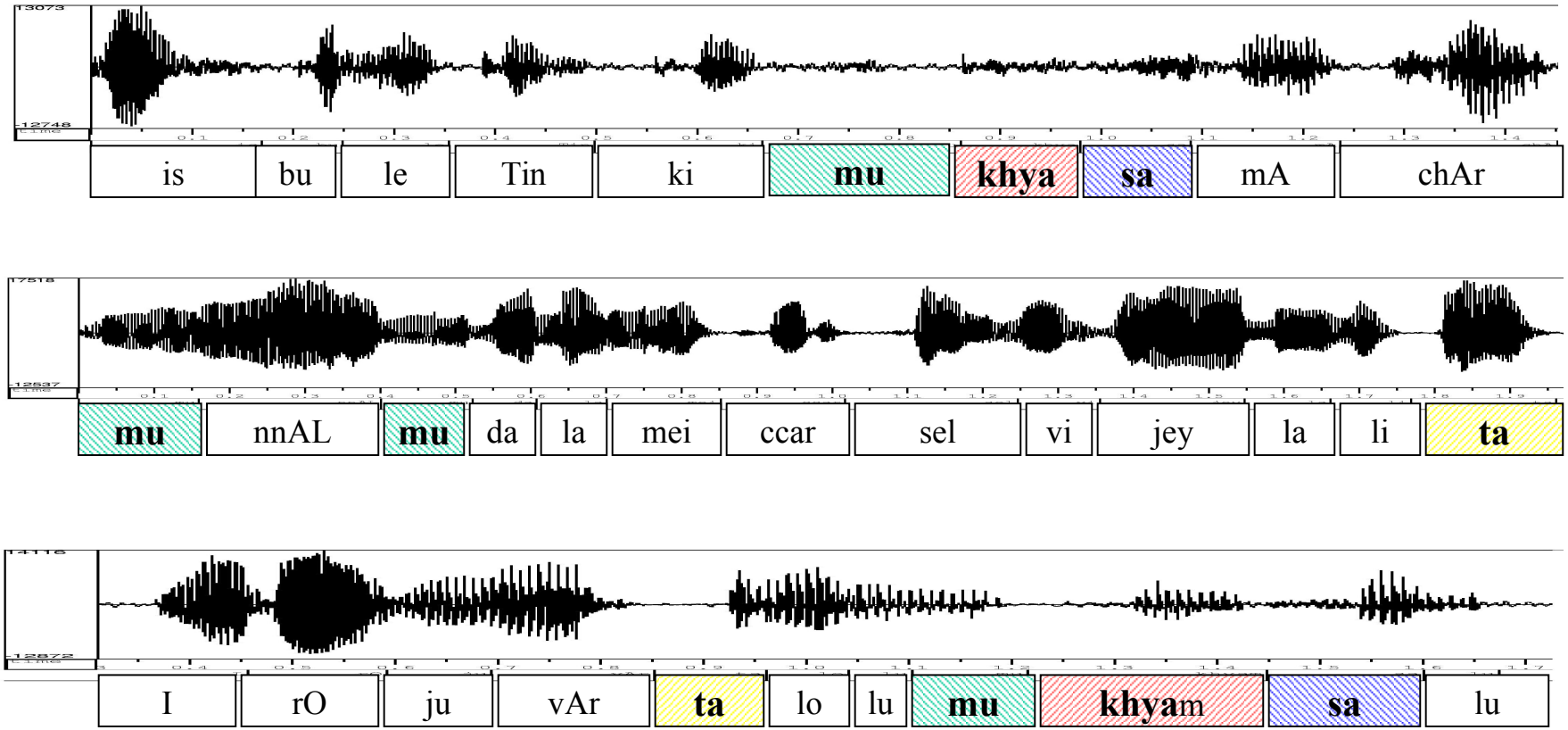
Basics of Speech Recognition

Dr. Suryakanth V. Gangashetty

International Institute of Information Technology (IIIT)
Hyderabad - 500 032, India
Email: svg@iiit.ac.in

Speech Signal-to-Symbol Transformation



Phonetic engine: **Capable of speech signal-to-symbol transformation independent of vocabulary and language**

What is Speech Technology ?

The tools and techniques for making a computer ‘understand’ human speech, and respond to it.

Why Speech Technology ?

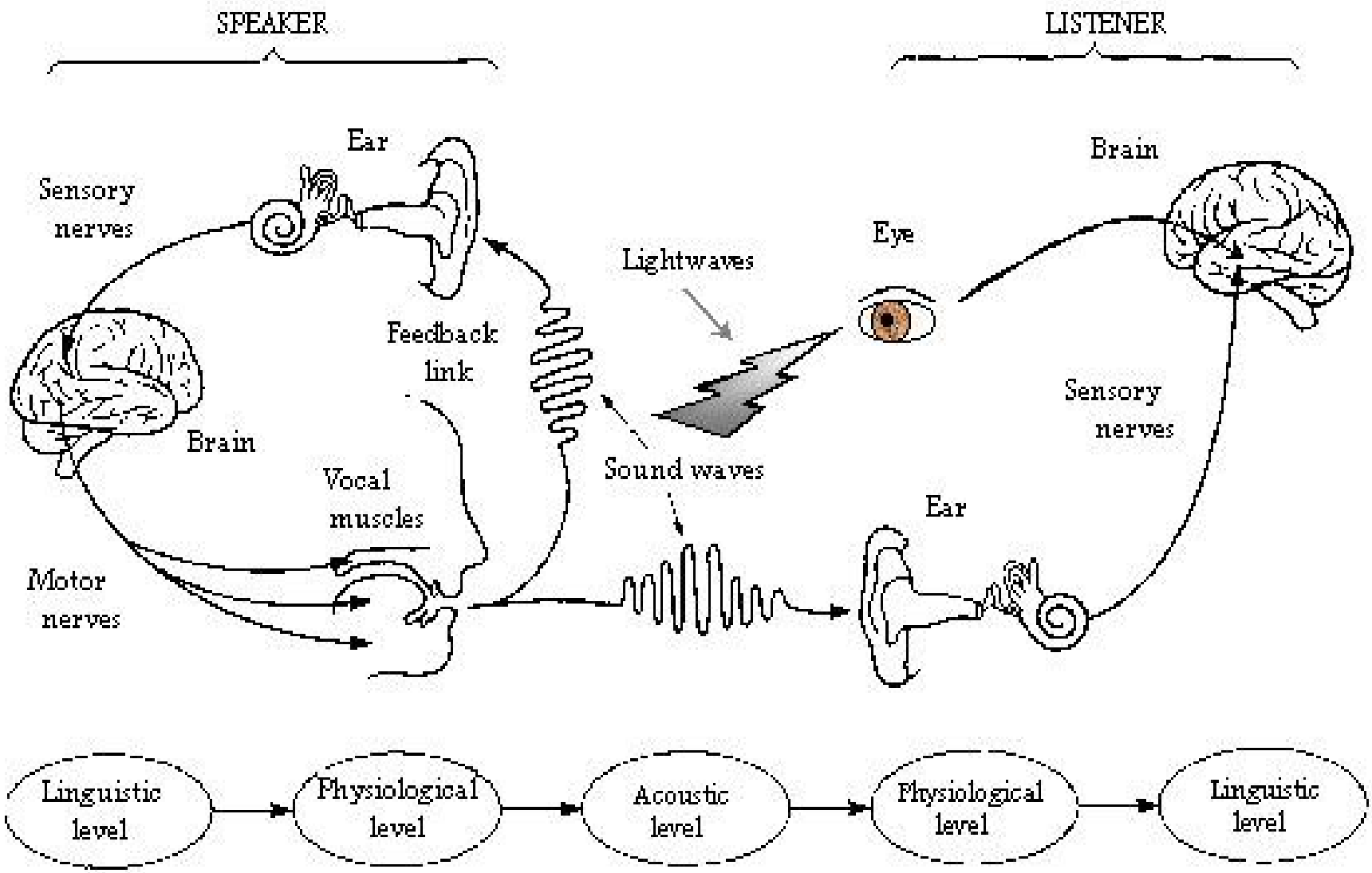
Speech is the most natural, versatile and convenient form of communication.

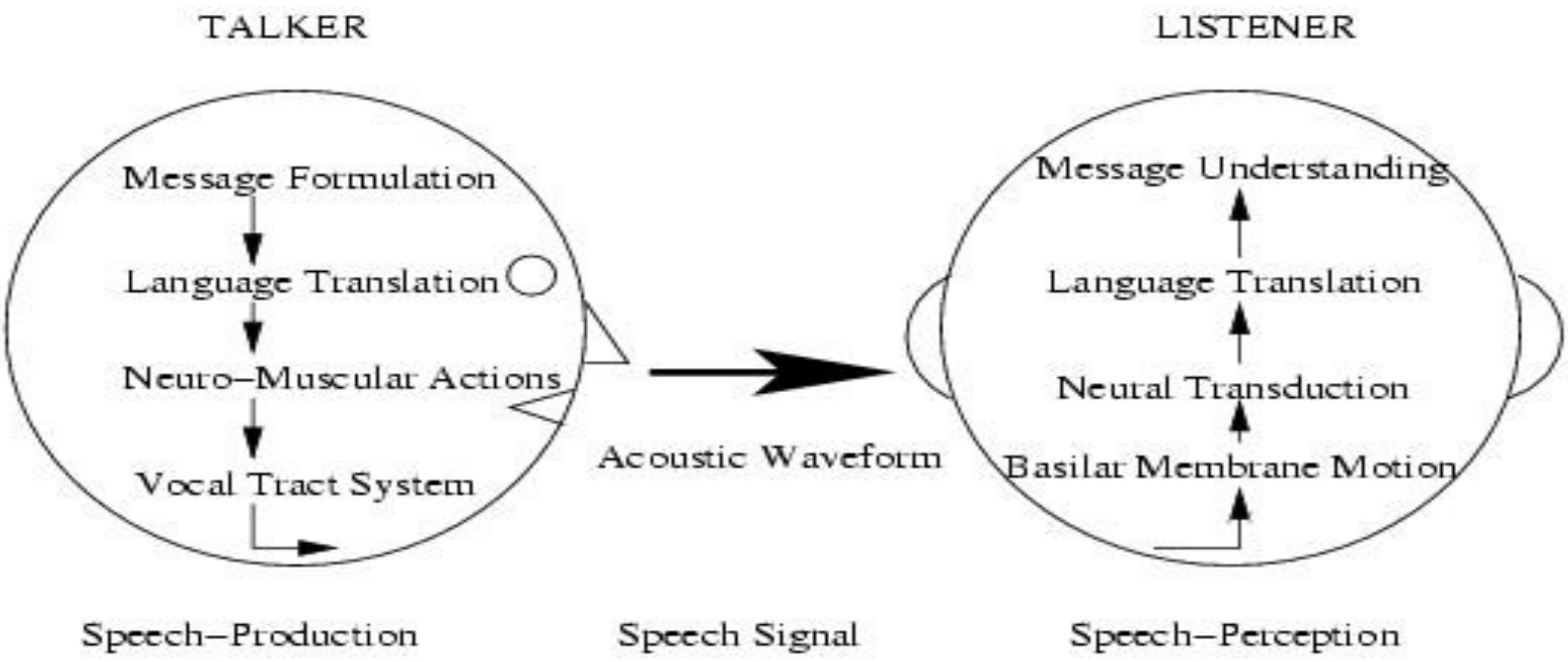
Hence, it is best suited as an interface between humans and machines.

Speech Technology: An Interdisciplinary Approach

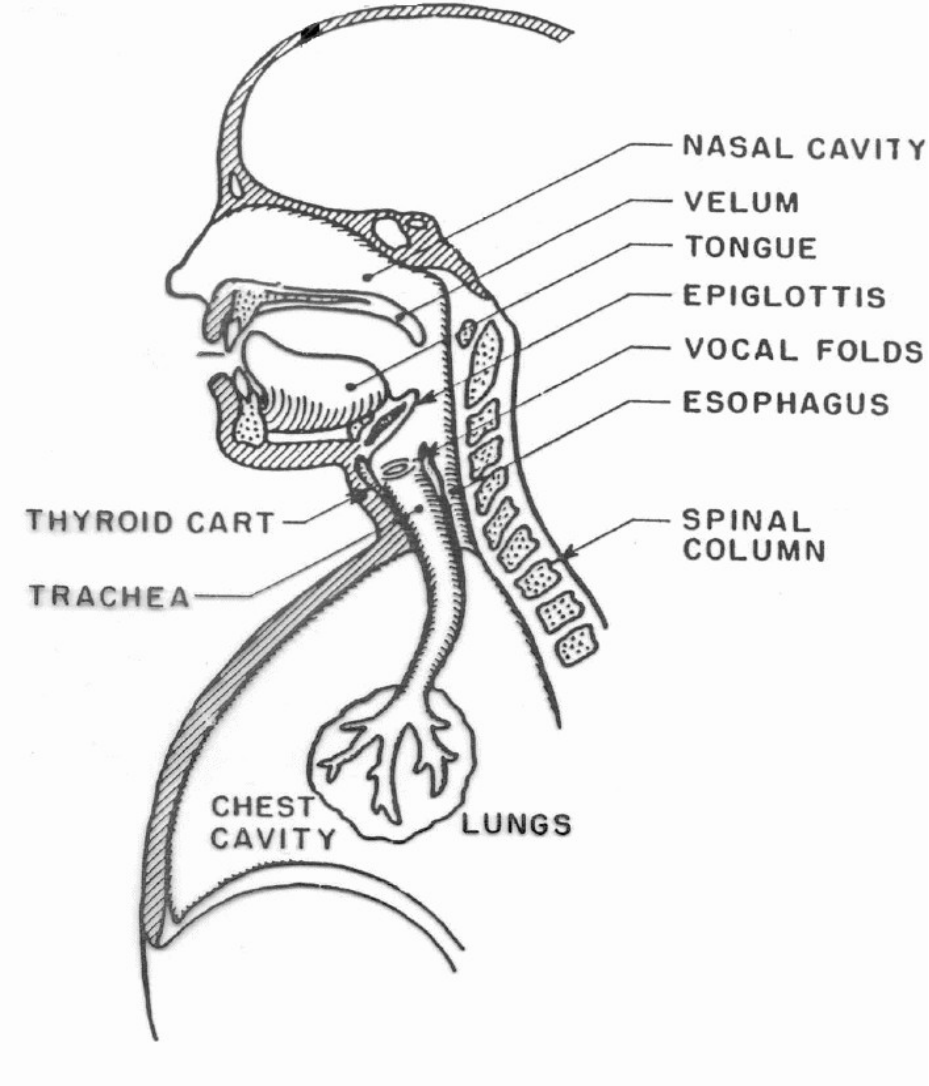
- Signal processing
- Acoustics
- Pattern recognition
- Information theory
- Linguistics
- Physiology
- Computer Science

Speech Production and Perception Process

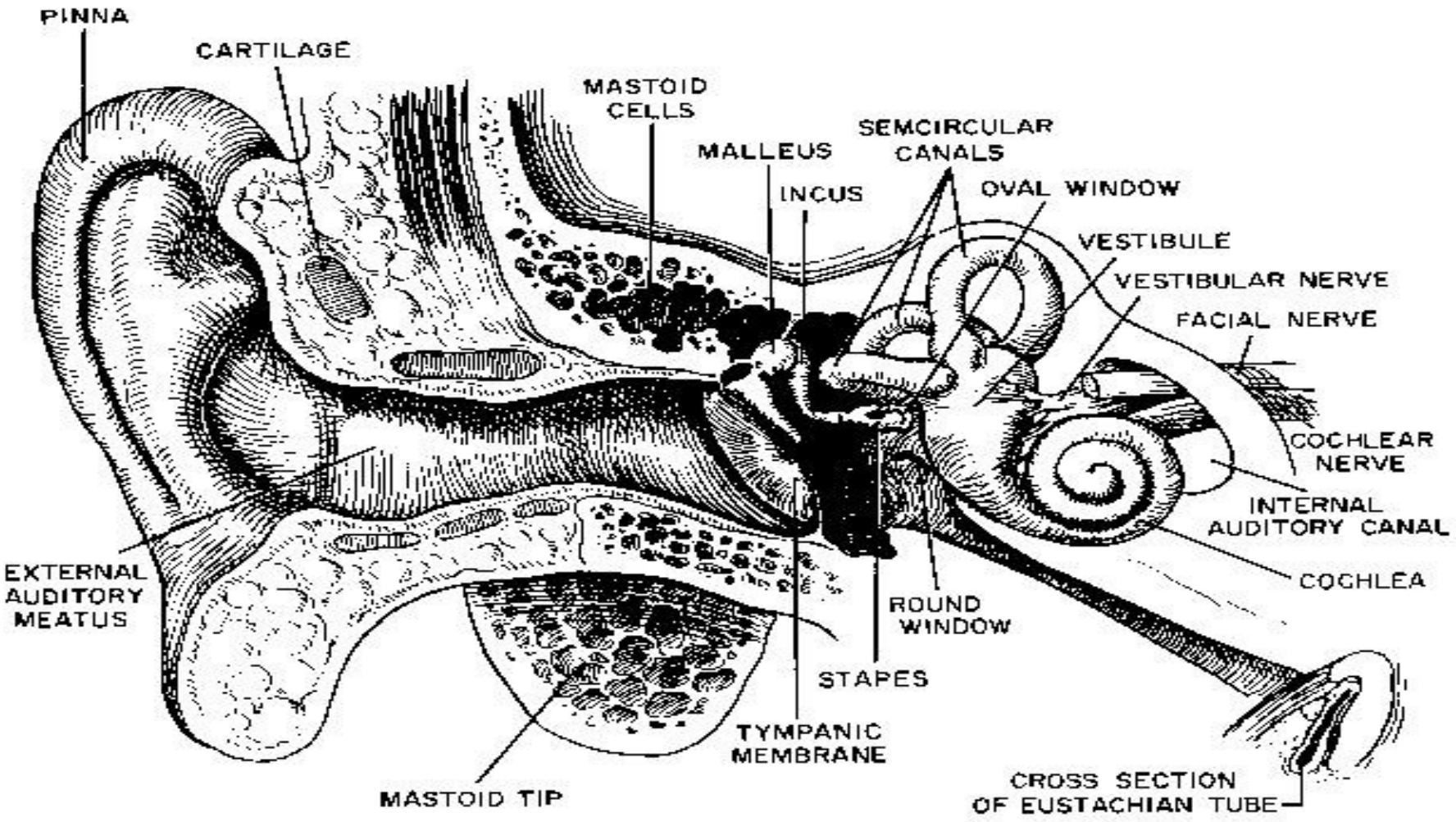



TALKER
LISTENER
Message Formulation
Language Translation
Neuro–Muscular Actions
Vocal Tract System
Acoustic Waveform
Message Understanding
Language Translation
Neural Transduction
Basilar Membrane Motion
Speech–Production
Speech Signal
Speech–Perception

Speech Production Mechanism



Structure and Mechanism of the Human Ear



Different Types of Sounds

Sound units in Indian Languages

Characters representing vowel sounds

Short vowels	/a/(अ)	/i/(इ)	/u/(उ)	/e/(ए)	/o/(ओ)
Long vowels	/a:/(आ)	/i:/(ई)	/u:/(ऊ)	/e:/(ए.)	/o:/(ओ.)
Diphthongs	/ai/(ऐ)		/au/(औ)		

Characters representing Consonant-Vowel (CV) combinations with vowel /a/(अ)

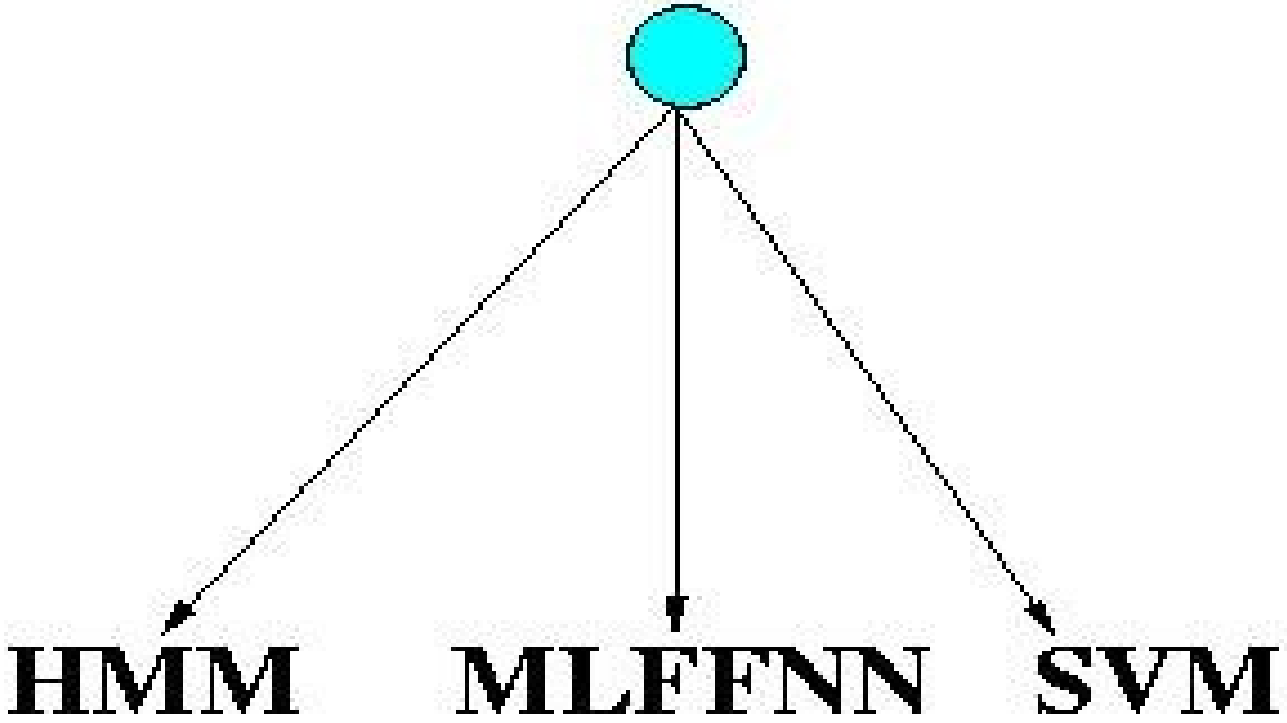
Place of articulation	Manner of articulation				Nasals	Semivowels	Fricatives
	Unvoiced		Voiced				
	Unaspirated	Aspirated	Unaspirated	Aspirated			
Velar	/ka/(क)	/kha/(ख)	/ga/(ग)	/gha/(घ)	/kna/(ङ)	----	/ha/(ह)
Palatal	/cha/(च)	/chha/(छ)	/ja/(ज)	/jha/(झ)	/chna/(ञ)	/ya/(य)	/sha/(श)
Alveolar	/Ta/(ट)	/Tha/(ठ)	/Da/(ड)	/Dha/(ढ)	/Tna/(ण)	/ra/(र)	/shha/(ष)
Dental	/ta/(त)	/tha/(थ)	/da/(द)	/dha/(थ)	/na/(न)	/la/(ल)	/sa/(स)
Bilabial	/pa/(प)	/pha/(फ)	/ba/(ब)	/bha/(भ)	/ma/(म)	/va/(व)	----

< Back

Issues in the Development of a Phonetic Engine

1. Choice of subword unit of speech
2. Number of subword unit classes
3. Frequency of occurrence of subword units in different languages
4. Variability in the characteristics of speech signal for a subword unit
5. Acoustic similarity among subword units
6. Classification models for recognition of subword units
7. Representation of subword units of speech
8. Compression of large dimensional pattern vectors of the subword units
9. Recognition of subword units in continuous speech
10. Recognition of subword units in multiple languages

Classification Models



Soft Computing

- Soft computing consists of several computing paradigms, including neural networks, fuzzy set theory, approximate reasoning, and derivative-free optimization methods such as genetic algorithms and simulated annealing

Soft Computing Approach

An innovative approach to constructing computationally intelligent systems. An emerging approach to computing which parallels the remarkable ability of human mind to reason and learn in an environment of uncertainty and imprecision

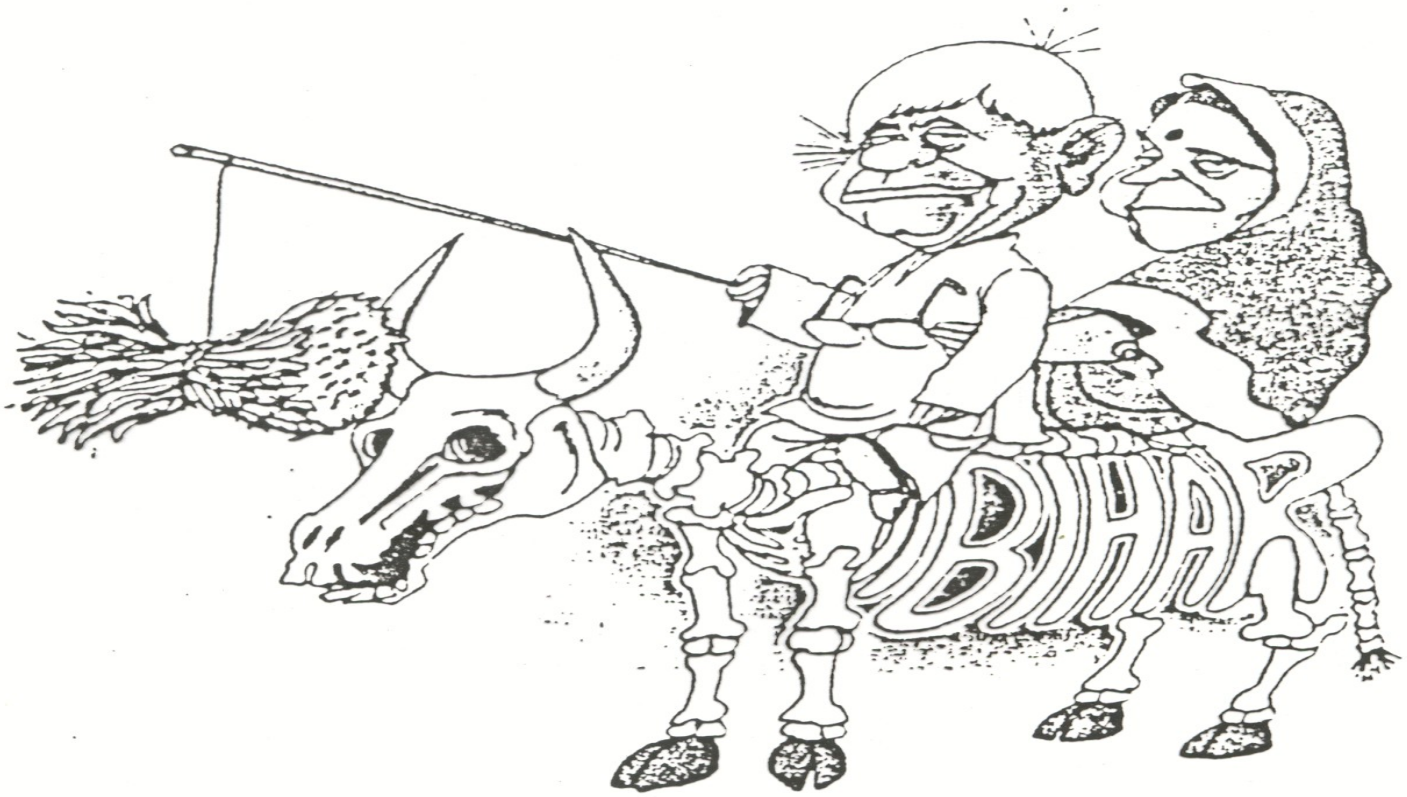
- Complex real-world problems require intelligent systems that combine knowledge, techniques and methodologies from various sources
- These intelligent systems are supposed to possess humanlike expertise within a specific domain, adapt themselves and learn to do better in changing environments, and explain how they make decisions or take actions

Can Computers be Intelligent ?

Intelligence: Creativity, Skill, Consciousness, Emotion and Intuition

- Mid 1900's – Alan Turing
- Machines could be created that would mimic processes of the human brain
- There was nothing the brain could do a well-designed computer could not
- Fifty years later his statements are still visionary

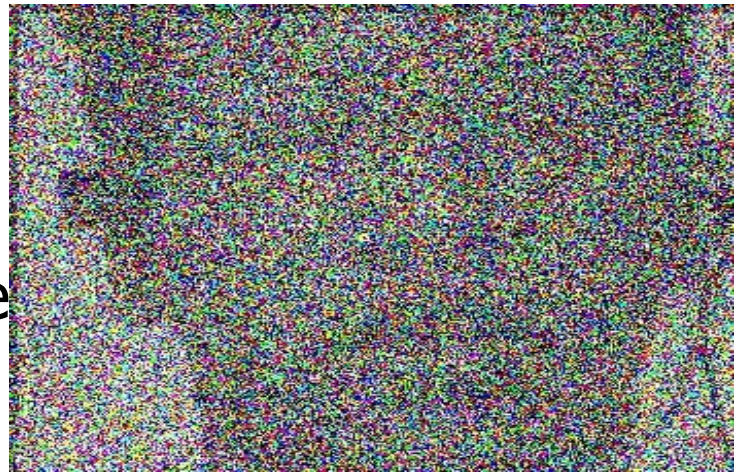
Humans Perceive Cues, Edges

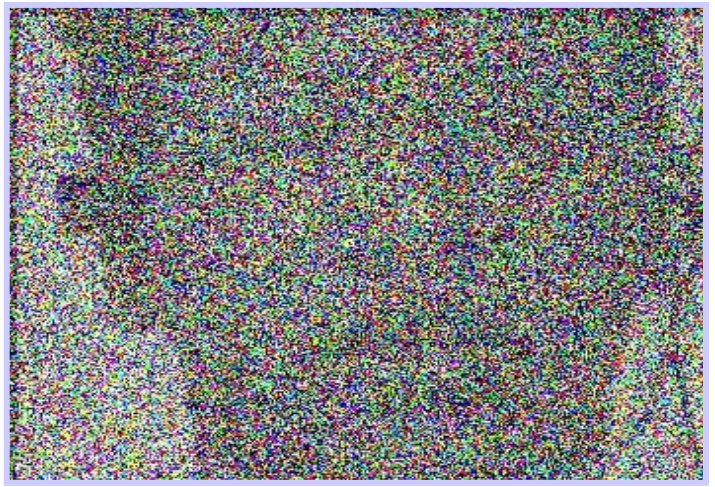


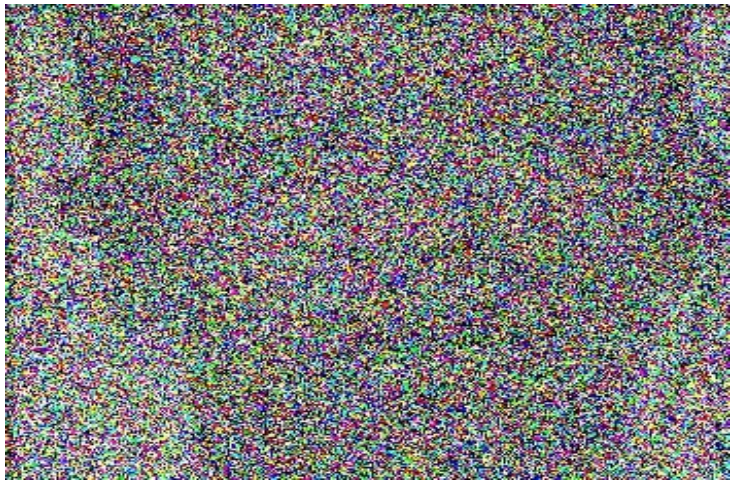
Robust Perception Against Noise

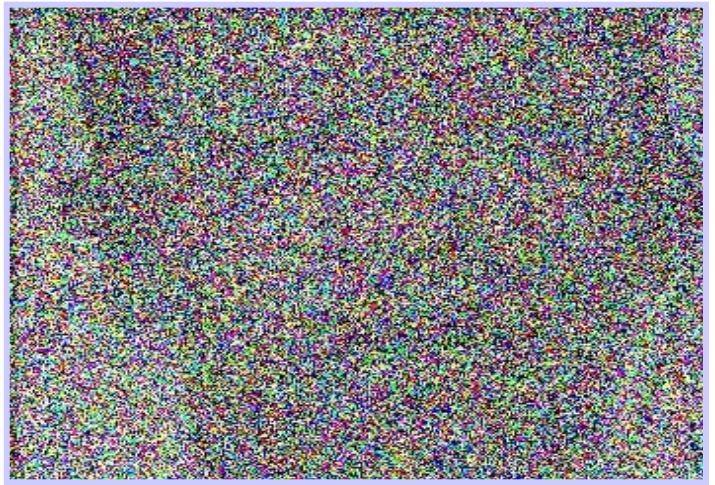
Less noise

More noise









Still Frame

Video Sequence

Humans Perceive Cues, Edges

Can Computers be Intelligent ?

Intelligence: Creativity, Skill, Consciousness, Emotion and Intuition

- Mid 1900's – Alan Turing
- Machines could be created that would mimic processes of the human brain
- There was nothing the brain could do a well-designed computer could not
- Fifty years later his statements are still visionary

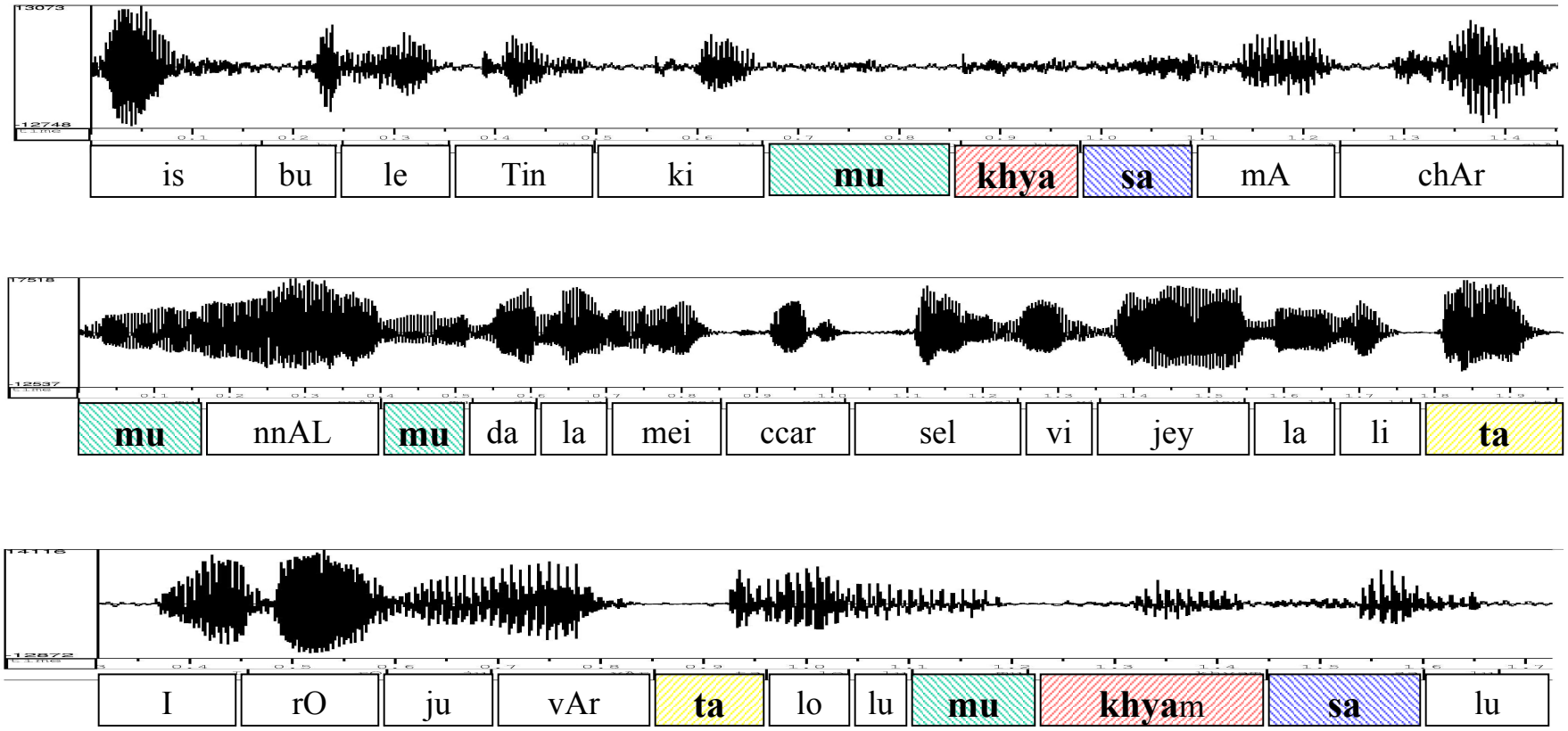
Perception: Human Vs Machines

	Machine	Human
Representation	Pixels/samples Discrete numbers (data)	Cues, symbols and interrelations (patterns)
Processor	Single	Multiple (neurons)
Processing	Sequential (local)	Parallel and distributed (local and global)

Characteristics of Human Problem Solving

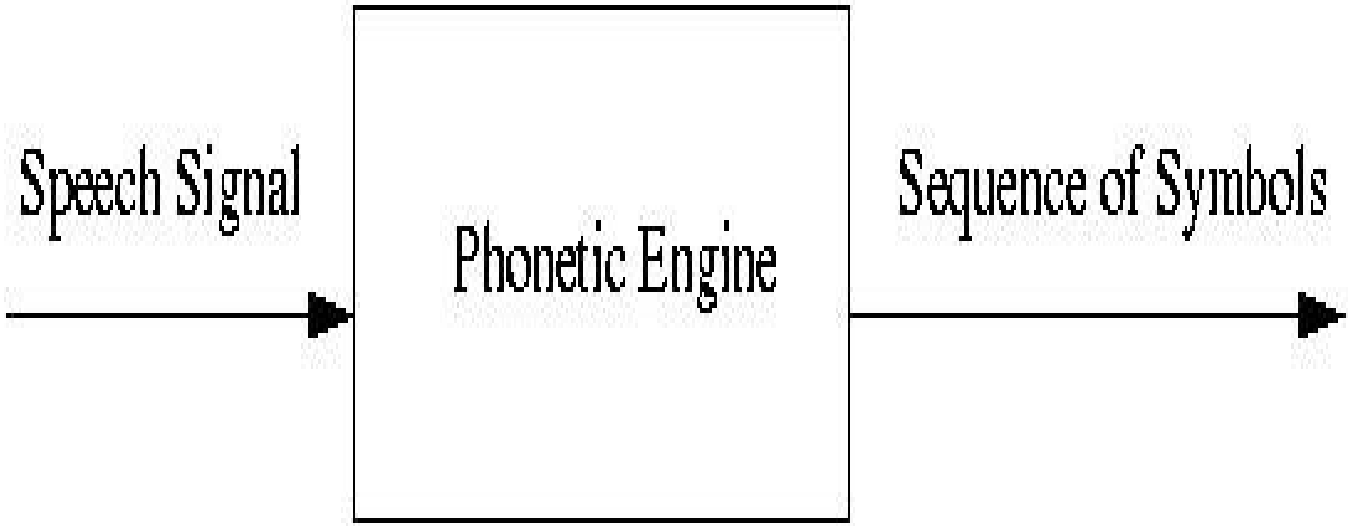
- Essentially *pattern* processing instead of data processing
- Integration of local and global patterns
- Delayed decisions
- Nonuniqueness of solutions

Speech Signal-to-Symbol Transformation



Phonetic engine: **Capable of speech signal-to-symbol transformation independent of vocabulary and language**

Phonetic Engine



Approaches to Speech Signal-to-Symbol Transformation

- **Approach based on segmentation and labeling**
  - **Segmentation of continuous speech signal into regions of subword units**
  - **Assignment of labels to the segmented regions using a subword unit classifier**

- **Approach based on spotting subword units in continuous speech**
  - **Detection of anchor points in continuous speech**
  - **Assignment of labels to the segments around the anchor points using a subword unit classifier**

HMM Framework

- **Assumes that the model (topology and density functions) actually reflects the structure of the data**

  - Learning the structure from the data would be a better idea

- **Incremental model optimization approach in an Maximum Likelihood (ML) framework simplifies the training process**

  - Discriminative information is not used in training. All rival state sequences (corresponding to other models) are not considered during optimization of parameters for a given model

Soft Computing Approach

An innovative approach to constructing computationally intelligent systems. An emerging approach to computing which parallels the remarkable ability of human mind to reason and learn in an environment of uncertainty and imprecision

- Complex real-world problems require intelligent systems that combine knowledge, techniques and methodologies from various sources
- These intelligent systems are supposed to possess humanlike expertise within a specific domain, adapt themselves and learn to do better in changing environments, and explain how they make decisions or take actions

Can Computers be Intelligent ?

Intelligence: Creativity, Skill, Consciousness, emotion and Intuition

- Mid 1900's – Alan Turing
- Machines could be created that would mimic processes of the human brain
- There was nothing the brain could do a well-designed computer could not
- Fifty years later his statements are still visionary

- Soft computing consists of several computing paradigms, including neural networks, fuzzy set theory, approximate reasoning, and derivative-free optimization methods such as genetic algorithms and simulated annealing

- Soft computing consists of several computing paradigms, including neural networks, fuzzy set theory, approximate reasoning, and derivative-free optimization methods such as genetic algorithms and simulated annealing

Biological Neural Networks

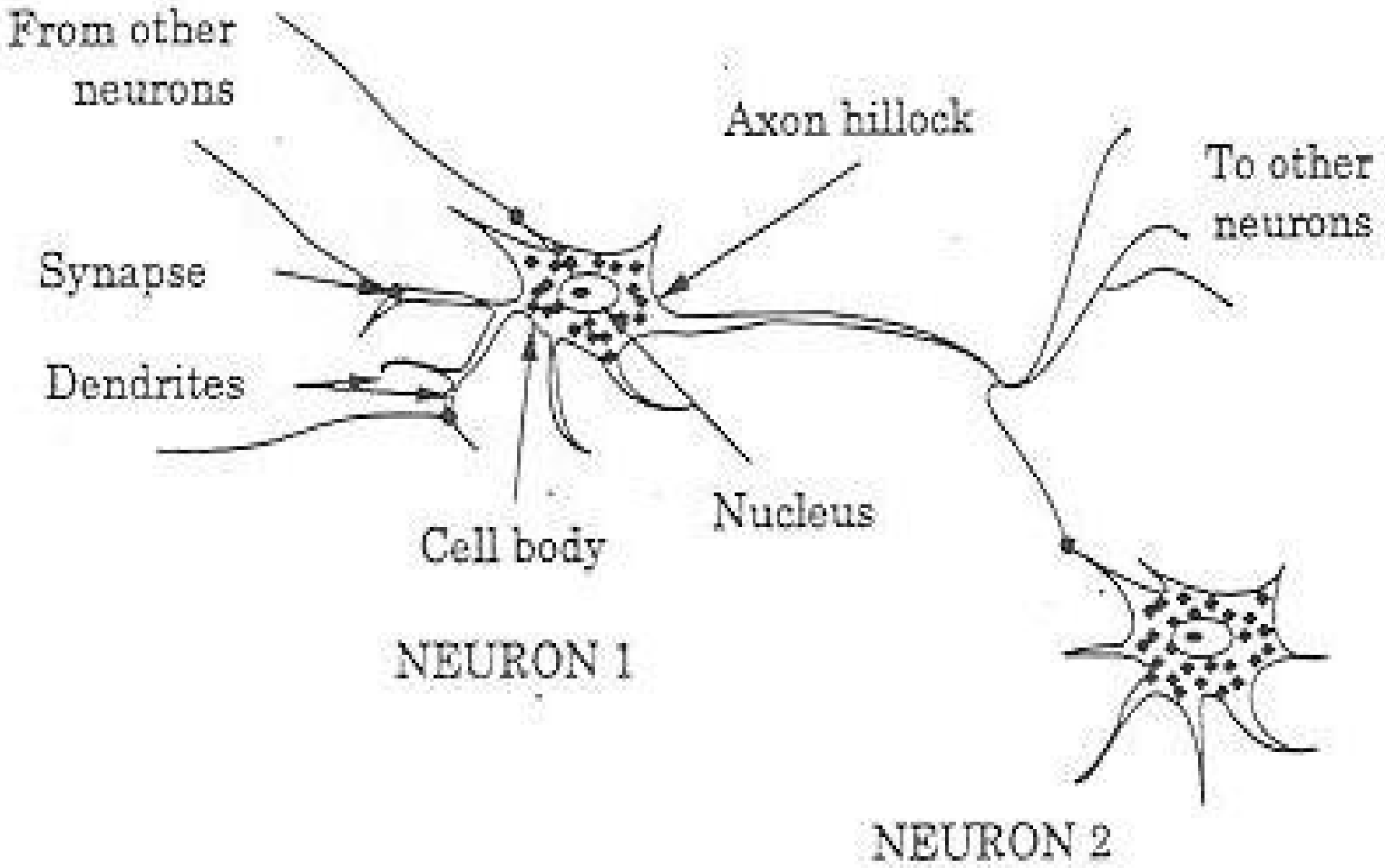


Figure 1.1 Schematic diagram of a typical neuron or nerve cell.

Biological Neural Networks

- Structure and function: Neurons, interconnections, dynamics for learning and recall
- Features: Robustness, fault tolerance, flexibility, ability to deal with variety of data situations, collective computation
- Comparison with computers: Speed, processing, size and complexity, fault tolerance, control mechanism
- Parallel and Distributed Processing (PDP) models

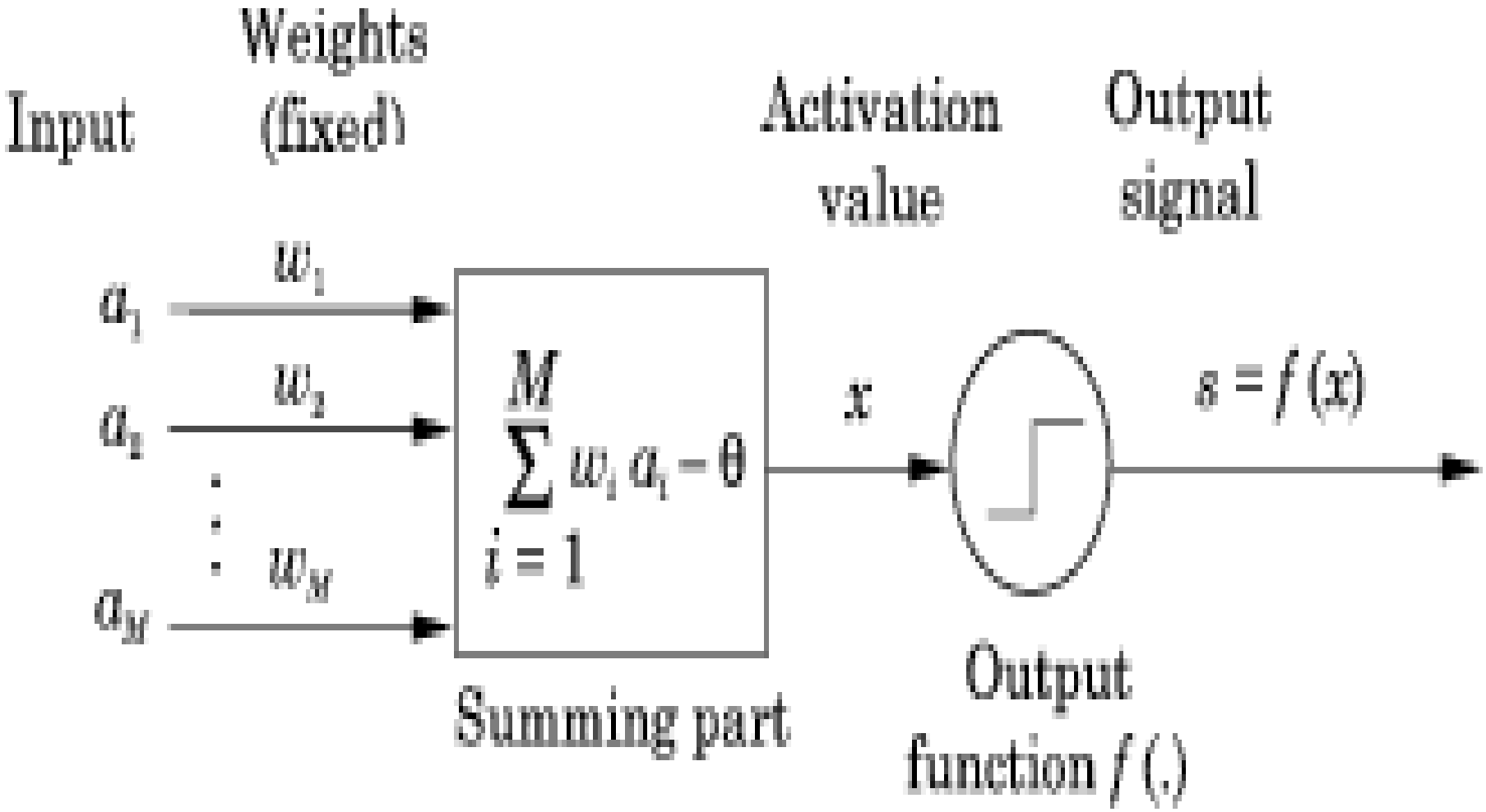
Artificial Neural Networks (ANN)

- Problem solving: Pattern recognition tasks by human and machine
- Pattern vs data
- Pattern processing vs data processing
- Architectural mismatch
- Need for new models of computing

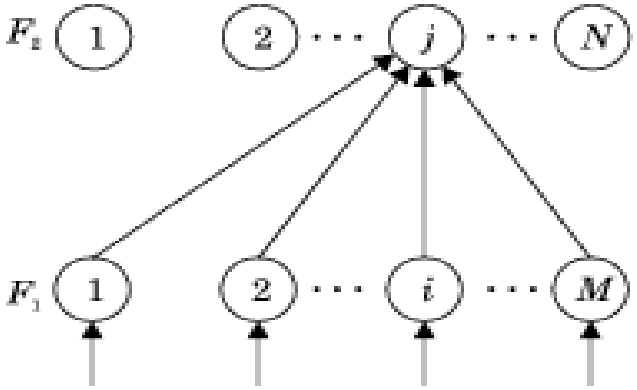
Basics of ANN

- ANN terminology: Processing unit
- Interconnection, operation and update (input, weights, activation value, output function, output value)
- Models of neurons: MP neuron, perceptron and adaline
- Topology
- Basic learning laws

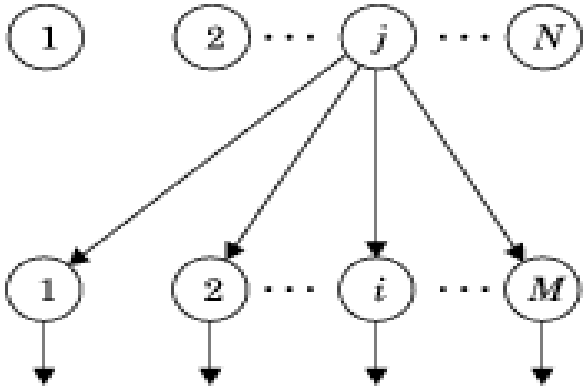
Model of a Neuron



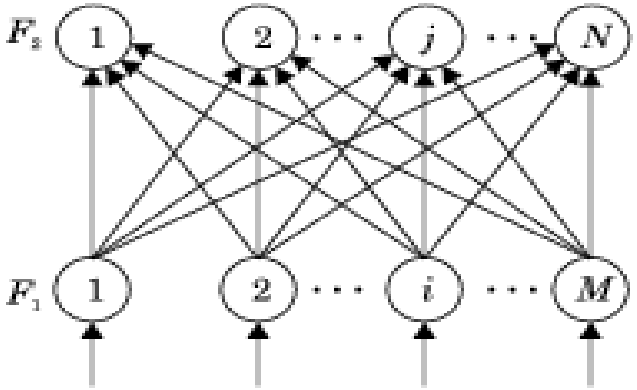
Topology



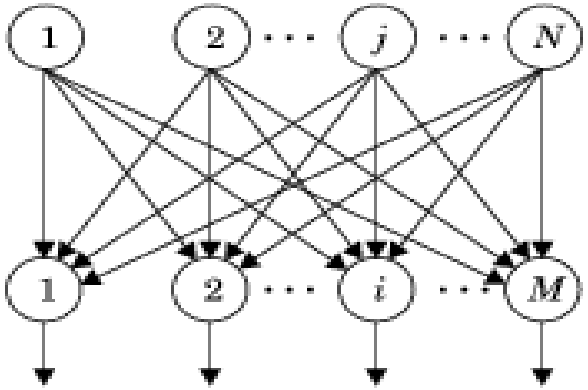
(a) Instar



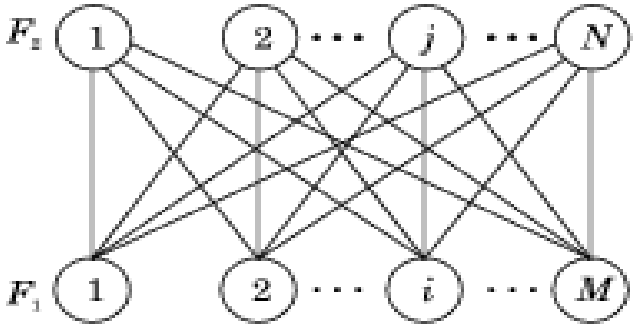
(b) Outstar



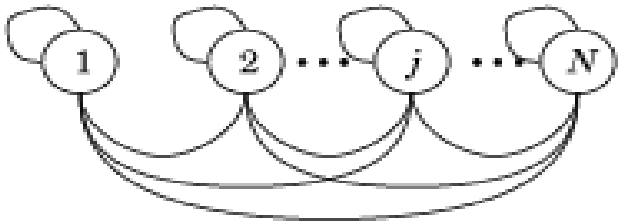
(c) Group of instars



(d) Group of outstars



(e) Bidirectional associative memory



(f) Autoassociative memory

<back

Basic Learning Laws

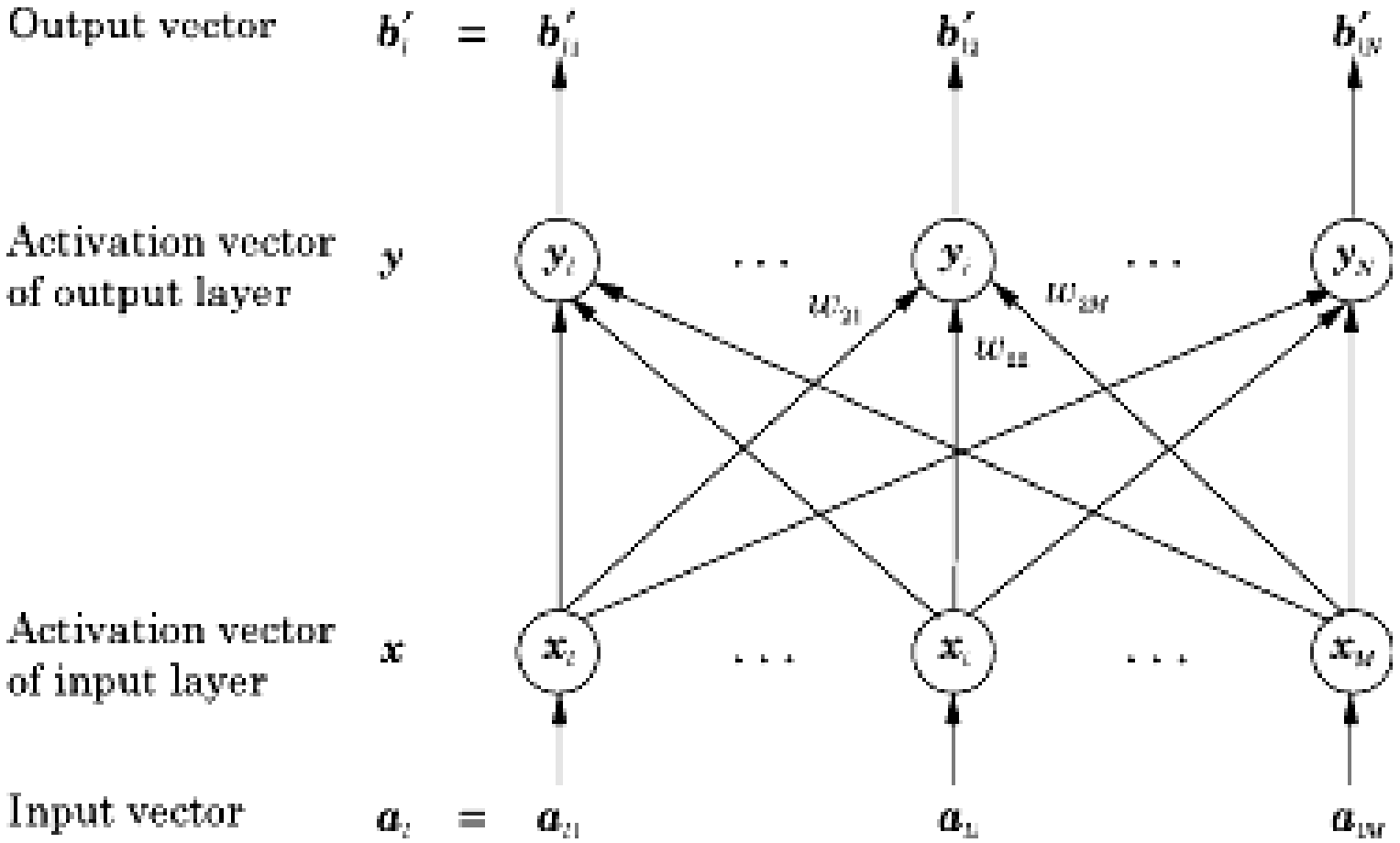
Learning law	Weight adjustment Δw_{ij}	Initial weights	Learning
Hebbian	$\Delta w_{ij} = \eta f(\mathbf{w}_i^T \mathbf{a}) a_j$ $= \eta s_i a_j$, for $j = 1, 2, \ldots, M$	Near zero	Unsupervised
Perceptron	$\Delta w_{ij} = \eta [b_i - \text{sgn}(\mathbf{w}_i^T \mathbf{a})] a_j$ $= \eta (b_i - s_i) a_j$, for $j = 1, 2, \ldots, M$	Random	Supervised
Delta	$\Delta w_{ij} = \eta [b_i - f(\mathbf{w}_i^T \mathbf{a})] \dot{f}(\mathbf{w}_i^T \mathbf{a}) a_j$ $= \eta [b_i - s_i] \dot{f}(x_i) a_j$, for $j = 1, 2, \ldots, M$	Random	Supervised
Widrow-Hoff	$\Delta w_{ij} = \eta [b_i - \mathbf{w}_i^T \mathbf{a}] a_j$, for $j = 1, 2, \ldots, M$	Random	Supervised
Correlation	$\Delta w_{ij} = \eta b_i a_j$, for $j = 1, 2, \ldots, M$	Near zero	Supervised
Winner-take-all	$\Delta w_{kj} = \eta (a_j - w_{kj})$, k is the winning unit, for $j = 1, 2, \ldots, M$	Random but normalised	Unsupervised
Outstar	$\Delta w_{jk} = \eta (b_j - w_{jk})$, for $j = 1, 2, \ldots, M$	Zero	Supervised

<back

Functional Units and Pattern Recognition Tasks

- Feedforward ANN
  - Pattern association
  - Pattern classification
  - Pattern mapping/classification
- Feedback ANN
  - Autoassociation
  - Pattern storage (LTM)
  - Pattern environment storage (LTM)
- Feedforward and Feedback (Competitive Learning) ANN
  - Pattern storage (STM)
  - Pattern clustering
  - Feature map

Two Layer Feedforward Neural Network (FFNN)



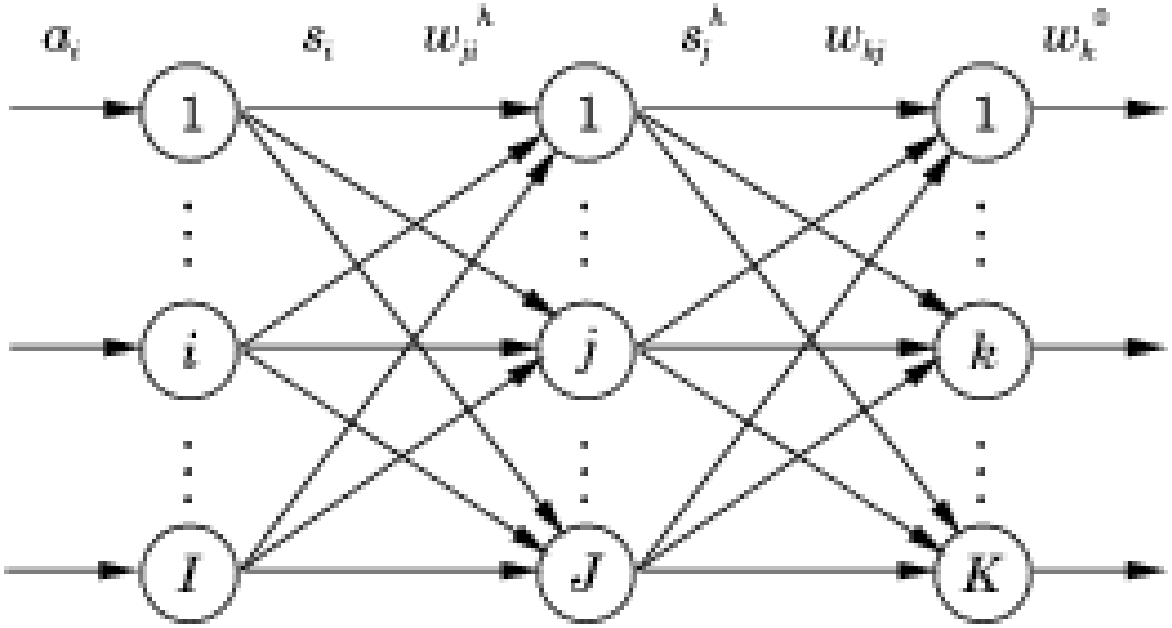
PR Tasks by FFNN

- Pattern association
  - Architecture: Two layers, linear processing, single set of weights
  - Learning:, Hebb's (orthogonal) rule, Delta (linearly independent) rule
  - Recall: Direct
  - Limitation: Linear independence, number of patterns restricted to input dimensionality
  - To overcome: Nonlinear processing units, leads to a pattern classification problem
- Pattern classification
  - Architecture: Two layers, nonlinear processing units, geometrical interpretation
  - Learning: Perceptron learning
  - Recall: Direct
  - Limitation: Linearly separable functions, cannot handle hard problems
  - To overcome: More layers, leads to a hard learning problem
- Pattern mapping/classification
  - Architecture: Multilayer (hidden), nonlinear processing units, geometrical interpretation
  - Learning: Generalized delta rule (backpropagation)
  - Recall: Direct
  - Limitation: Slow learning, does not guarantee convergence
  - To overcome: More complex architecture

Perceptron Network

- Perceptron classification problem
- Perceptron learning law
- Perceptron convergence theorem
- Perceptron representation problem
- Multilayer perceptron

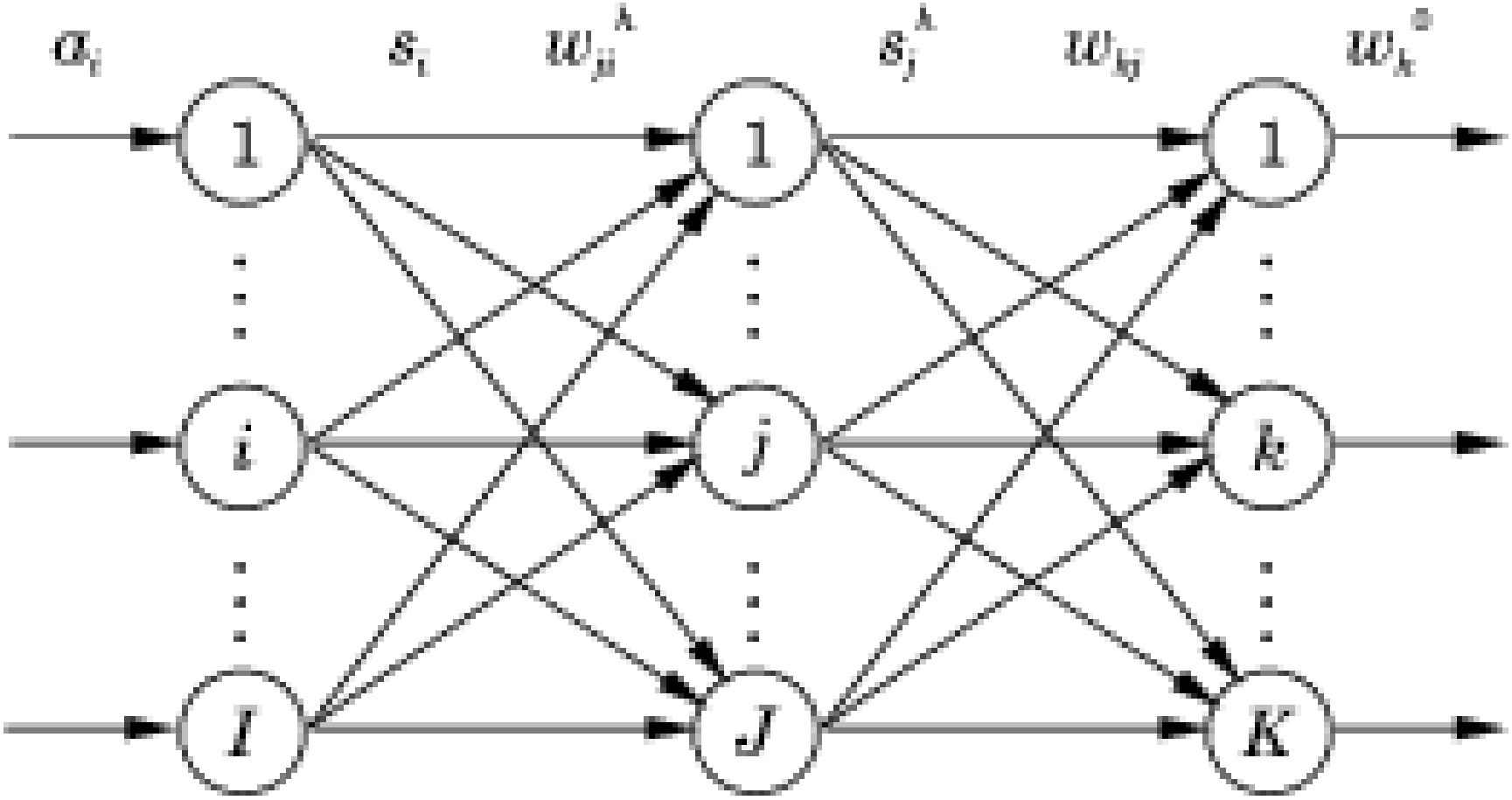
Generalized Delta Rule (Backpropagation Learning)



$$\Delta wkj = \eta s_k^o s_j^h, s_k^o = \left(b_k - f_k^o \right) \dot{f}_k^o$$

$$\Delta w_{ji}^h = \eta s_j^h a_i, s_j^h = \dot{f}_j^h \sum_{k=1}^{K} w_{kj} s_k^o$$

Multilayer FFNN



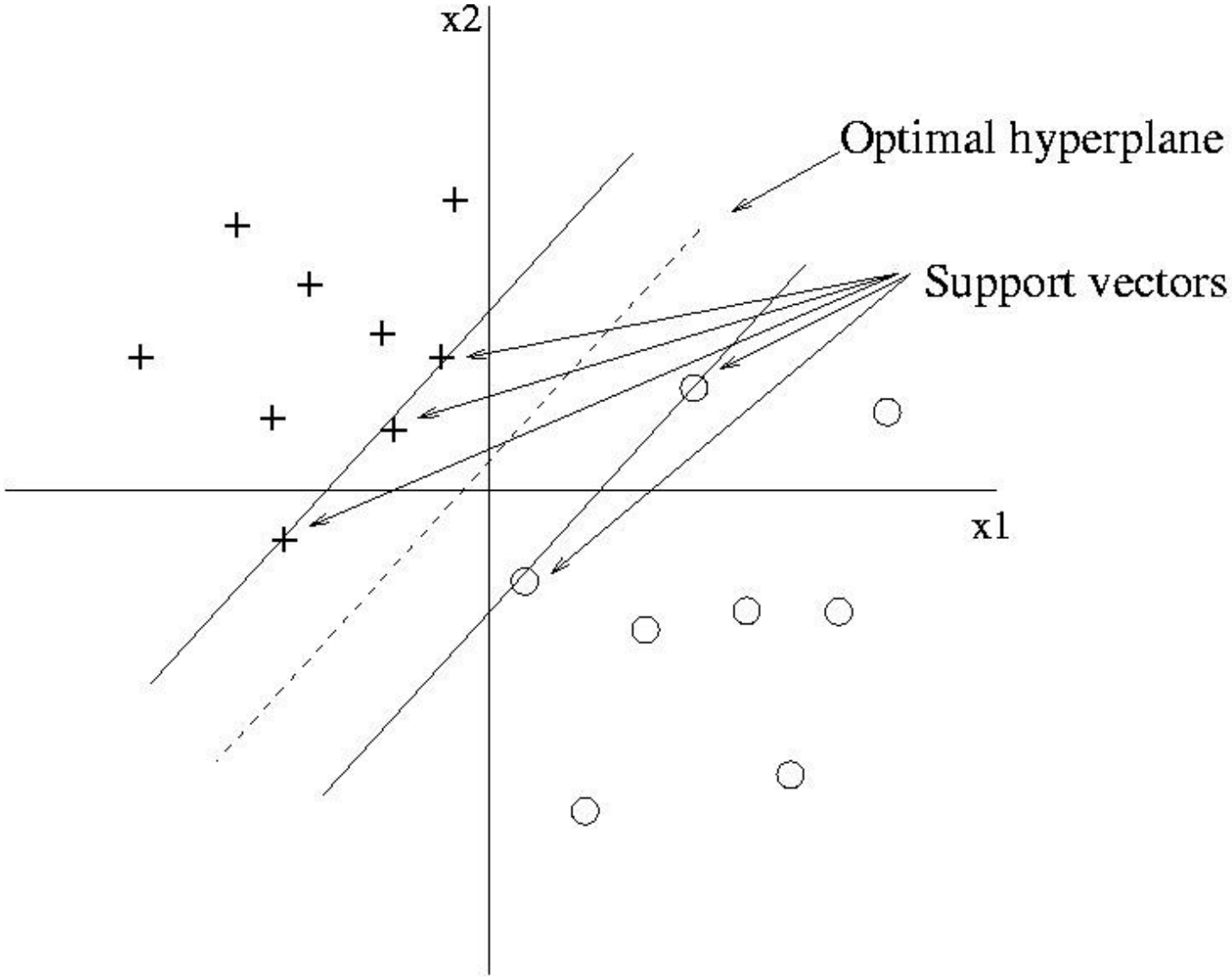
Artificial Neural Networks

- **Ability to form complex decision surfaces by using discriminatory learning algorithms**
- **MLFFNN model trained using backpropagation algorithm, provides a computationally efficient solution to the pattern classification problem**
- **To attain a better classification performance we need to build problem-domain knowledge into the design of the MLFFNN model, and tune the design parameters**
- **Difficult to arrive at the optimal parameters of the MLFFNN models for complex speech recognition tasks. Incorporation of problem-domain knowledge is also a difficult proposition**

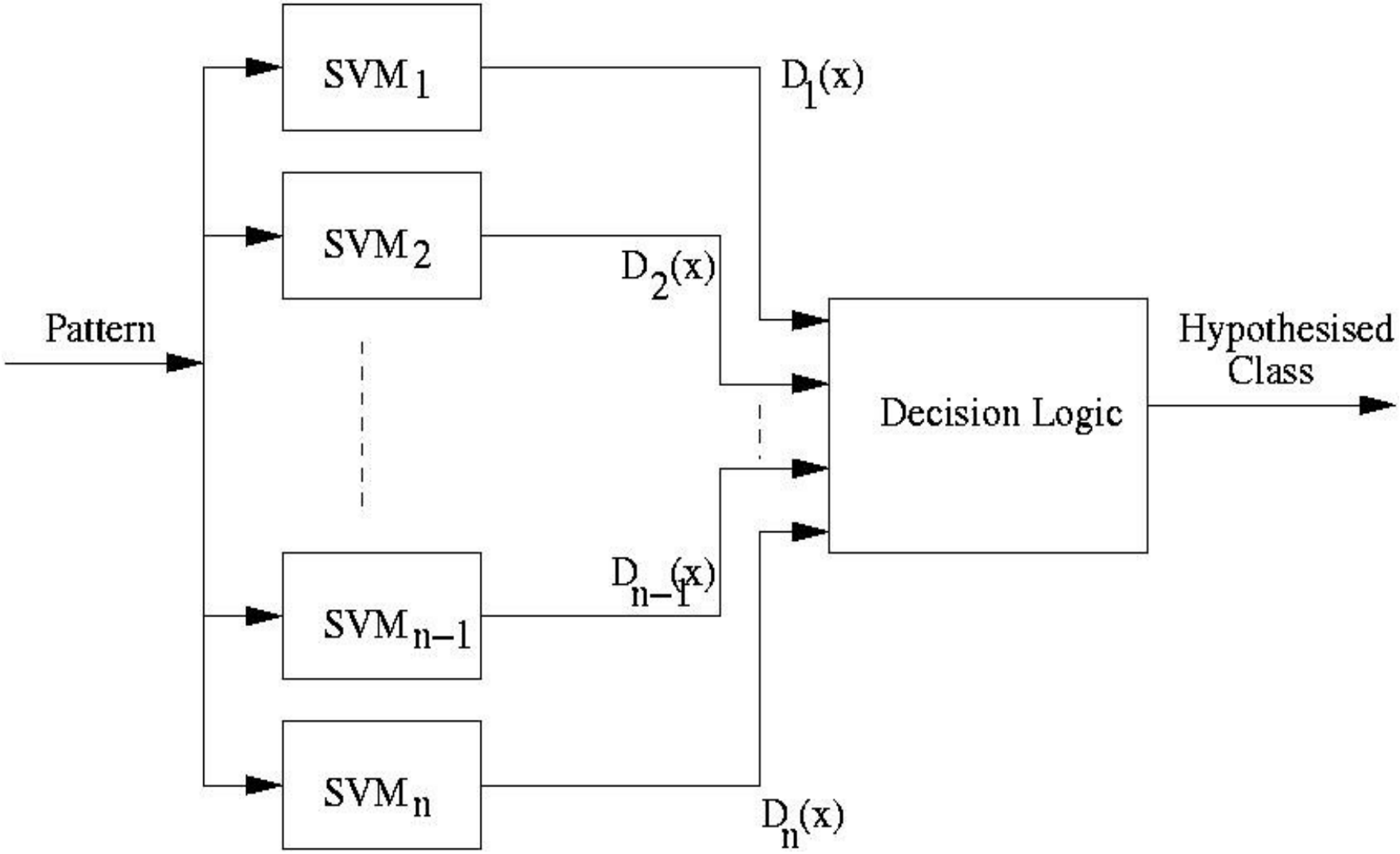
Support Vector Machines

- **Theory of a support vector machine (SVM) avoids the need for heuristics often used in the design of conventional MLFFNN models**
- **Possible to achieve better performance with no problem-domain knowledge built into the design of a machine**
- **Capable of discriminating confusable classes**

Concept of Support Vector Machine



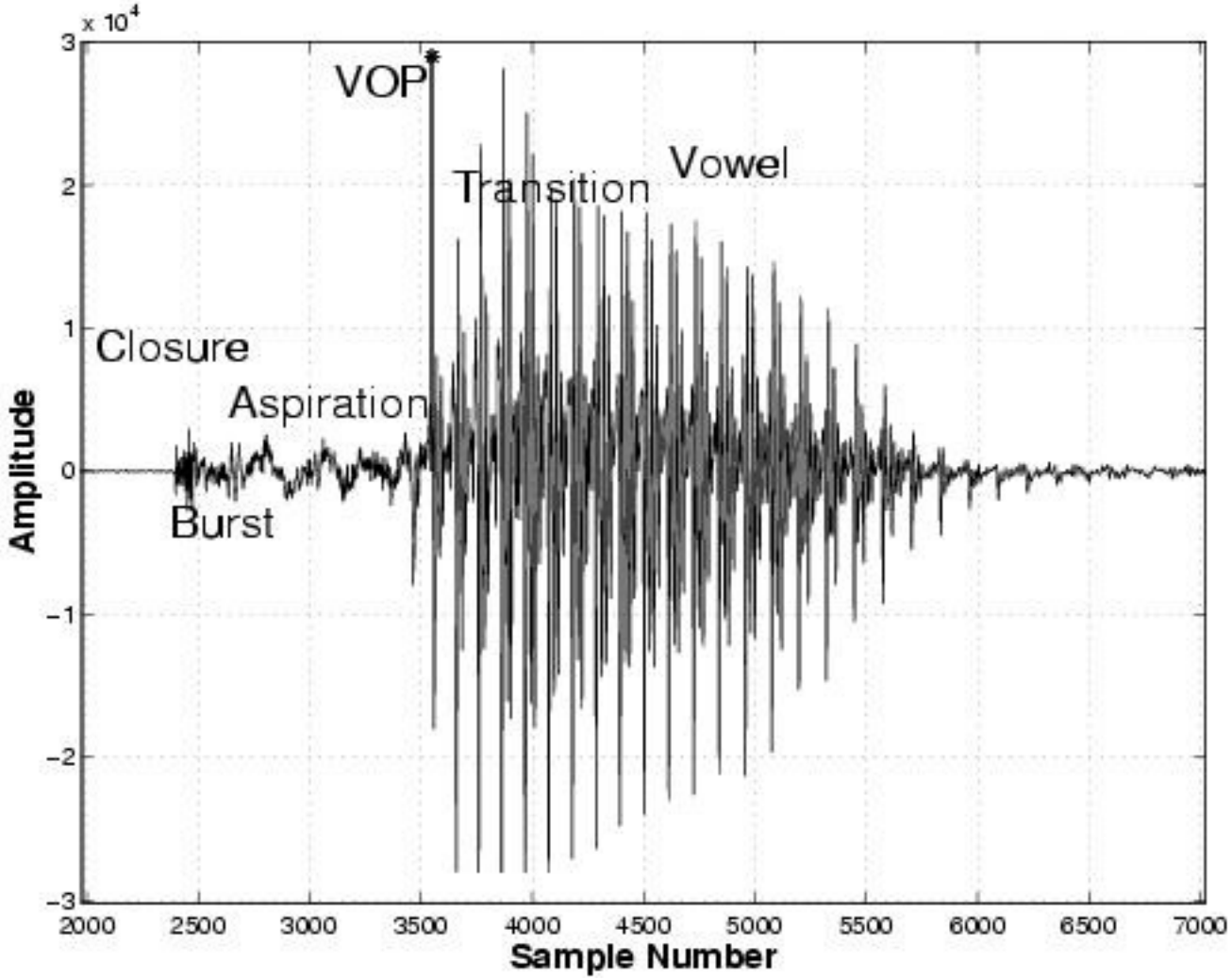
SVMs for Handling Multiple Classes



Issues in Recognition of CV Units using SVMs

- **SVM models are suitable for classification of fixed dimensional patterns**
- **Classification using SVM is computationally intensive**

Significant Events in a CV Unit

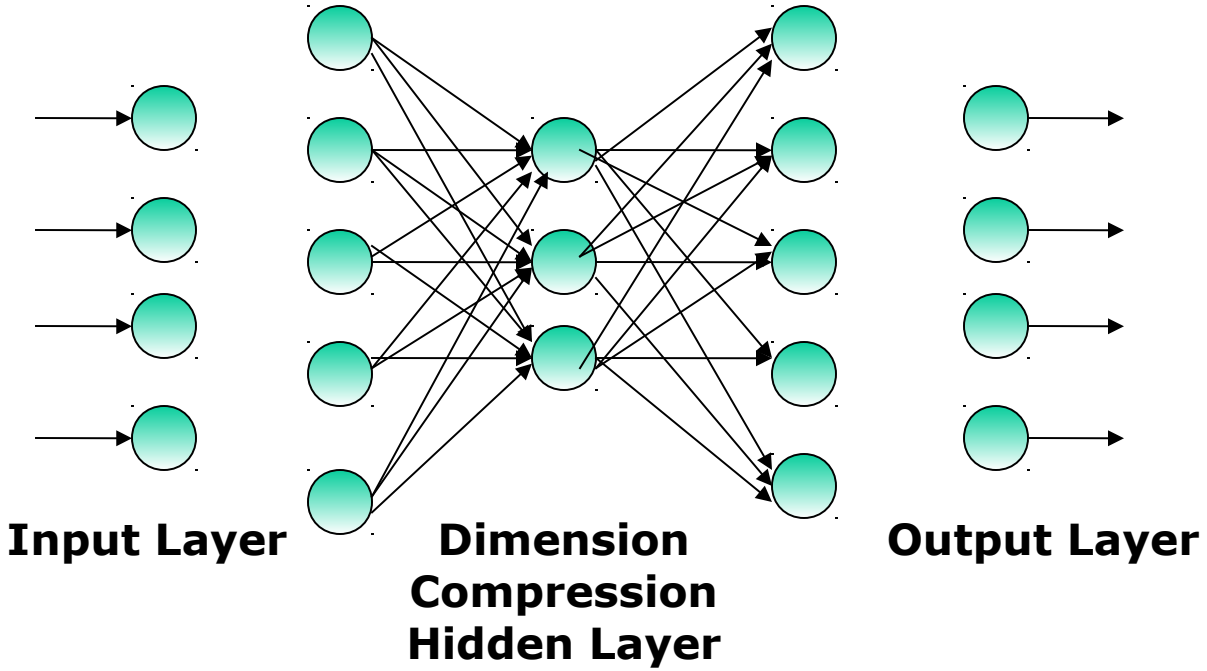


Autoassociation Neural Network (AANN)

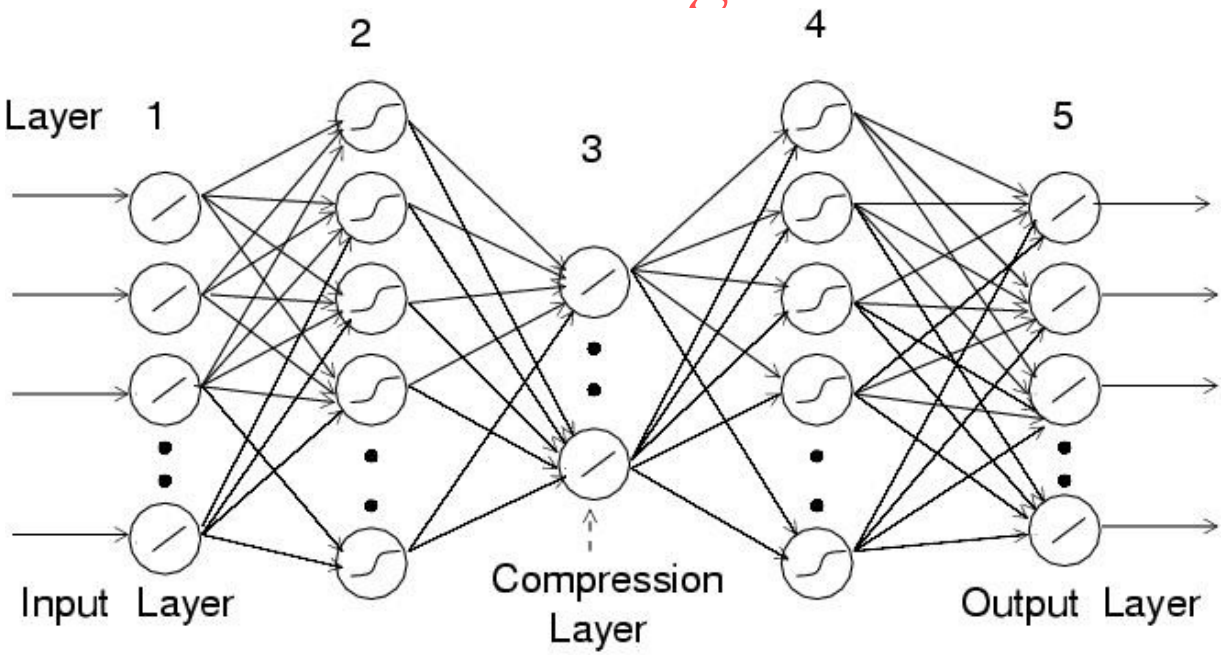
- Architecture
- Nonlinear PCA
- Feature extraction
- Distribution capturing ability

Autoassociation Neural Network (AANN)

- Architecture



VOP Detection using AANN Models



- **Autoassociative neural network (AANN) models are used for capturing the distribution of data without making any assumptions about the underlying distribution**
- **One AANN is trained to capture the distribution of the consonant region of a CV unit**
- **Another AANN is trained to capture the distribution of the vowel region of a CV unit**

System for Detection of VOPs using AANNs

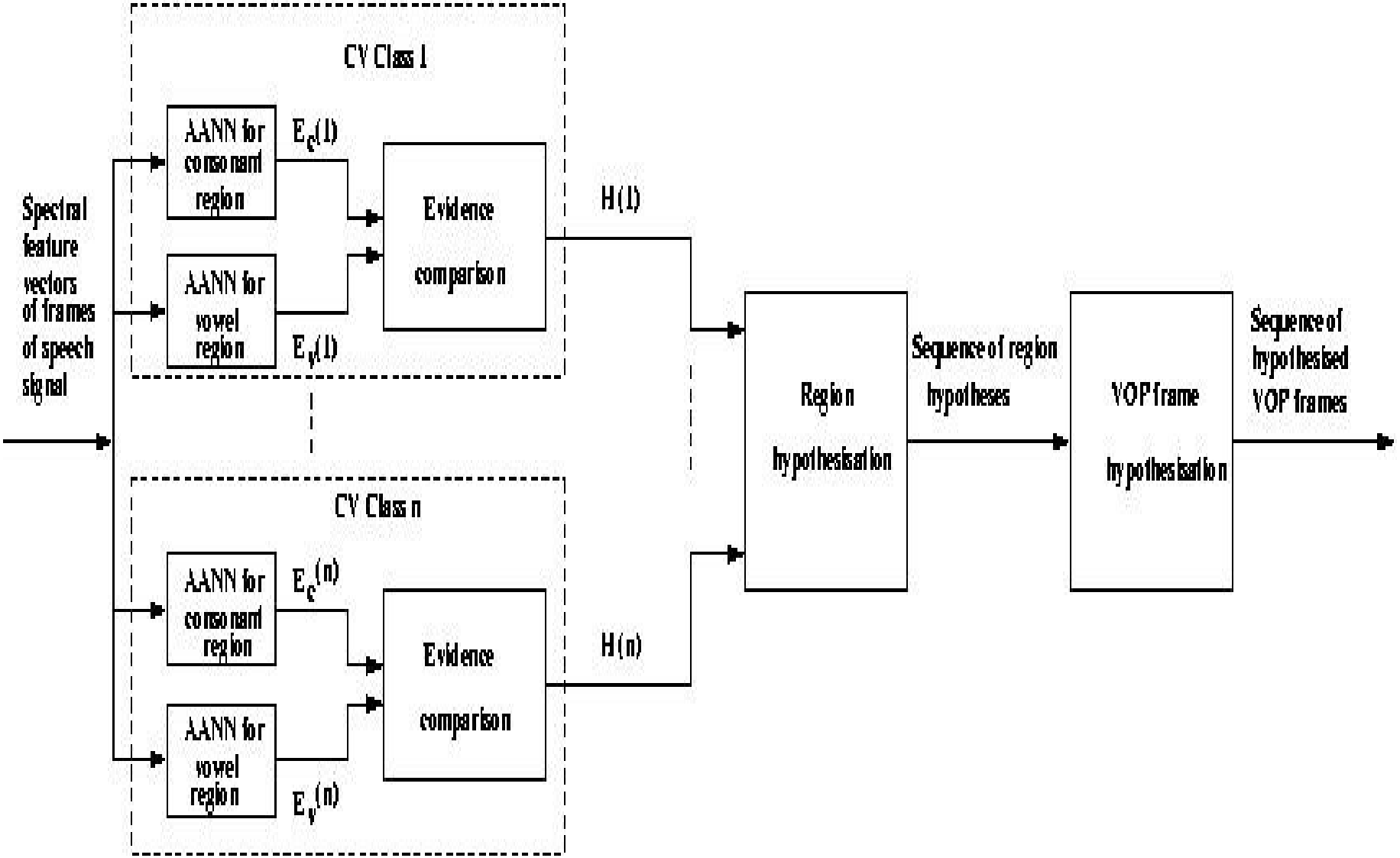
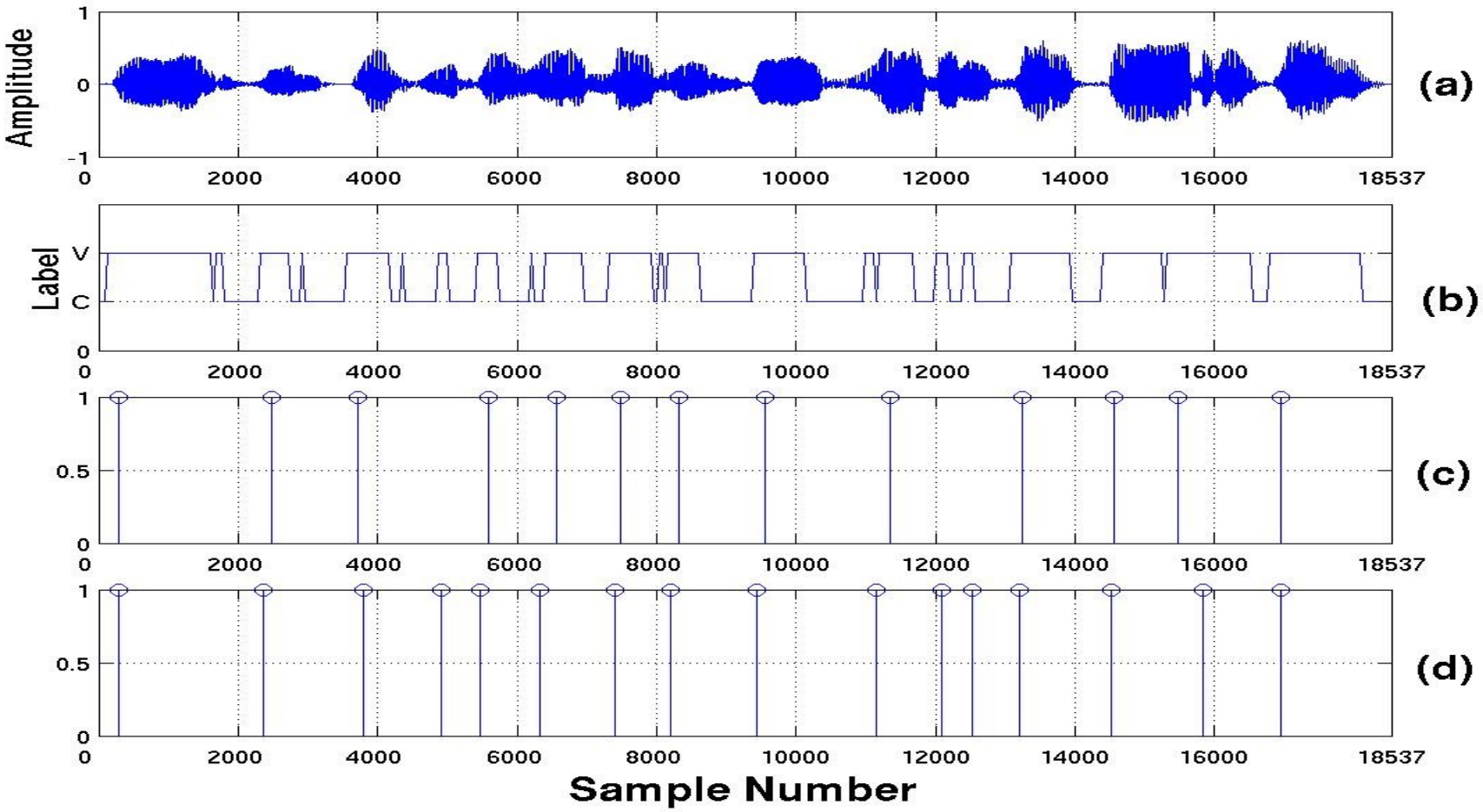


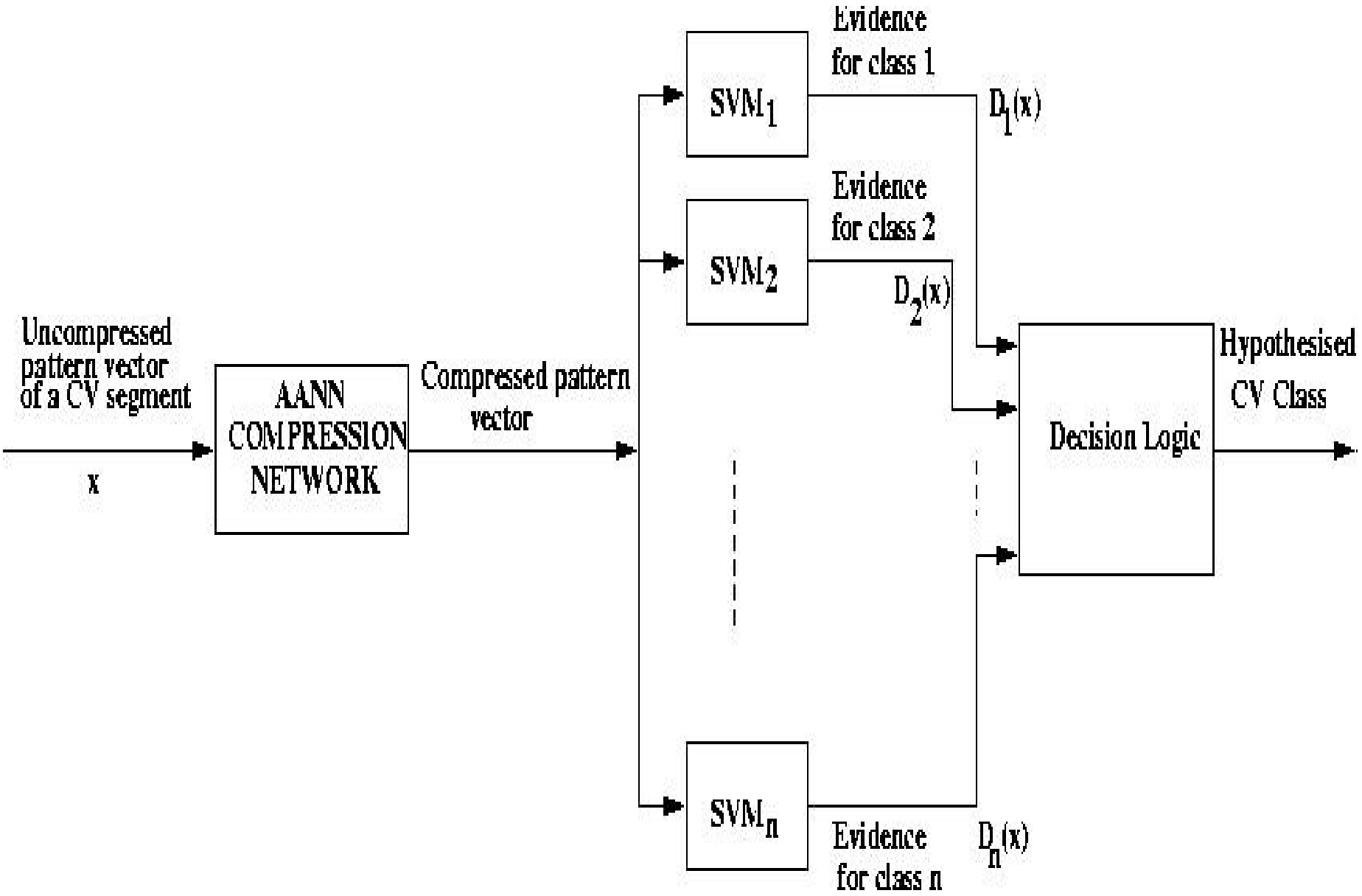
Illustration of Detection of VOPs



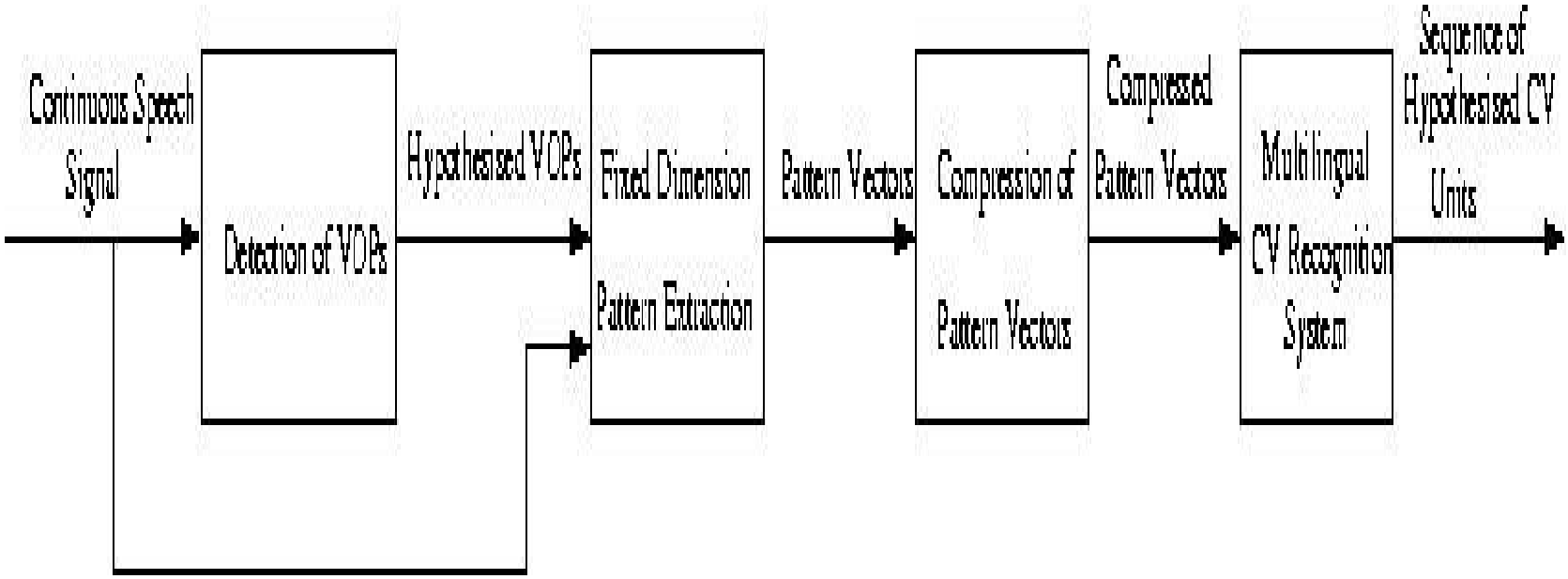
(a) Waveform, (b) Hypothesised region labels for each frame, (c)Hypothesised VOPs, and (d) Manually marked (actual) VOPs for the Tamil language sentence /kArgil pahudiyilirundu UDuruvalkAraRhaL/

Broadcast News Corpus of Indian Languages

Classification of CV Segments using SVMs

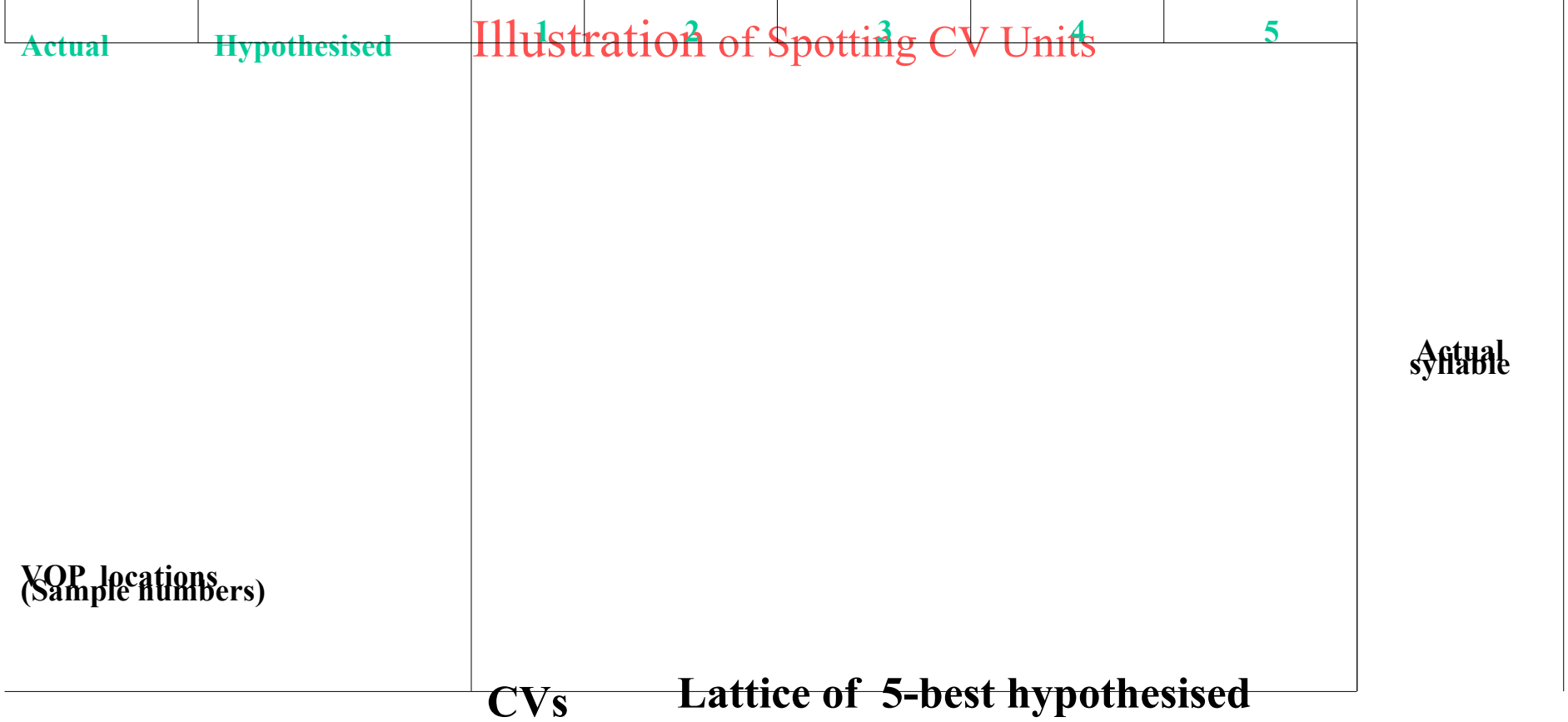


System for Spotting CV Units



- **The system gives a 5-best performance of about 74.63% for spotting CV units in 300 test sentences containing 3,924 syllable-like units**

Illustration of Spotting CV Units



References

- B.Yegnanarayana, “Artificial Neural Networks”, Prentice-Hall of India, New Delhi, 1999.
- L. R. Rabiner and B. -H. Juang “Fundamentals of Speech Recognition”, Englewood Cliffs, New Jersey: PTR Prentice Hall, 1993.

Thank You