

ssalehi@aut.ac.ir

%

*

[]:

ASR

[]

[]

)

(

[]

)

(

[]

[]

[]

[]

¹ Automatic Speech Recognition
⁵ Speaker Dependent/Independent
⁹ Mismatch
¹³ Variations
¹⁷ Redundancy
²¹ Alternative Pronunciation

² Isolated Word
⁶ Lippmann
¹⁰ Clean/Controlled
¹⁴ Background Noise
¹⁸ Multicondition Training

³ Continuous Speech
⁷ Task
¹¹ Noise
¹⁵ Reverberation
¹⁹ Speech Enhancement

⁴ Spontaneous
⁸ Perplexity
¹² Robust
¹⁶ Inter-Speaker
²⁰ Parameter

ASR

.[]

ASR

)

(

[]

OO II HH_2 HH_1
 i X_i

$$\Delta x_i = -\gamma \frac{\partial E}{\partial x_i} \quad ()$$

$$\frac{\partial E}{\partial x_i} = \frac{\partial E}{\partial z_{1m}} \frac{\partial z_{1m}}{\partial net_{z_1}(m)} \frac{\partial net_{z_1}(m)}{\partial x_i} = \dot{z}_{1m} W_{im}^I \quad ()$$

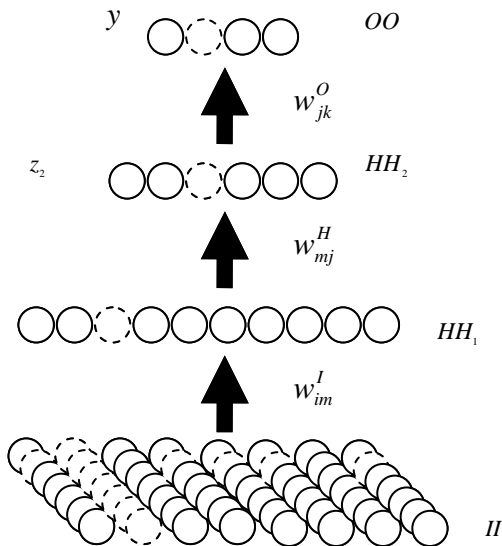
SSE

$$\dot{y}_k = (\hat{y}_k - d_{y_k}) f'(net_{y_k}(k)) \quad 1 \leq k \leq OO \quad ()$$

$$\dot{z}_{3j} = \left(\sum_{k=1}^{OO} w_{jk}^O \dot{y}_k \right) f'(net_{z_3}(j)) \quad 1 \leq j \leq HH_3 \quad ()$$

$$\dot{z}_{2m} = \left(\sum_{j=1}^{HH_3} w_{mj}^{H_2} \dot{z}_{3j} \right) f'(net_{z_2}(m)) \quad 1 \leq m \leq HH_2 \quad ()$$

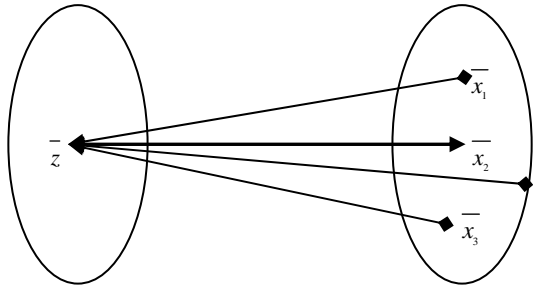
$$\dot{z}_{1l} = \left(\sum_{m=1}^{HH_2} w_{lm}^{H_1} \dot{z}_{2m} \right) f'(net_{z_1}(l)) \quad 1 \leq l \leq HH_1 \quad ()$$



[]

()

[]



$$X_i = X_i + \sum_t \Delta_t x_i \quad 1 \leq i \leq II \quad ()$$

$$X_i = X_i + \sum_N \sum_t \Delta_t x_i \quad 1 \leq i \leq II \quad (\wedge)$$

[]

[]

%

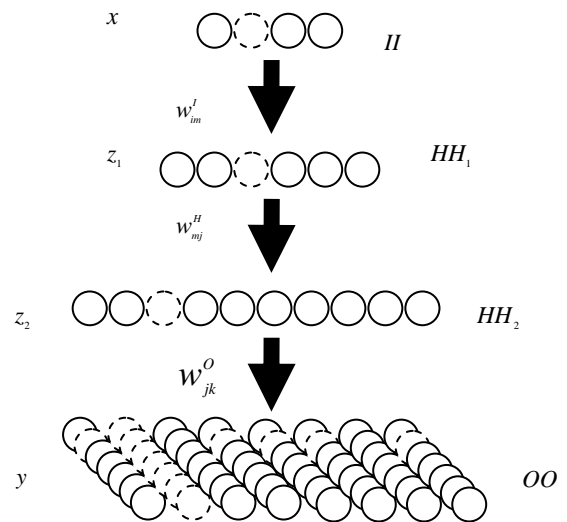
%

[]

) kHz

Hz

(



$$\bar{k}_1 = \frac{1}{M} \sum_{i=1}^M x_i \quad ()$$

$$\bar{q}_i = x_i - \bar{k}_1 \quad ()$$

$$\bar{k}_2 = \frac{1}{M} \sum_{i=1}^M q_i^2 \quad ()$$

(MFCC)

$$x_i^{normalized} = \left\{ \frac{q_{2n}}{\sqrt{k_{2n}}}, n = 1, 2, \dots, k, \dots, N \right\} \quad ()$$

M
N

)

(

LFBE

MFCC

LFBE

(N)

$$\Delta^i \{u_t\} = \Delta^{i-1} \{u_{t+1}\} - \Delta^{i-1} \{u_{t-1}\}, \Delta^0 \{u_t\} = u_t \quad ()$$

MFCC

MFCC

LFBE

MFCC

()

L_i

() ()

*

MFCC

C_i LFBE

$$L_i = \log(fb_{bank}_i) \quad i = 2, \dots, 14 \quad ()$$

$$C_i = \sum_{j=2}^{14} A_j \log(fb_{bank}_j) \quad i = 1, \dots, 12 \quad ()$$

$$C_{13} = \log(Energy) \quad ()$$

³⁴ Bark Scale
³⁸ Normalization

³⁵ Mel Scale
³⁹ Acoustic Model

³⁶ Mel Frequency Cepstral Coefficients
⁴⁰ MultiLayer Perceptron

³⁷ Logarithmic Filterbank Energy

()

| | | | | |
|-----|-----|-----|-----|-----|
| | | | | |
| % / | % / | % / | % / | % / |

()

φ

[]

[]

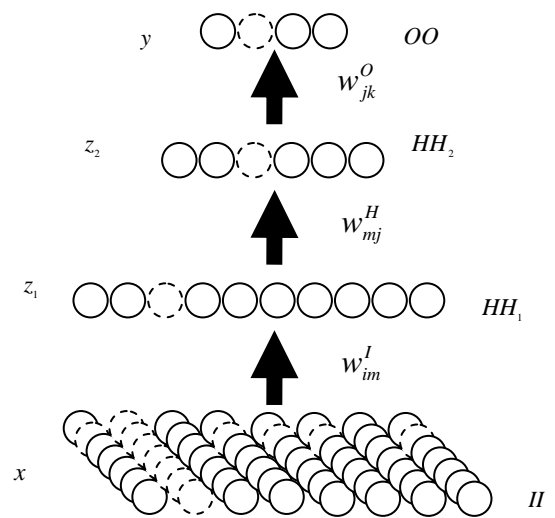
[]

% /

[]

[]

[]



⁴¹ Nguyen-Widrow
⁴⁵ Insertion Ratio
⁴⁹ Basin of Attraction

⁴² Correction Ratio
⁴⁶ Substitution Ratio

⁴³ Accuracy Ratio
⁴⁷ Mismatch

⁴⁴ Deletion Ratio
⁴⁸ Attractor

MLP

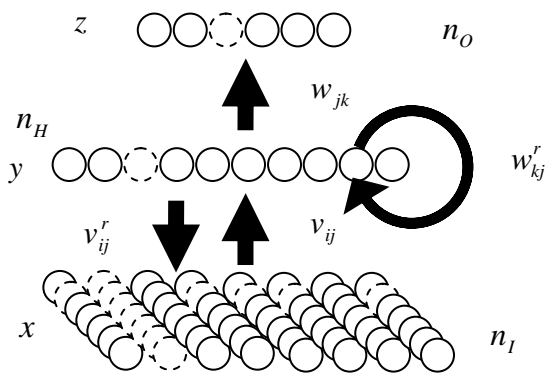
[]

()

)

(

*



| | |
|--------------|-------------|
| | |
| r=dah/'\$ | h/r=dah/'\$ |
| r=da/'\$ | |
| ah/r=dah/'\$ | |
| h/rdah/'\$ | |
| h/r=deh/'\$ | |
| h/r=da/'\$ | |
| . := :/ :' | :\$ |

$$\hat{z}(k, n) = \frac{1}{n} \sum_{m=1}^n z(k, m) \quad k = 1, \dots, n_o$$

$$\dot{z}(k, n) = (\hat{z}(k, n) - d_z(k)) f'(z(k, n)) \quad k = 1, \dots, n_o$$

$$w_{jk} = w_{jk} + \eta \dot{z}(k, n) y(j, n) \quad k = 1, \dots, n_o, j = 1, \dots, n_H$$

$$\dot{y}(j, n) = f'(Y(j, n)) \left(\sum_{k=1}^{n_o} \dot{x}(k, n+1) v'_{jk} + \right.$$

$$\left. \sum_{k=1}^{n_o} \dot{z}(k, n) w_{jk} + \sum_{i=1}^{n_H} \dot{y}(i, n+1) w'_{ij} \right) \quad j = 1, \dots, n_H \quad ()$$

$$w'_{ij} = w'_{ij} + \eta \dot{y}(j, n) y(i, n-1) \quad i = 1, \dots, n_H, j = 1, \dots, n_H$$

$$v_{ij} = v_{ij} + \eta \dot{y}(j, n) x(i, n-1) \quad i = 1, \dots, n_I, j = 1, \dots, n_H$$

$$\dot{x}(i, n) = (1 - \gamma) \dot{x}(i, n+1) +$$

$$\gamma \left(\sum_{j=1}^{n_H} \dot{y}(j, n) v_{ij} \right) f'(x(i, n+1)) \quad i = 1, \dots, n_I$$

$$v'_{ij} = v'_{ij} + \eta \dot{x}(k, n) y(j, n) \quad k = 1, \dots, n_I, j = 1, \dots, n_H$$

$$n = N_0 - 1, \dots, 1$$

()

()

$$\hat{x}(k, N_0) = \frac{1}{N_0} \sum_{m=1}^{N_0} x(k, m) \quad k = 1, \dots, n_I$$

$$\dot{x}(k, N_0) = (\hat{x}(k, N_0) - d_x(k)) f'(X(k, N_0)) \quad k = 1, \dots, n_I \quad ()$$

)

n () (

n+1

N₀

() N₀

$$y(j, n) = f \left(\sum_{i=0}^{n_I} x(i, n) v_{ij} + \sum_{k=1}^{n_H} y(k, n-1) w'_{kj} \right) \quad j = 1, \dots, n_H$$

$$z(k, n) = f \left(\sum_{j=1}^{n_H} y(j, n) w_{jk} \right) \quad k = 1, \dots, n_o \quad ()$$

$$x(i, n+1) = (1 - \gamma) x(i, n) + \gamma \left(\sum_{j=1}^{n_H} y(j, n) v'_{ij} \right) \quad i = 1, \dots, n_I$$

$$\gamma \quad n = 1, \dots, N_0, 0 \leq \gamma \leq 1$$

)

(γ=0.7

()

| | | | | | |
|-----|-----|-----|-----|-----|--|
| | | | | | |
| % / | % / | % | % / | % | |
| % / | % | % / | % / | % / | |

| | | | | | |
|-----|-----|-----|-----|-----|--|
| | | | | | |
| % / | % / | % / | % | % / | |
| % / | % / | % / | % / | % / | |
| % / | % / | % / | % / | % / | |

HMM

)

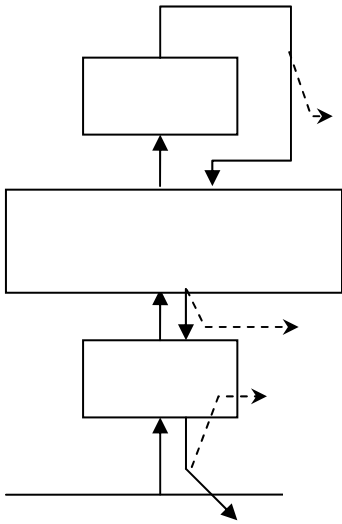
(
)

.[]

(

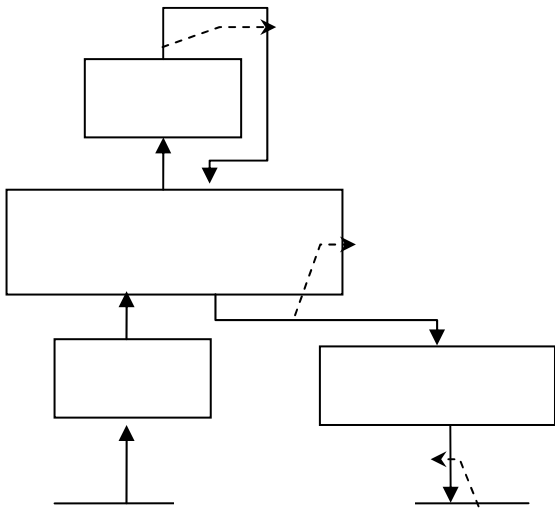
()

.[]



)

(



()

()

()

| | | | | | |
|-----|-----|-----|-----|-----|--|
| | | | | | |
| % / | % / | % / | % / | % / | |
| % / | % / | % / | % / | % / | |

()

MLP

*

] ASR

()

.[

.[]

:

.[]

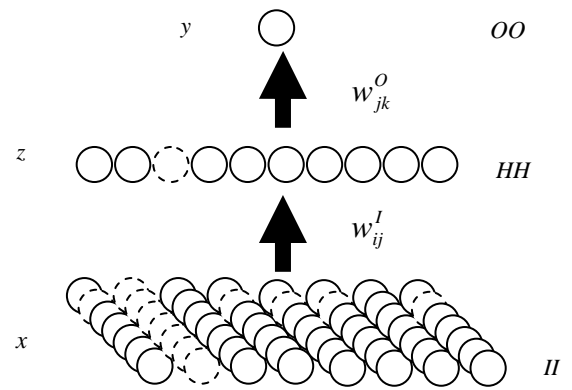
" / "

" "

" / "

" "

.[]



()

| | | | | | |
|-----|---|-----|-----|-----|--|
| | | | | | |
| % / | % | % / | % / | % / | |
| % / | % | % / | % / | % / | |

[]

()

- [13] Bijankhan M., Seikhzadeghan J., Roohani M.R., Samareh Y., Lucas K. Tebyani M., FARSDAT-the speech Database of Farsi Spoken Language; Proc. of SST94 1994;826-831.

[]

LHCB MFCC

- [15] Nguyen D., Widrow B., Neural Networks for Selflearning Control systems; IEEE Control Systems Magazine 1990; 10:18-23.
- [16] Koerner E., Gewaltig M.O., Koerner U., Richter A., Rodemann U., A Model of Computation in Neocortical Architecture; Neural Networks Elsevier Science 1999;12: 989-1005.
- [17] Koerner E., Tsujino H., Masutani T., A cortical type modular neural network for hypothetical reasoning; Neural Networks , Elsevier Science 1997; 10: 791-814.
- [18] Koerner E., Matsumoto G., Cortical architecture and self-referential control for brain-like computation; Engineering in Medicine and biology Magazine, IEEE 2002; 21:121-133
- [19] Wan E, Nelson AT; Networks for Speech Enhancement. In: Handbook of Neural Networks for Speech Processing, 1998:541-541.
- [20] Ghosen J., Bengio Y., Bias Learning, Knowledge sharing; IEEE Trans. On Neural Networks 2003; 14:84-108.
- [21] Mesulam M.M., From Sensation to Cognition; Brain, Oxford Univ. Press 1998, 121:1013- 1052.

[]

- [23] Saul L.K., Jordan M.I.; Attractor dynamics in feed forward neural networks; Neural Computation, Massachusetts Institute of Technology 2000; 12: 1313-1335.
- [24] Trappenberg, T., Continuous attractor neural networks. In L. N. de Castro & F. J. V. Zuben (Eds.), Recent developments in biologically inspired computing. Hershey, PA: Idee Group; 2003.
- [25] Wu Y., Pados D.A.; A feedforward bidirectional associative memory; IEEE Trans. On Neural Networks 2000;11:42.

- [1] Bimbot F., Chollet G., Paoloni A., Assessment methodology for speaker identification and verification systems: An overview of SAM-A Esprit project 6819 - Task 2500. *ESCA Workshop on Automatic Speaker Recognition Identification and Verification* 1994; 75-82.
- [2] Lippmann R.P., Speech recognition by machines and humans; *Speech Communication* 1997; 22:1-15.
- [3] Gong, Y.; Speech recognition in noise environments: a survey; *Speech Communication* 1995; 16:261-291.
- [4] Miller G.A., Licklider J.C.R., The intelligibility of interrupted speech; *Journal of the Acoustic Society of America* 1950; 22:167-173.
- [5] Fletcher H., Speech and Hearing in Communication. *Journal of the Acoustic Society of America* 1953; 28:164-172.
- [6] Furui S., Recent advances in robust speech recognition; *Speech Communication* 1997; 22:27-39.
- [7] Lockwood P., Boudy J., Experiments with non-linear Spectral Subtractor (NSS), Hidden Markov Models and the projection, for robust speech recognition in cars; *Speech Communication* 1992; 11: 215-228.
- [8] Diamantaras K.I., Neural networks and principal component analysis, In: Handbook of neural network signal processing; CRC Press, 2002.
- [9] Cooke M., Morris A., Green P., Recognizing Occluded Speech; *ESCA Tutorial and Workshop on the Auditory Basis of Speech Perception* 1996, Keele University, 15-19.
- [10] Jensen C.A., Reed R.D., Marks R.J., Inversion of Neural Networks: Algorithms and Applications; *IEEE Neural Networks* 1999; 87:1536-1549.
- [11] Williams R.J., Inverting a Connectionist Network Mapping by Backpropagation of Error; *Proc. 8th Annu. Conf. Cognitive Science*.