markov Model

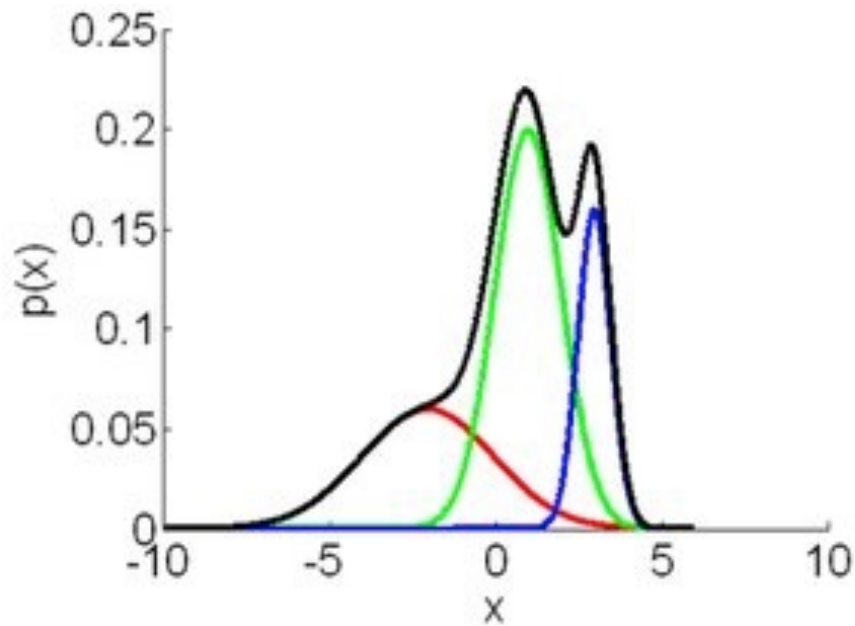


Visible Speech



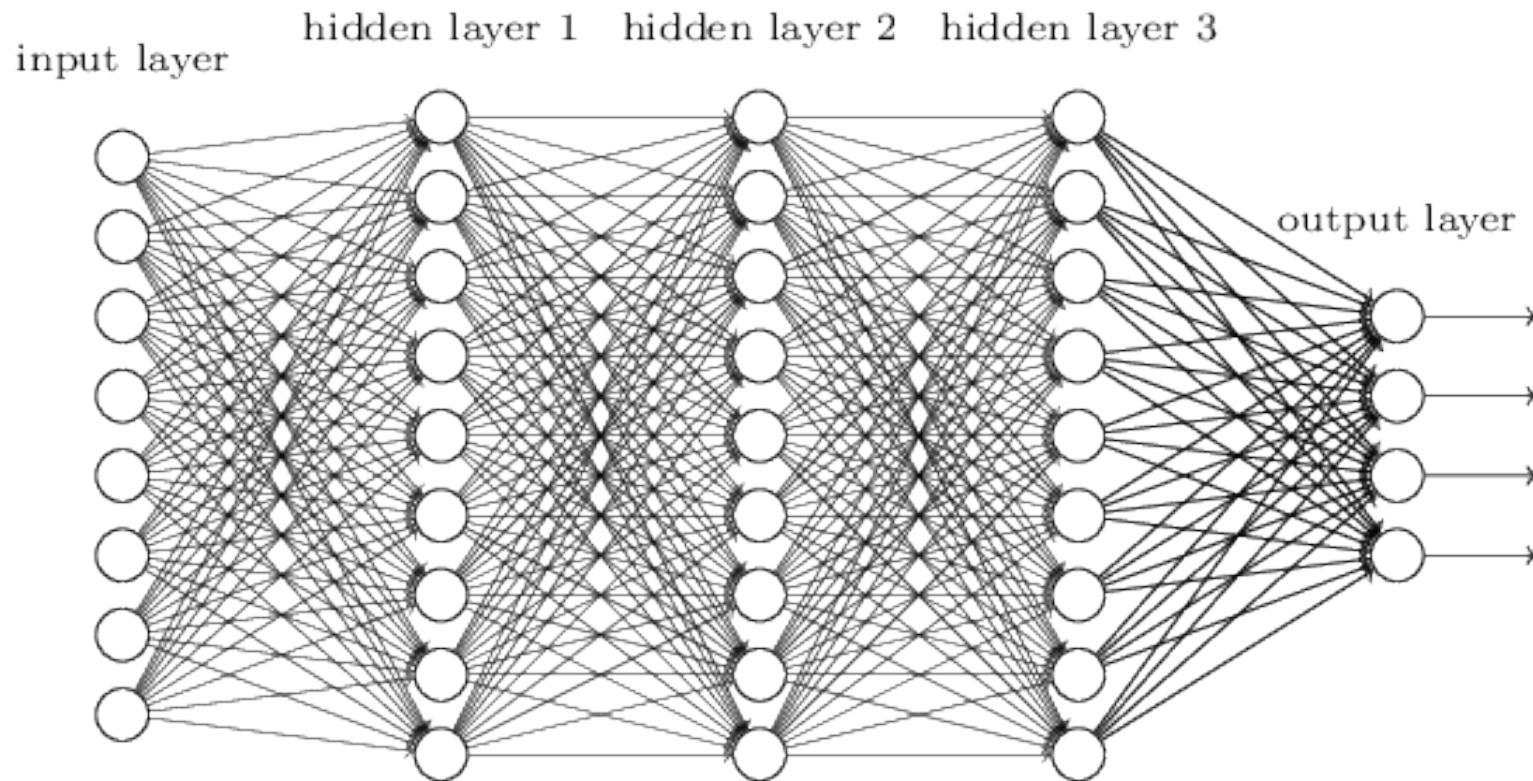
spectrogram

Gaussian Mixture Model (GMM)



$$p(x) = \sum_{j=1}^K w_j \cdot N(x | \mu_j, \Sigma_j)$$

Deep Neural Network (DNN)



DNN ..

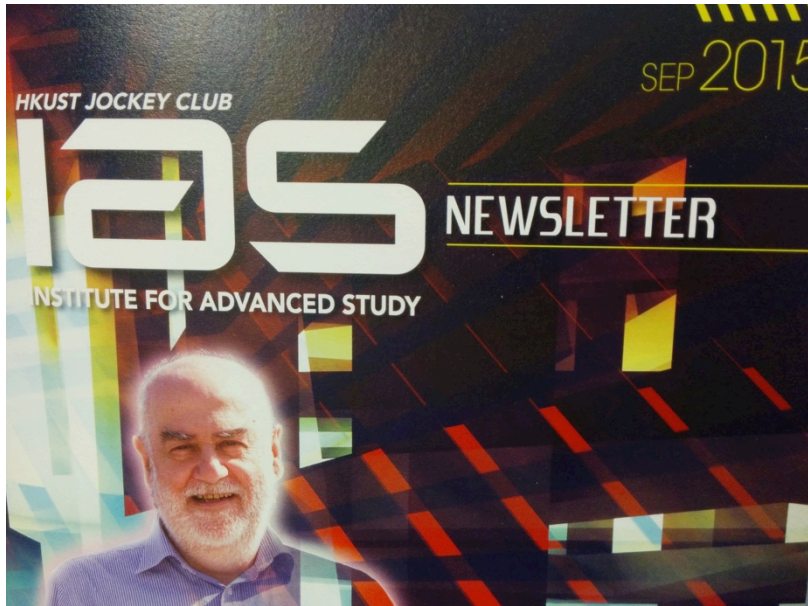
- #input units: 500 – 1000
- #hidden layers: 5 – 7
- #hidden units / layer : 1K – 2K
- #output units: 5K – 10K
 - recently we shows how to train 60K output units and get better performance—a big net!
- Total #weights: 30M – 40M (150M)

ASR Progress

- [Switchboard](#): 309-hr spontaneous conversations

YEAR	LAB-SCHOOL / TECHNIQUE	Word Error Rate
1992		>70.0%
1997	(CMU) GMM + extra 2000-hr data	35.1%
< 2005	GMM + “careful engineering” + extra data	< 20%
2011	Microsoft / DNN	16.1%
2014	IBM / DNN+CNN	10.4%
2015	IBM / maxout+CNN+DNN+NNLM+extra data	8%
	HUMAN	~4%

Epilogue



leader among scientific institutions
become one of the best
institutions in the world for sure.
happening. We see so many scholars
interesting talks and lectures in
mathematics and many other subjects.

My advice to students is not to believe in people who tell you that something is impossible. The major breakthroughs in science have always proven the naysayers wrong.