

**YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

85071

**GRAFİK PROGRAMLAMA KULLANARAK
KEPSTRUM ANALİZİ VE YAPAY SİNİR AĞI İLE
KONUŞMACI TANIMA**

Elek. Yük. Müh. Orhan ÖZHAN

**F.B.E. Elektrik Mühendisliği Anabilim Dalında
Hazırlanan**

DOKTORA TEZİ

**T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ**

85071

Tez Savunma Tarihi : 30 Nisan 1999

Tez Danışmanı : Prof. Dr. Halit PASTACI (YTÜ)

Jüri Üyeleri : Doç. Dr. Fikret GÜRGEN (BÜ)

Doç. Dr. Tamer ÖLMEZ (İTÜ)

Prof. Dr. Halit PASTACI
Doç. Dr. Fikret GÜRGEN
Doç. Dr. Tamer ÖLMEZ

İSTANBUL, 1999

İÇİNDEKİLER

SİMGE LİSTESİ	iv
KISALTMA LİSTESİ ve YABANCI SÖZCÜKLER	vi
ŞEKİL LİSTESİ	viii
TABLO LİSTESİ	x
ÖNSÖZ	xi
ÖZET	xii
ABSTRACT	xiii
1 GİRİŞ	1
1.1 Tipografi	1
1.2 Önceki Çalışmalar	1
1.3 Motivasyon ve Amaç	6
1.4 Yöntem	7
1.4.1 Konuşma verileri	8
1.4.2 Öznitelik çıkarma	8
1.4.3 SOM	11
1.5 Tezde Getirilen Yenilikler	13
1.5.1 Ayırık SOM Yapay Sinir Ağları	13
1.5.2 İki geçişli öznitelik çıkarma	14
1.5.3 Diferansiyel temel periyod dedektörü	14
1.5.4 Glotal uyarım faktörünün bulunması	14
1.6 Tezin Bölümleri	14
1.6.1 Giriş	14
1.6.2 Temel Bilgiler	15
1.6.3 Öznitelik Çıkarma	15
1.6.4 Eğitim ve Tanıma	15
1.6.5 Tartışma ve Öneriler	16
2 TEMEL BİLGİLER	17
2.1 Konuşma İşareti	17
2.2 Ses Üretimi	19
2.3 Doğrusal Konuşma Modeli ve Homomorfik Analiz	27
2.4 SOM Yapay Sinir Ağı	32
3 ÖZNİTELİK ÇIKARMA	41
3.1 Giriş	41
3.2 Çerçeveleme ve Pencereleme	42
3.3 Öznitelik Uzayı	50
3.4 Analiz	52
3.4.1 Yarım perde seslendirme problemi	53

3.4.2	Birinci geiř.....	56
3.4.2.1	ereveleme ve pencereleme.....	61
3.4.2.2	Kepstrum analizi.....	62
3.4.2.3	Temel peryod dedeksiyonu.....	64
3.4.2.4	Öznitelik vektörü oluřturulması ve dosyalama.....	67
3.4.3	İkinci geiř.....	68
4	EĐİTİM ve TANIMA.....	72
4.1	Giriř.....	72
4.2	Eđitim.....	73
4.3	Tanıma Algoritması.....	76
4.4	Glotal Uyarım Katsayısının Belirlenmesi.....	79
4.5	İstatistik Zenginleřtirme.....	83
4.6	LabVIEW ile Tanıma.....	85
5	SONULARIN TARTIřILMASI ve ÖNERİLER.....	93
5.1	YSA Mimarisi.....	93
5.2	SOM Performansı.....	94
5.3	Glotal Uyarım Miktarı.....	96
5.4	Yarım Perde Seslendirmesi.....	96
5.5	LabVIEW.....	97
5.6	İleriye Dönük Öneriler.....	98
6	KAYNAKLAR.....	100
EK1	KONUřMACILAR.....	105
EK2	SOM_PAK YAZILIMININ KULLANIMI.....	109
EK3	TANIMA SONULARI.....	113
	ÖZGEMİř.....	120

SİMGE LİSTESİ

A_i	Akustik borularda i -nci borunun kesit alanı
$a(k,S)$	Bir S konuşması ve k glotal uyarım faktörü için performans ağırlığı
$\alpha(n)$	n -nci adımda (SOM yapay sinir ağı) öğrenme hızı katsayısı
$\beta(n)$	n -nci adımda (SOM yapay sinir ağı) Gauss güncelleme fonksiyonu
$c_1, \dots, c_{19}$	Konuşma işaretinin 1.nci, ..., 19.ncu kepsrum bileşenleri
C_j	j -nci kod defteri
C_{jk}	j -nci kod defterinin k -nci kod vektörü
$d(f_i, C_{jk})$	f_i çerçevesi ile C_{jk} arasındaki uzaklık
$d(X, Y)$	X, Y vektörleri arasındaki uzaklık
Φ	Öznitelik haritası
F_0	Konuşma işaretinin temel frekansı
$F_1, F_2, \dots$	Konuşma işaretinin 1.nci, 2.nci, ... formantları
$\mathcal{F}^{-1}(\cdot)$	Ters Fourier dönüşümü operatörü
G	Tanıma güvenirlik oranı
$H(w,t)$	Artikülasyon filtresinin zamana bağımlı (basınç) transfer fonksiyonu
$H_1(w,t)$	Sadece ağız boşluğu açıkken artikülasyon filtresinin zamana bağımlı (basınç) transfer fonksiyonu
$H_2(w,t)$	Sadece burun boşluğunun zamana bağımlı (basınç) transfer fonksiyonu
$\angle H(n\omega_0)$	n -nci glotal uyarım harmoniğinde artikülasyon filtresinin gösterdiği faz farkı
$i(\mathbf{x})$	SOM haritasında $\mathbf{x}$ giriş vektörü için kazanan nöron indeksi
k	Glotal uyarım faktörü
K_i	1) Akustik borularda " i "nci bölümden " $i-1$ "nci bölüme akustik yansıma katsayısı; 2) " i "nci konuşmacı
$\Lambda_i(\mathbf{x})$	Kazanan nöron $i(\mathbf{x})$ etrafında ağırlık yenileme bölgesi
μ_{T_0}	Temel periyodun (T_0) ortalama değeri
$n[A S]$	Verilen bir S konuşması için A konuşmacısına ait tanıma puanı
$P(A S)$	Verilen bir S konuşmasını A konuşmacısının söyleme olasılığı
p_i	Artikülasyon filtresi transfer fonksiyonunun " i "nci kutbu
$p(k)$	Glotal uyarım performans fonksiyonu
Q	Rezonans filtresinin kalite faktörü: rezonans frekansı $\div$ band genişliği

R	$\Lambda_{i(x)}$ güncelleme bölgesinin yarıçapı
σ_{T_0}	Temel peryodun (T_0) standard sapması
$s(n)$	Konuşma işareti
T_0	Konuşma işaretinin temel peryodu
$w(n)$	Pencere fonksiyonu
$\mathbf{w}_j(n)$	n -anında j -nöronunun (sinaptik) ağırlığı
z^{-1}	Ayrık-zaman birim gecikme operatörü



KISALTMA LİSTESİ ve YABANCI SÖZCÜKLER

ANN	Artificial Neural Network (Yapay Sinir Ağı)
DARPA-ISTO	Defense Advanced Research Agency – Information Science and Technology
DTW	Dynamic Time Warping (Dinamik Zaman Deformasyonu)
EDVAC	Electronic Discrete Variable Computer. Von Neumann tarafından McCulloch-Pitts nöron modeline dayanarak yapılan bilgisayarın kod adı
ENIAC	Electronic Numeric Integrator And Computer. von Neumann tarafından yapılan ilk bilgisayarın kod adı
GMM	Gaussian Mixture Model (Gauss Karışım Modeli)
GUI	Graphical User Interface
HMM	Hidden Markov Model (Gizli Markov Modeli)
KNN	K-Nearest Neighbor (sınıflama algoritması)
LPC	Linear Predictive Coding (Doğrusal Öngörü Kodlaması)
LVQ	Learning Vector Quantizer
MIT	Massachusetts Institute of Technology
MLP	Multi-Layer Perceptron (Çok Katmanlı Algılayıcı)
MSS	Merkezi Sinir Sistemi
NIST	National Institute of Standards and Technology
NPS	Normal Perde Seslendirmesi
PCM	Pulse Code Modulation
RBF	Radial Basis Function
SOM	Self-Organizing Map (Kendi kendini düzenleyen harita)
SOFM	Self-Organizing Feature Map (Kendi kendini düzenleyen öznetelik haritası)
SubVI	Sub virtual instrument (alt sanal cihaz)
YPS	Yarım Perde Seslendirmesi
YSA	Yapay Sinir Ağı
TIMIT	Texas Instruments + Massachusetts Institute of Technology
VI	Virtual Instrument (sanal cihaz)
VQ	Vector Quantization (Vektör Nicelemesi)
XOR	Exclusive OR lojik işlemi

Alveoler yamaç	Kesici dişlerle sert damak arasındaki bölge
Antialias (filtre)	Nyquist frekansından yüksek frekansları durduran alçak geçiren filtre
Articulation	Konuşma üretimi, artikülasyon
Arytenoid	Larinksin arka tarafında bulunan kıkırdak
Big-Endian	Bir bayttan daha büyük sayısal veriyi yazıp okurken önce yüksek mertebeli bayttan başlayan sayısal format; MOTOROLA formatı
Cricoid	Larinksin alt tarafında bulunan kıkırdak
Larynx	Gırtlak. Nefes borusunun üstünde ses tellerini içinde barındıran kıkırdaksı yapı, larinks
Little-Endian	Bir bayttan daha büyük sayısal veriyi yazıp okurken önce düşük mertebeli bayttan başlayan sayısal format; INTEL formatı
Notch (filtre)	Dar bir frekans bandını durduran filtre; çentik filtre
Palate	(Sert) damak
Perceptron	Algılayıcı. En temel yapay sinir ağı elemanı
Pitch Halving	Yarım Perde Seslendirmesi
Shift register	LabVIEW'da for ve while döngülerinde bir sayısal değeri bir ötelemeden diğer ötelemeye aktarmak için kullanılan kayar kayıt
Subroutine	Özel bir işlem için programlanmış, üst düzey program kodları tarafından çağrılan alt program
Sinaps	Bir sinir hücresinin aksonuyla diğerinin dendriti arasındaki bağlantı, <i>synapse</i>
Thresholding	Eşikleme işlemi. Bir süreçte kriter olarak bir "eşik" değeri belirleyip bundan büyük ve küçük değerleri kabul ve reddetmek
Thyroid	Larinksin üst tarafında bulunan kıkırdak
Trachea	Nefes borusu
Thyroid kıkırdağı	Larinksin üst tarafında bulunan kıkırdak
Unvoiced speech	Ses tellerinin titreşimi olmadan üretilen (konuşma)
Velum	Yumuşak damak
Virtual Instrument	Sanal cihaz
Vocal chords	Ses telleri
Vocal folds	Ses kıvrımları. Yapılarından ötürü ses tellerine verilen diğer bir isim
Voiced speech	Ses tellerinin titreşimi ile üretilen (konuşma)

ŞEKİL LİSTESİ

Şekil 1.1	Bir konuşmacı tanıma sisteminin prensip şeması	2
Şekil 1.2	TIMIT konuşmalarından <i>saI</i> cümlesinin spektrogram ve kepstrogramı	3
Şekil 1.3	İki kademeli öznitelik çıkarıcısı.....	9
Şekil 1.4	Temel periyod dedektörlerinin karşılaştırması	10
Şekil 1.5	Bir SOM kullanarak sınıflama.....	11
Şekil 1.6	Geliştirilen konuşmacı tanıma sistemi	12
Şekil 2.1	Konuşma işareti zaman grafiği	17
Şekil 2.2	/a/ sesinin frekans spektrumu.....	18
Şekil 2.3	Artikülasyon organları.....	20
Şekil 2.4	Konuşma sırasında konuşma organlarının konumu ve konuşma yolunun aldığı şekiller	22
Şekil 2.5	/a,e,i,o,ö,u,ü/ ses dizisinin üç ayrı şekilde söylenmesinden elde edilen spektrogram ve kepstrogramlar	23
Şekil 2.6	Kişilerdeki söyleyiş farklarının spektrograma yansımaları.....	25
Şekil 2.7	Kişilerdeki söyleyiş farklarının kepstrograma yansımaları.....	26
Şekil 2.8	Konuşma modeli	29
Şekil 2.9	Homomorfik analiz	30
Şekil 2.10	Artikülasyon filtresi modeli.....	32
Şekil 2.11	Öznitelik haritası.....	35
Şekil 2.12	İki konuşmacının temel frekanslarının SOM haritaları.....	36
Şekil 2.13	SOM yapay sinir ağı komşuluk topolojileri	37
Şekil 2.14	Öznitelik haritasında ağırlıkların güncellenmesi	39
Şekil 3.1	(a) Çerçeveleme ve pencereleme işlemleri, (b) Yaygın olarak kullanılan bazı pencereler.....	42
Şekil 3.2	Dikdörtgen, Hamming, Hanning, Blackman ve Bartlett pencerelerinin frekans genlik karakteristiği.....	44
Şekil 3.3	Çeşitli pencerelerde spektral sızıntı	46
Şekil 3.4	Çeşitli pencerelerle işaret çözünürlüğü	47
Şekil 3.5	Çerçeveleme ve pencereleme işlemlerinin LabVIEW platformunda gerçekleştirilmesi	49

Şekil 3.6	Öznitelik olarak kullanılan reel kepstrum katsayıları	51
Şekil 3.7	Glotal uyarım	53
Şekil 3.8	Aynı çerçeve içinde peşpeşe gelen iki sinüzoidal tonun güç spektrumu	54
Şekil 3.9	2Tone sanal cihazının diyagramı	55
Şekil 3.10	1nciGecis.vi panosu	57
Şekil 3.11	1nciGecis.vi diyagramı	59
Şekil 3.12	1nciGecis.vi'in bağımlı olduğu SubVI hiyerarşisi	60
Şekil 3.13	SesOku.vi diyagramı	62
Şekil 3.14	Homomorfik analiz (kepstrum analizi)	63
Şekil 3.15	Kepstrum analizinde T_0 ve $2T_0$ 'ın değişik kombinasyonları	65
Şekil 3.16	LabVIEW ile gerçekleştirilmiş yarım dalga doğrultucusu	66
Şekil 3.17	DiffPitchBul.vi diyagramı	67
Şekil 3.18	2nciGecis sanal cihazının panosu	69
Şekil 3.19	2nciGecis.vi diyagramı	70-71
Şekil 4.1	Eğitmeden sonra bir konuşmacı için oluşturulan SOM haritası	74
Şekil 4.2	Bazı SOM haritalarının kT_0 katmanları	75
Şekil 4.3	Eğitim ve tanıma işlemleri	77
Şekil 4.4	Tani.vi ile konuşmacı tanıma	79
Şekil 4.5	Performansın k faktörü ile değişimi	83
Şekil 4.6	Tani.vi 'in subvi hiyerarşisi	85
Şekil 4.7	Tani.vi diyagramı iki çerçeveli bir "sequence"den meydana gelir	87-89
Şekil 4.8	InitSpeakers sanal cihazı	90
Şekil 4.9	3-DUZaklik sanal cihazı	91
Şekil 4.10	3-D Array.vi diyagramı	92

TABLO LİSTESİ

TABLO 3-1	Bazı Pencerelelerin Matematiksel İfadeleri.....	42
TABLO 3.2	Pencerelelerin Spektral Özellikleri.....	45
TABLO 3.3	C Kodu ile Çerçeveleme ve Pencereleme.....	50
TABLO 3.4	İnciGecis.vi'da Kullanılan SubVI 'lar	61
TABLO 4.1	Konuşmacılara Ait Eğitim Konuşmaları.....	73
TABLO 4.2	Güvenirlik Faktörünün Skala ile Değişimi	80
TABLO 4.3	Tanıma Performansının Skala Faktörü ile Değişimi	82
TABLO 4.4	Yeniden Eğitimden Sonra Tanıma Puanları.....	84
TABLO E1-1	Konuşmacılar	105
TABLO E3-1	Skala Faktörü ile Güvenirlik Oranının Değişimi	114
TABLO E3-2	Skala Faktörü ile Performansın Değişimi.....	116
TABLO E3-3	Yeniden Eğitimden Sonra Tanıma Puanları.....	118

ÖNSÖZ

ODTÜ'de Yüksek Lisans programını bitirmemle doktora programına başlamam arasında onyediyıl geçti. Bu zaman zarfında doktora yapma aşkı daima içimde kaldı. Uzun yıllar Türkiye dışında kalmak bu sevdadan bir şey eksiltmedi. YTÜ'de bu rüyanın hakikat olmasından gerçekten mutluluk duyuyorum.

Beni yaşıma bakmadan bu çalışmaya teşvik eden Elektrik Mühendisliği Bölümü hocalarıma özellikle Prof. Dr. Hüseyin Çakır'a, tez çalışması boyunca daima yanımda olan, sabırla bana yön veren tez danışmanım Prof. Dr. Halit Pastacı'ya duyduğum minnet ve şükran her türlü tarifiin üzerindedir. Onların teşviklerini ve cesaretlendirmelerini hep hatırlayacağım.

Doktora programına başladığımda konuşma tanıma üzerinde çalışma yapmak düşüncem yoktu. Kendisinden aldığım "Konuşma İşareti İşleme" ve "Yapay Nöron Ağları" derslerinde bana bu konuyu sevdiren ve bu iki disiplin arasında bir konu seçmemde büyük etkisi olan Doç. Dr. Fikret Gürge'n'i anmadan edemem. Yine tez süresince bir fahri danışman olarak kıymetli fikirlerini esirgemeyen Dr. Yusuf Kaya Gülüt'e de teşekkür ederim.

Son olarak, bu uzun çalışma boyunca bana katlanan, benimle beraber aynı sıkıntıları yaşayan, bunaldığım zamanlarda moral ve destek veren, rahatlarından fedakarlık eden aileme de şükranlarımı sunuyorum.

Orhan Özhan

ÖZET

Veri sınıflama amacıyla Kendi Kendini Düzenleyen bir tek Harita (SOM) yaygın bir şekilde denetleyicili bir eğitim ile birlikte kullanılır. Konuşmacı tanıma için bu çalışmada denetleyicili bir eğitim olmadan %98 doğrulukla ayrık SOM'lar kullanılmıştır. Metinden bağımsız konuşmacı tanımayı gerçekleştirmek için hem eğitimde hem de tanımda ayrık SOM'lar kullanıldı. TIMIT konuşma veri tabanından farklı yaş, eğitim ve lehçe gruplarına mensup oniki konuşmacı seçildi. İlk eğitimden sonra denetleyicili bir eğitim kullanılmadı. Ünlü seslerden elde edilen temel peryod T_0 ve iki ila yirminci kepsturm katsayıları öznelilik vektörü olarak seçildi. 10 ms aralıklarla üst üste binmiş 32 ms uzunluktaki konuşma çerçevelerinden kepsturm katsayıları elde edildi. Konuşmacı tanımda en iyi glotal uyarım katkısını belirlemek için, deneyler sırasında kepsturm katsayılarına dokunmadan temel peryod skala faktörü değiştirildi. Kadın konuşmacılarda çok miktarda yarım peryod seslendirmesi ile karşılaşıldı. Yarım peryod seslendirmesinden kaynaklanan hatalı T_0 değerlerinden sakınmak için iki geçişli bir öznelilik çıkarma yöntemi tasarlandı. Birinci geçiş sırasında yeni bir diferansiyel temel peryod çıkarıcısı normal perde ile birlikte görülen yarım perde etkisini ortadan kaldırdı. İkinci geçişte fizyolojik olarak olanaksız görülen veri, istatistik yöntemle ayıklandı. En iyi tanıma için glotal uyarım katkısı $0.028T_0$ olarak bulundu. Bu uyarımla %98.3 tanıma skoru elde edildi. Öznelilik çıkarma ve konuşmacı tanıma LabView grafik dili kullanılarak gerçekleştirildi. GUI ortamında yapılan Arge sırasında LabVIEW'nun yazılı programlamadan daha efektif ve değerli bir araç olduğu görüldü.

Anahtar kelimeler: Metinden bağımsız konuşmacı tanıma, SOM, grafik programlama, yarım perde seslendirmesi, glotal uyarım, temel peryod çıkarma, kepsturm analizi

ABSTRACT

A single Self-Organizing Map (SOM) is commonly followed by supervised training for data classification purposes. In this work individual SOM's were utilized without supervised training with 98% accuracy for text-independent speaker identification. Individual SOM's were trained and used to identify speakers independently of the speech being uttered. To realize text-independent speaker identification, twelve speakers belonging to different age, education and dialect groups were selected from TIMIT speech database. No supervised training was used after initial SOM training. Scaled fundamental period T_0 and second through twentieth cepstral coefficients obtained from voiced speech were selected as the feature vector. Cepstrum coefficients were obtained from 32 ms speech frames overlapping every 10 ms. Keeping cepstral coefficients intact, fundamental period scale was varied throughout the experiments to determine the best contribution of glottal excitation to speaker identification. Female speech was observed to manifest a lot of pitch halving. To avoid erroneous T_0 values due to pitch halving, a two-pass feature extraction scheme was devised. During the first-pass a novel differential pitch extractor eliminated the effects of half-pitch sounds occurring with normal-pitch sounds. The second-pass statistically eliminated outliers that are physiologically implausible. Glottal contribution for best identification was found to be $0.028T_0$ for which an identification score of 98.3% was attained. Feature extraction and speaker identification were implemented using LabVIEW graphical language. During research and development on graphical user interface platforms, LabVIEW proved more efficient and valuable than text-based programming.

Key words: Text-independent speaker identification, SOM, graphical programming, pitch halving, glottal excitation, pitch extraction, cepstrum analysis

1 GİRİŞ

1.1 Tipografi

Bu tezde, farklı amaçlar için aşağıdaki fontlar kullanılmıştır:

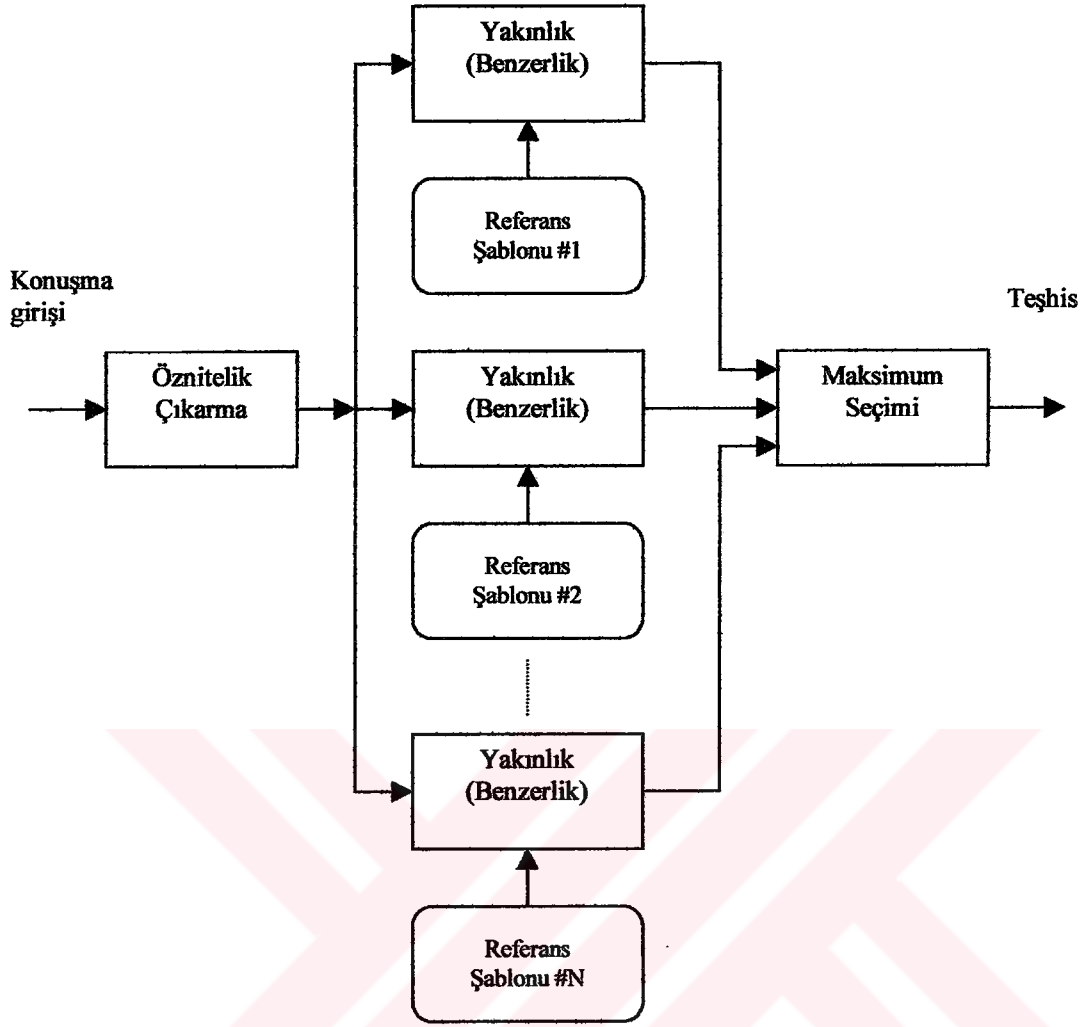
1. Tez metni: "Times New Roman",
2. Başka yazarlardan yapılan alıntılar: "Tahoma, (tırnak içinde)",
3. Konuşma fonemleri, konuşma adları: "*Times New Roman (italik)*",
4. LabVIEW fonksiyon ve sanal cihaz isimleri, program kodu ve dosya isimleri: "Courier New".

Bunun dışında vurgu yapılmak istenen yerler **kalin harflerle**, dikkat çekilmek istenen yerler *italik harflerle* yazılmıştır.

1.2 Önceki Çalışmalar

"Konuşmacı tanıma (speaker recognition), konuşma işaretinde saklı kişilik bilgilerine dayanarak kimin konuştuğunu otomatik olarak tanıma işlemidir. Konuşmacı tanıma iki bölüme ayrılır: a) Konuşmacı tasdik etme (speaker verification) ve b) konuşmacı teşhis etme* (speaker identification). Her iki durumda sistem, kayıtlı konuşmacıların sesleri ile eğitilir. Konuşmacı tasdik sisteminde, bir kişi sisteme belli bir kimlik iddiasıyla başvurur. Sistem ondan belli bir metni veya rastgele bir metni söylemesini ister. Bu konuşma iddia edilen kimliğe ait bilgilere uygunsa kimlik iddiası kabul edilir; uymuyorsa reddedilir. Konuşmacı teşhisinde red veya kabul durumu yoktur; konuşma kayıtlı konuşmacılardan hangisinin sesine uyuyorsa ona sınıflanır. Konuşmacı tanıma başka bir açıdan metne bağımlı ve metinden bağımsız diye de ayrılabilir. Birinci yöntem hem eğitim hem de tanıma sırasında konuşmacıdan aynı metni söylemesini bekler. İkincisi ise sabit bir metnin söylenmesine bağlı kalmaz (Furui, 1996)." Şekil 1.1'de görüldüğü gibi konuşmacı tanıma, öznelilik çıkarma ve referans şablonları ile kıyaslama kısımlarından oluşur.

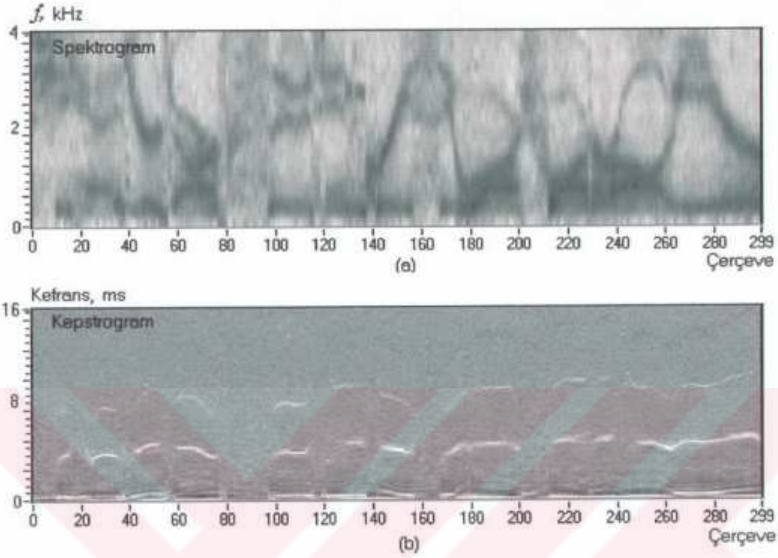
* *Recognition* ve *identification* kelimeleri birbirlerine çok yakın kelimeler olduğu için bunların Türkçeleştirilmesinde birincisine tanıma, ikincisine teşhis demek durumunda kaldık.



Şekil 1.1 Bir konuşmacı tanıma sisteminin prensip şeması. Referans şablonları eğitim sırasında elde edilir ve bilinmeyen konuşmanın kime ait olduğunu tayin etmede kullanılır (Furui, 1996'dan uyarlanmıştır).

Konuşmacı tanımanın güvenlik ve askeri (Clifford, 1991), diyalog halindeki konuşmacıları otomatik olarak etiketlemek (Yu ve Gish, 1993) ve adli tıp (Gregory, 1986) gibi uygulamaları vardır. Bu konu ses bilimi, ses organlarının anatomisi, yapay sinir ağları ve veri sınıflama gibi disiplinleri ilgilendirir. Bu disiplinlerin her biri kendi başına birer bilimdir. Ses gibi, içinde aynı anda bir çok bilgiyi barındıran karmaşık bir işaretten her şeyi atıp yalnız kimlik bilgisini çıkarmak için, bu sayılan disiplinlerin birikimlerini kullanmak doğaldır.

Konuşmacı kimliğini belirleyen iki etken vardır (Nolan, 1983): Kişinin sahip olduğu konuşma organlarının anatomik yapısı ve kişinin bu anatomik yapıyı zaman içinde kendine özgü bir şekilde kullanması. Ses telleri, ses yolunun uzunluğu, ağız ve burun boşluklarının



Şekil 1.2 TIMIT konuşmalarından *sal* cümlesinin (a) spektrogram ve (b) kepstrogramı (Birim çerçeve uzunluğu = 32 ms).

büyüklüğü, tam veya noksan dişler, dudak yapısı gibi anatomik durumlar birinci faktöre girer. İkinci faktör ise zamanla "kazanılmış", öğrenilmiş etkenlerdir: lehçe, telaffuz ve etnik farklılıklar gibi. Birinci grup faktörler formantlar ve temel frekans gibi sesin statik özelliklerini, ikinci gruba giren faktörler formant geçişleri ve koartikülasyon gibi dinamik özellikleri belirler. Konuşmanın zamanla değişimi (temporal karakteristikler) spektrogram, kepstrogram gibi araçlarla gösterilebilir (Şekil 1.2). Farklı ortamlarda yetişmiş tek yumurta ikizlerinde bile ikinci etkenden dolayı konuşma farklılıkları görülür. Bu noktadan hareketle artikülasyon organlarını kontrol eden anatomik modellerle konuşmacı tanıma çalışmaları yapılmıştır (Flanagan, 1972, Flanagan ve Rabiner, 1973). Sesin dinamik özellikleri üzerinde maalesef görüş birliği sağlanamamıştır. Atal (1976), kısa süreli spektral öznelitik vektörlerinin önemli miktarda kişilik bilgisi içerdiğini gösteren deneyler yaptı. Bu çalışmadan uzun konuşma bilgisinin, performansı biraz arttırmakla birlikte, konuşmacı tanıma için zaruri olmadığı sonucu doğdu. Daha sonra Furui (1994) VQ (Vector Quantization) kod defterlerini ergodik HMM (Hidden Markov Model) yapılarıyla

karşılaştırıp o da tanıma için sadece çıkış yoğunluk rakamlarının önemli olduğu sonucuna vardı. Netice olarak bu araştırmacılar, konuşmacının sadece spektrumunu modellemenin tanıma için yeterli olduğunu, spektrumun temporal örüntülerinin (pattern) gereksiz olduğunu söylediler (Colombi, 1996). Bu ise yukarıda işaret ettiğimiz ses özelliklerinin ikincisi ile, "kazanılmış" dinamik özelliklerle, çelişiyordu. Bu çelişkiyi gidermek için Levinson vd. (1983), aralarında Markov'un yaptığı deneyler de bulunan, çok temel deneylere göndermelerde bulundu. Bu deneyler bazı HMM mimarilerinin, dilin yapısını öğrendiklerini, dolayısıyla konuşmacı farklılıklarını modellemeye uygun olmadıklarını ortaya koydu. Bu ikilemin etkisinin hala sürmekte olduğunu düşünüyoruz: Mesela çok yakın bir geçmişte, bu kanının zıddına Sönmez vd. (1997) konuşma işaretinin temporal bir karakteristiği olan temel frekans (pitch) izlemek için istatistik bir model geliştirmişler ve prosodi tabanlı tanımda kullanmışlardır.

Konuşma çalışmalarının tarihi akışı içinde konuşmacı tanıma, konuşma tanımdan sonra gelmektedir. Bunun için kaçınılmaz olarak konuşma tanımdan bazı teknikleri ödünç almıştır. Wolf (1972) evvelce kaydedilmiş sesleri tanımaya yönelik, metne bağımlı ses tanıma amacıyla DTW (Dynamic Time Warping: Dinamik Zaman Deformasyonu) kullandı. Soong (1985) metinden bağımsız konuşmacı tanıma için kümelendirme (clustering) yaklaşımı önerdi. Bu bir VQ yaklaşımıdır; konuşmacı eğitim vektörlerinin kümeleneşi o konuşmacıyı temsil eden model olur. Sınıflama da minimum kuantalama hatası ile belirlenir. VQ ile metinden bağımsız konuşmacı tanımda bir çok başarılı uygulama gerçekleştirildi (Matsui, 1992, Fakotakis, 1993 ve Zhu, 1994).

Baum vd. (1966) öncülüğünde gelişen HMM geçen on yılın konuşma tanıyıcılarına damgasını vurdu (Rabiner, 1989). HMM, konuşma işareti için akustik, fonem, kelime ve dil seviyelerinde modelleme ve tanıma yöntemleri ile birlikte ML (Maximum Likelihood) eğitim algoritmaları sunmaktadır. Mesela konuşmacı tanıma ile birlikte kimlik numaraları veya şifre kodları girilecekse, HMM önce konuşmacı tanıma, arkasından konuşma tanıma yapabilir. Poritz (1982) HMM yöntemlerini konuşmacı tanımda ilk uygulayanlardandır.

Sesin temporal karakteristiğini göz önüne almayan bir başka metinden bağımsız tanıma yaklaşımı da Reynolds'un (1992) getirdiği GMM'dir (Gaussian Mixture Model). "GMM modelinde Gauss bileşenlerinin spektral şekillerinin konuşmacıya bağlı olduğu gösterilmiştir. Bunlarla konuşmacı kişiliği efektif olarak modellenilebilir (Reynolds, 1995)." GMM tekniğinde

temporal konuşmacı karakteristikleri gözönüne alınmadığı için kısa konuşma örneklerinden modelleme/sınıflama yapılmaktadır. Bu yüzden GMM tek durumlu özel bir HMM olarak düşünülebilir.

"Bu çalışmaların hepsinde de bütün konuşma etkileşim ve geçişleri tek bir modelle temsil edilmiştir. Bu durum ise onlarca fonem modelinin veya binlerce tri-fon modelinin kullanıldığı konuşma tanıma ile bir tezat oluşturmaktadır.

Gizli Markov modelleri konuşma gözlemlerinde hatalı ve basit varsayımlarda bulunur (Colombi, 1996)."

Colombi'nin bu sözleri yukarıda sözünü ettiğimiz konuşmanın temporal bilgi içeriği ile ilgili diğer yaklaşıma karşı bir itirazdır. Biz de konuşmanın dinamiği olan temporal karakterinin konuşmacı tanımada korunması görüşündeyiz. Çalışma yöntemimiz Colombi ile örtüşmemekle birlikte konuşma işaretinin kimlik belirleyici unsurlarına bakışımız aynıdır. Onun yöntemi parametrik sınıflama yöntemi, bu çalışmanın yöntemi ise parametrik olmayan sınıflamadır.

Bir nonparametrik bilgi sınıflama yöntemi olan yapay sinir ağları bilgi depolayabilirler. Fischler ve Firschein (1987) bilgiyi şöyle tanımlar:

"Bilgi, dış dünyayı yorumlamak, tahmin etmek veya uygun bir cevap vermek için insan veya makina tarafından depo edilen malumat veya modellere denir."

Bilgi temsiliinin iki karakteri vardır: 1) Hangi malumat öne çıkarılacaktır, 2) bu malumat daha sonraki kullanım için fiziki olarak ne şekilde kodlanacaktır. Bu noktada yapay sinir ağlarıyla klasik örüntü sınıflayıcıları arasında bir fark ortaya çıkmaktadır: Klasik bilgi işleyen sınıflayıcılar çevre gözlemlerinin matematik bir modelini yapmakla işe başlar, gerçek verilerle bu modelin sağlamasını yapar, sonra tasarımı bu model üzerine kurar. Bunun aksine, yapay sinir ağında tasarım doğrudan doğruya gerçek veriler üzerine kurulur, veri adına (model değil) verinin kendisi konuşur. Böylece yapay sinir ağı sadece dış dünyanın bir örtülü modeli olmakla kalmaz, aynı zamanda bilgi işleme ünitesi rolünü de üstlenir (Haykin, 1994). Yukarıda değinilen HMM ve GMM, model-tabanlı sınıflayıcılara; VQ ve yapay sinir ağları ise parametrik olmayan sınıflayıcılara girer. Yapay sinir ağında çevre ile ilgili bilgi nöronların sinaptik ağırlıklarında depolanır.

Oglesby ve Mason (1991) konuşmacı tanımak için RBF (Radial Basis Function: Radyal Temel Fonksiyonlu) ağı kullandılar. Tanınacak her konuşmacı tek çıkışlı olan kendi ağıma sahipti. RBF ağları iki katmanlı, ikinci katman da lineer sınıflayıcı olduğu için eğitimleri MLP (Multi Layer Perceptron: Çok Katmanlı Algılayıcı)'dan on kere daha hızlıdır. Öznetelik olarak kısa süreli spektrum kullandılar. Sistemleri hem konuşmacı teşhisi hem de konuşmacı tasdiki için kullanılabiliyordu. Zeek (1996), HMM modelini tek katmanlı bir algılayıcı ile kullandı ve tanımda %88 doğruluk elde etti. Aynı çalışmasında bir de Öznetelik Uzak Yörüngesi – Yapay Sinir Ağı kullandı. Büyük bir veri tabanı ile %65 tanıma skoru sağladı. Aynı şartlardaki bir HMM %12-24 daha düşük performans gösterdi. Zeek'e göre (1996) "Öznetelik Uzak Yörüngesi görüntü işlemede temporal tanıma için geliştirildi ve bazı görüntü dizilerinde HMM'den daha yüksek performansa sahip olduğu gösterildi." Zeek'in (1996) temporal karakteristikler ve HMM'le ilgili sözleri Colombi'nin (1996) sözlerini ve bu çalışmanın yola çıkış motivasyonunu desteklemektedir. Sharma ve Mammone (1996) T-Netix, Inc. şirketinde kelimedeki küçük birimlerle çalışan metne bağımlı ticari bir konuşmacı tasdik sistemi geliştirdiler. NTN (Neural Tree Network) adını verdikleri sinir ağı ile kelime altı birimleri (sub-word units) modellediler. Leinonen vd. (1991) SOM (Self Organizing Map) ağı kullanarak ses bozukluklarını teşhis eden fonotopik haritalama yaptılar. Klinik hekimlerinin subjektif karar vermelerini önlemek için tasarlanan bu sistemde SOM, konuşmacılardan alınan ses örneklerinden fonotopik harita elde ediyordu. Bu haritalarda normal ses ile kısık, "çatlak" ses vb. ayırt edilebiliyordu. Kohonen (1988b) gerçek zamanda çalışan ve Fince sesleri yazıya çeviren bir daktilo tasarladı. Daktilo 12X8 büyüklüğünde bir SOM kullanıyordu.

1.3 Motivasyon ve Amaç

Literatür taraması sırasında konuşmacı tanıma yönelik çalışmaların büyük çoğunluğunu HMM, GMM gibi parametrik modellerin oluşturduğu görüldü. Parametrik olmayan sınıflama çalışmalarında da VQ ile MLP, RBF gibi yapay sinir ağları kullanma eğilimi gözlemlendi. SOM'u bulan Teuvo Kohonen'in (1988b) SOM tabanlı bir fonetik daktilo çalışması bulunmaktadır. Bu çalışma, konuşma tanıma tekniklerini kullanarak konuşmayı yazıya çeviriyordu. Bunun dışında SOM yapay sinir ağı ile gerçekleştirilmiş bir konuşmacı tanıma çalışmasına rastlanmadı.

Literatürde sınıflama amacıyla SOM kullanımı konusunda şu tavsiyelere rastlanmaktadır:

"... SOM'un amacı sadece giriş vektörlerinin metrik-topolojik bağıntılarını görülebilir hale getirmektir. Bu yüzden SOM kendi başına sınıflama veya karar verme amacıyla kullanılmamalıdır (Haykin, 1994)."

"... örüntü sınıflamada en iyi sonuçları alabilmek için öznitelik haritası, denetleyicili bir öğrenme ile birlikte kullanılmalıdır (Kangas vd. 1990)."

Bu tavsiyelerin hepsi de tek bir SOM kullanımını kastettiklerinden biz her konuşmacı için ayrı bir SOM kullanarak denetleyicili eğitime gerek kalmadan konuşmacıları sınıflayabileceğimizi düşündük. Bunu yapmaya yönelten sebep de SOM'ların giriş uzayının istatistiğini öğrenebilme yetenekleridir. Her SOM kendi konuşmacısını yeterince öğrenebilirse ayrıca denetleyicili bir eğitime gerek kalmayacaktır. Sınıflamada, bilinmeyen konuşma oluşturulan SOM'larla birer birer karşılaştırılacaktır. Bilinmeyen konuşma en fazla "yakınlık" gösteren SOM'a sınıflanacaktır. Buna paralel olarak konuşmacı belirlemede ses telleri uyarımı ile artikülasyon karakteristiklerinin en iyi karışımını da bulmak istiyoruz.

1.4 Yöntem

Şekil 1.2 bir TIMIT konuşmasının spektrogram ve kepstrogramını göstermektedir. Spektro-gram ve kepstrogram ardarda gelen zaman dilimlerinde ses işaretinin Fourier ve kepstrum analizlerinin gösterimidir. Şekil 1.2(a)'da Fourier analizi kullanılarak artikülasyon filtresinin frekans genlik cevabının temporal görünüşü elde edilmiştir; şekilde siyah izler formantlardır. Şekil 1.2(b)'de homomorfik analiz kullanarak konuşmanın kepstrumu elde edilmiştir; kepstrogramın ortasındaki beyaz izler temel periyodu göstermektedir. Bu izler kişiden kişiye farklı şekiller alır. Konuşma spektrumunun 128 bileşeni, formantlar, formant genişlikleri, belli frekans bantları içindeki enerjiler öznitelik olarak kullanılabilir. Fakat kepstrum, spektrumdan daha fazla sıkıştırılmış bilgi içerdiğinden ve kepstrum formant spektrumundan önce elde edildiğinden öznitelik vektörlerimizi kepstrum analizi ile elde etmeye karar verdik.

1.4.1 Konuşma verileri

Eğitim ve tanıma için konuşma veri tabanı olarak TIMIT veri tabanı kullanılmıştır. TIMIT, DARPA-ISTO (Defense Advanced Research Project Agency - Information Science and Technology) sponsorluğunda MIT (Massachusetts Institute of Technology), SRI (Stanford Research Institute) ve TI (Texas Instruments)'ın ortak çabalarıyla oluşturulmuş okunmuş konuşmaları içeren bir kaynaktır. ABD'de konuşulan sekiz diyalekt bölgesinden seçilmiş kadın ve erkek 630 konuşmacı 10'ar adet cümle okumaktadırlar. 10 cümleden ikisi standard (*sa* cümleleri), beşi değişken (*sx* cümleleri), üçü ise her konuşmacı için bir tanedir (*si* cümleleri). TIMIT gürültüsüz bir konuşma kaynağıdır ve konuşmalar 8 kHz ile örneklenmiştir. TIMIT verileri Pennsylvania Üniversitesi Linguistic Data Consortium, <http://www.ldc-upenn.edu> adresinden INTERNET aracılığı ile elde edilebilmektedir. EK1'de TIMIT konuşma tabanı ile ilgili ayrıntılı bilgi verilmiştir.

1.4.2 Öznitelik çıkarma

Kullanılan sınıflama yöntemi ne kadar iyi olursa olsun, sınıflamaya esas olan öznitelikler doğru seçilmemişse sınıflama yönteminden herhangi bir performans beklenemez. Ses işareti zengin bir bilgi kaynağıdır ve bu kaynaktan sadece kişilik tanımamıza yarayacak ve yetecek olanları almak zorundayız. Kişilik belirten özniteliklerin yanısıra mesela linguistik veya yöresel öznitelikleri almanın bizim amacımıza hiç bir faydası ve katkısı yoktur; belki yanıltıcı sonuçlara bile yolaçabilir.

Konuşmacı tanıma için ses işaretini zaman domeninde 32 ms'lik dilimlere (çerçeve) ayırdık. Sonra spektral sızıntıyı azaltmak için bu çerçeveleri Hamming penceresiyle çarptık. Ses işaretinden öznitelikler elde edilip bunlardan sadece ünlü kısımları eğitim ve tanıma için kullandık. Kişilik belirleyebilecek /r/, /s/, /k/ gibi ünsüzler ise bu çalışmada kullanılmadı. Şekil 1.2(b)'de beyaz izler ünlü kısımları gösterir.

Buradan hareketle bu tezde sınıflamaya temel olacak öznitelikler:

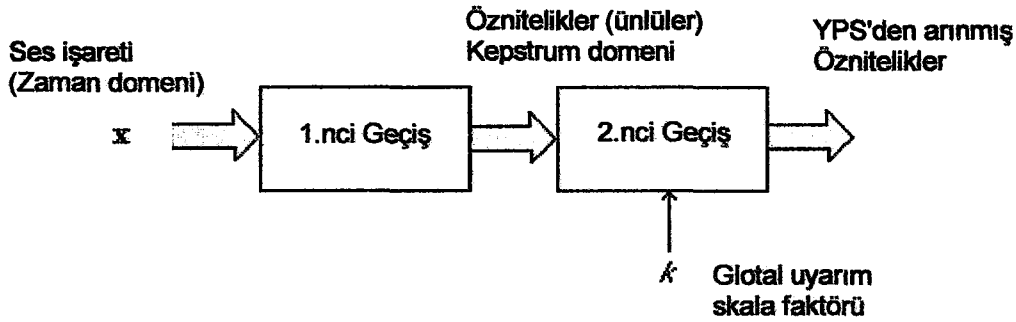
1. Glotal uyarım: kT_0 , (Glotal uyarım skala faktörü x Temel periyod)
2. $c_1, \dots, c_{19}$ kepstrum bileşenleri

olarak belirlenmiştir. En iyi tanıma için glotal uyarım faktörü k 'nın alması gereken değer

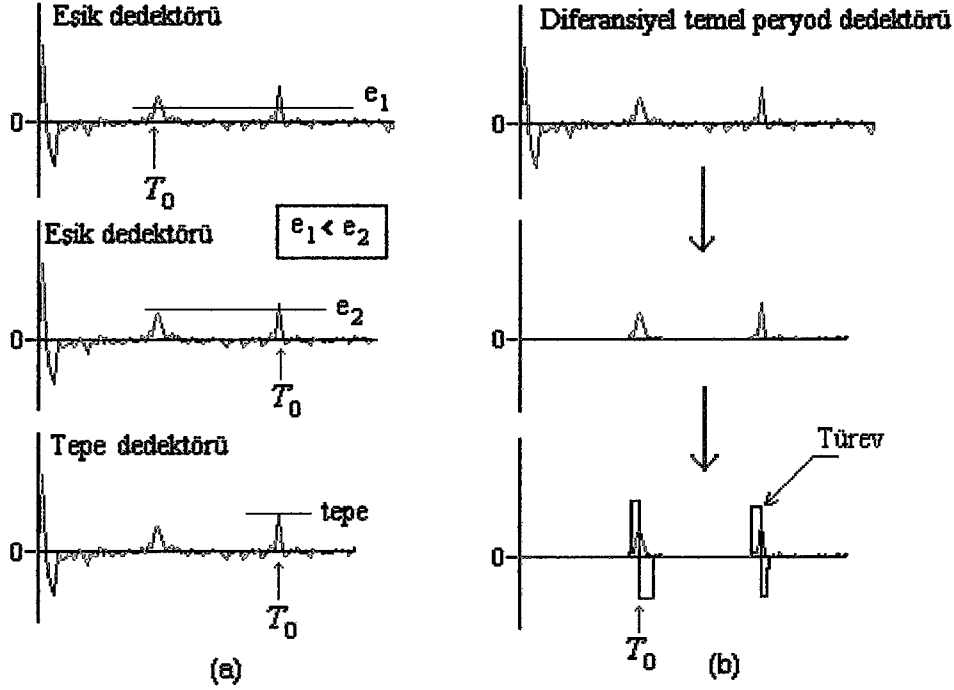
belirlenip, bu deęer konuřmacı tanımaya esas alınmıřtır.

řekil 1.2(b)'deki kepstrogramda temel peryodu temsil eden beyaz çizginin üst tarafında görölen (temel peryodun iki katına eřit uzaklıktaki) soluk beyaz iz bu çalışmada Yarım Perde Seslendirmesi diye adlandırdığımız bir olaydır. Bölüm 3'te etraflıca ele alacağımız bu olay temel peryod bulmayı güçleştirir. YPS (Yarım Perde Seslendirmesi) bazen normal seslendirmeden daha kuvvetli olur ve temel peryod deęeri $2T_0$ olarak bulunur. Bu hata farkedilmezse kadın sesi erkek sesi olarak sınıflandırılabilir. Literatürde YPS'nin azaltılması veya yok edilmesi hakkında doyurucu bir kaynak bulamadık. Bu yüzden YPS etkisini azaltmak için iki kademeli öznitelik çıkarma işlemi geliřtirdik (řekil 1.3). Bunlar 1.nci Geçiř ve 2.nci Geçiř diye adlandırıldı

1.nci Geçiř, zaman domenindeki konuřma iřaretini homomorfik analizle kepstrum domenine dönüřtürür. İřaretin sadece ünlü kısımları alınıp ünsüz kısımları atılır. Yukarıda belirtilen öznitelikler elde edilir. YPS etkisini azaltmak için 1.nci Geçiř'te bir diferansiyel temel peryod dedektörü geliřtirdik. Normal perde ve yarım perde ikisi birden bulunur ve normal perdenin genlięi yarım perde genlięinden küçük olduęu zaman diferansiyel temel peryod dedektörü temel peryodu hatasız olarak belirler. Bunun için "tepe dedektörü" veya "eřik dedektörü" kullanılması durumunda temel peryod olarak $2T_0$ 'ın atanması durumu ile karřılařılabilir.



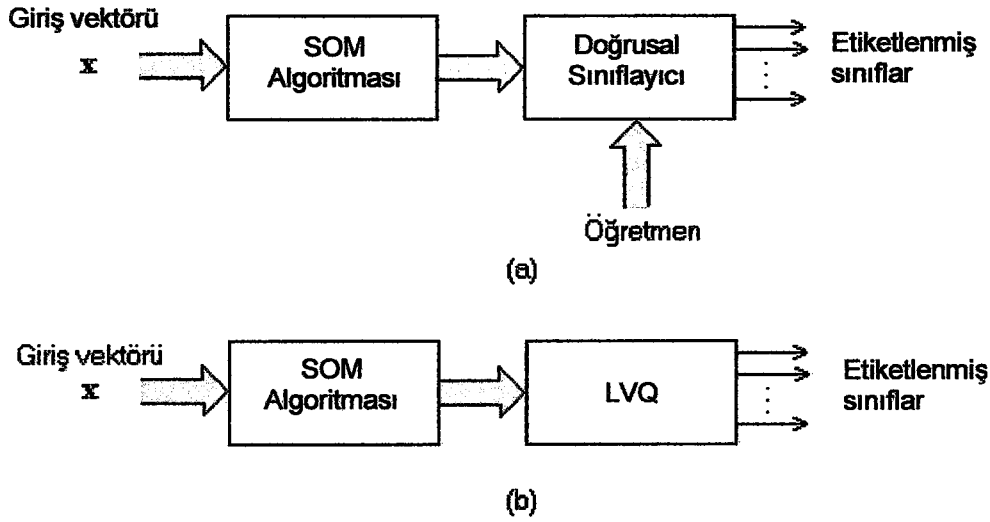
řekil 1.3 İki kademeli öznitelik çıkarıcısı.



Şekil 1.4 Temel periyod dedektörlerinin karşılaştırması. (a) Eşik dedektörü ve tepe dedektörü; (b) Diferansiyel temel periyod dedektörü.

Şekil 1.4'de periyod dedektörleri karşılaştırılmıştır. Şekil 1.4(a)'da bir konuşma keştrumu üzerinde eşik dedektörünün kullanılan eşik değeri duyarlılığı ve tepe dedektörünün doğru temel periyodu yakalayamama durumları gösterilmiştir. Diferansiyel temel periyod dedektöründe ise yirminci keştrum bileşeninden sonra türev alınmaktadır. Normal perde ile yarım perde bir arada olsalar bile türevin ilk pozitiften negatife geçişi T_0 olarak alındığı için $2T_0$ noktasında bulunan yarım perde seslendirmesi etkisiz hale getirilmiş olur (Şekil 1.4(b)).

Buna rağmen bazen normal perde bulunmayıp, onun yerine yarım perde bulunabilir. O zaman diferansiyel pitch dedektörünün $2T_0$ 'a karar vermekten başka bir çaresi yoktur; "1.nci Geçiş" bloğunun çıkışında temel periyod $2T_0$ olarak okunur. 2.nci Geçiş bloğu bu durumu düzeltir. "2.nci Geçiş" bütün konuşma boyunca temel periyodun ortalama ve standard sapma değerlerini hesaplar. (Ortalama değer ∓ 1.75 Standard sapma değeri) dışında kalan temel periyod değerlerine sahip öznitelik vektörleri reddedilir ve kayıttan çıkarılır. Temel periyod glotal uyarım faktörü k ile çarpılır ve öznitelik vektörüne kT_0 kaydedilir.



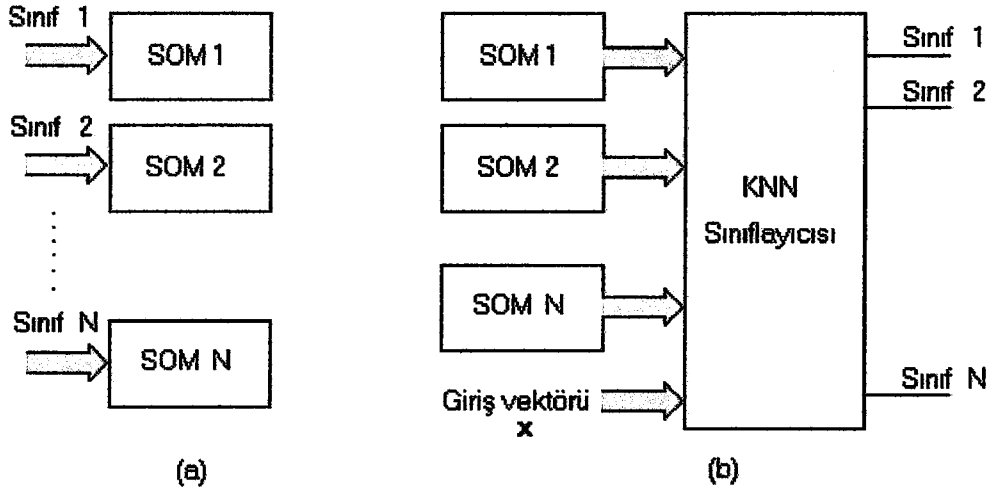
Şekil 1.5 Bir SOM kullanarak sınıflama. SOM'un arkasından (a) doğrusal sınıflayıcı, (b) LVQ kullanılabilir (Simon Haykin: *Neural Networks – A Comprehensive Foundation*, 1994'ten uyarlanmıştır).

1.4.3 SOM

Paragraf (1.3)'te değinildiği gibi SOM yapay sinir ağları veri sınıflandırma için değil veri içinde saklı olan öznitelikleri görsel hale getirmek için kullanılmışlardır (Kangas vd., 1990). Haykin (1994) SOM'u sınıflama amacıyla kullanmayı tavsiye etmemekte, eğer kullanmak gerekiyorsa SOM'dan sonra bir doğrusal sınıflayıcı veya LVQ (Learning Vector Quantizer) kullanılmasını önermektedir (Şekil 1.5). Bu şekilde sınıf sınırlarının "ince ayarı" yapılmış olmaktadır (Kohonen, 1988a). Burada dikkat edilmesi gereken nokta bu yazarlar sınıflama için "bir tek" SOM'un kullanılması halini kastetmektedirler. Bu çalışmamızda yenilik olarak birçok sınıf için bir tek SOM kullanmak yerine her sınıf için ayrı bir SOM kullanmayı seçtik. Bu yaklaşımla her SOM'u kendi verisiyle eğittik. Böylece bir sınıfın SOM'u sadece kendi sınıfının içindeki öznitelikleri öğrenip diğer sınıfların özniteliklerinden etkilenmemektedir. Tek SOM'la sınıflar arasında oluşan geçiş bölgeleri ayrık SOM eğitiminde bulunmamaktadır.

Bu yaklaşım SOM eğitiminden sonra bir eğitim daha gerektirmemektedir. Önce her sınıfın eğitim verileriyle o sınıfın SOM'una sınıfın sahip olduğu öznitelikler öğretilmiştir (Şekil 1.6a). Tanıma sırasında KNN (K-Nearest Neighbor, K-En yakın komşu) sınıflayıcısı oluşturulan ayrık SOM'ları kullanarak bilinmeyen giriş vektörünü sınıflandırılmaktadır. Bu şekilde SOM'un giriş vektör uzayının özniteliklerini öğrenebilme özelliğini kullanırken, ayrık SOM'lar kullanıp sınıfların birbirlerinin özniteliklerini olumsuz yönde etkilemeleri

önlenmiştir.



Şekil 1.6 Geliştirilen konuşmacı tanıma sistemi. (a) Eğitim: TIMIT veri tabanından konuşmacıların SOM haritaları oluşturulur; (b) Tanıma: Bu haritaları kullanarak bilinmeyen konuşmanın kime ait olduğu bulunur.

Tanıma sistemimiz eğitim ve test bölümlerinden oluşmaktadır. Şekil 1.6 sistemin blok şemasını göstermektedir. Her TIMIT konuşmacısının SOM'u o konuşmacının *sx* cümleleriyle eğitildi. Konuşmacıların *sx* cümleleri birbirlerinden farklı olduğu için metinden bağımsız bir eğitim gerçekleştirilmiş oldu. Test sırasında *sa*, *sx* ve *si* cümlelerinin hepsi de kullanıldı. On cümleden sadece *sa1* ve *sa2* cümleleri sabit, diğer sekiz cümle değişken olduğu için tanıma da metinden bağımsız tanıma gerçekleştirilmiş oldu.

Bu düzenekle glotal uyarım faktörünün çeşitli değerleri için eğitim ve tanıma işlemleri tekrarlandı. Deneylerin performansını belirleyen bir performans fonksiyonu tanımlayıp en iyi tanıma için gereken glotal uyarım faktörünü belirledik. Bu faktörü kullanarak hatalı tanıma görülen konuşmacıların SOM'larına ek konuşmalarla daha fazla bilgi aktarıldı. Bunun sonucunda "Ayrık SOM Konuşmacı Tanıyıcı"mız %98.3 tanıma performansına ulaştı.

SOM haritalarını oluşturmak için Helsinki University of Technology'deki SOM Programlama Grubu'nun SOM_PAK yazılımı (Kohonen vd., 1995), oluşturulan SOM haritalarını göstermek için bazen NENET Grubu'nun hazırladığı NENET 1.0b yazılımını

kullandık (Elomaa vd, 1998). Her iki programa da INTERNET aracılığıyla ulaşılabilmektedir.

Çalışmamızın ÖZNETLİK ÇIKARMA ve TANIMA kısımlarını LabVIEW 4.01 ortamında gerçekleştirdik. LabVIEW, WINDOWS işletim sisteminde çok hızlı uygulama geliştirmeye, veri toplama ve işlemeye yönelik bir grafik programlama dilidir. Kullanıcıyı WINDOWS'un alt seviye program arabirimlerine ve C veya C++ gibi dillere muhtaç etmeyecek, fakat isterse bu dillere ulaşma imkanı veren GUI (Graphical User Interface) birimlerine sahiptir. LabVIEW ile kullanıcı sanal bir cihaz tasarlar. LabView buna VI (Virtual Instrument) adını vermiştir. Her VI bir cihaz panosundan, bir de diyagramdan oluşur. Pano sanal cihazın dış görünümüdür. Diyagram ise bu cihazın nasıl çalışması gerektiğini belirleyen grafik programıdır. Tezimizde birkaç NENET grafik çıkışları hariç, tamamen kendi yazdığımız (daha doğrusu çizdiğimiz) VI'lerden aldığımız grafikleri kullandık. Şekil 1.2'deki spektrogram ve kepstrogram yazdığımız bir sanal cihazın panosundan alınmıştır.

1.5 Tezde Getirilen Yenilikler

Bu tez çalışmasında üç orijinallik ve bir de buluş bulunmaktadır. Buluş en iyi tanıma için gerekli olan global uyarım faktörünün bulunmasıdır. Her ne kadar aşağıdaki listeye almasak da burada LabVIEW'yu anmakta fayda görüyoruz. Bu çalışma gibi analiz ve grafik ağırlıklı gerçek-zaman veya offline uygulamalara örnek olması açısından LabVIEW kullanımı da bir orijinallik olarak düşünülebilir. Aslında LabVIEW'nun bizatihi kendisi orijinaldir. LabVIEW platformunda varılmak istenen hedefe çok daha hızlı ulaşılabileceği inancındayız.

1.5.1 Ayırık SOM yapay sinir ağları

SOM literatüründe SOM'un sınıflar arası öznetelik farklılıklarını görebilmek amacıyla kullanılması tavsiye edilmektedir. Ayrıca bütün sınıflar için bir tek SOM kullanılması durumu söz konusudur. Tanıma için, elde edilen SOM'un ikinci bir eğitimden geçirilmesi önerilmektedir. Biz her sınıf için ayırık SOM kullandık. Elde edilen SOM'ları ek bir eğitime tabi tutmadık. Elde edilen SOM'lar ek eğitime gerek kalmadan KNN gibi basit bir

sınıflama algoritması ile kullanılabilirler. Böylece konuşmacı tanıma probleminde Ayrık SOM Yapay Sinir Ağları ile %98.3 tanıma başarısı elde edilmiştir.

1.5.2 İki geçişli öznitelik çıkarma

Fizyolojik bir olgu olduğunu düşündüğümüz YPS (Yarım Perde Seslendirmesi) temel peryodu bulurken daima karşımıza bir problem olarak çıktı. Kadın konuşmacılarda konuşmanın bazı kısımlarında temel peryod yerine iki katı değerindeki yarım perde peryodu yakalanıyordu. Bu hem eğitim açısından hem de tanıma açısından seste "erkekleşme eğilimi" anlamına gelmektedir. Kepstrumdan temel peryod çıkarırken bu durumu azaltacak önlem aldık (paragraf 1.4.3). Ancak yine bazı konuşma çerçevelerinde bu önlemin yalnız başına yeterli olmadığı görüldüğü için **İki Geçişli Öznitelik Çıkarma** yöntemi geliştirildi. Birinci geçişte YPS'ye karşı bir ön işlem uygulanmakta, İkinci Geçişte konuşmanın bütününde temel peryod istatistiği bulunarak bununla YPS ayıklanmaktadır.

1.5.3 Diferansiyel temel peryod dedektörü

Diferansiyel temel peryod dedektörü, normal perde ile yarım perde bir arada bulundukları zaman normal perdeyi ayırabilen basit bir türevsel temel peryod dedektörüdür. Normal ve yarım perde genliklerine duyarlı olmayan bu dedektör, kepstrumun perde tepelerinin eğimlerinde ilk pozitiften negatife geçiş anını temel peryod olarak yakalar.

1.5.4 Glotal uyarım faktörünün bulunması

Glotal Uyarım Faktörü temel peryodun öznitelik vektöründe ne ölçüde temsil edileceğini belirler. Temel peryodun öznitelik vektöründe aşırı temsil edilmesi hiç temsil edilmemesinden daha kötü sonuçlar doğurur. En iyi tanıma için bu faktörün sahip olması gereken değerin 0.028 civarında olduğunu bulduk.

1.6 Tezin Bölümleri

Tezin bölümlere dağılışı ana hatlarıyla aşağıdaki gibidir:

1.6.1 Giriş

Önceki çalışmalar ve tez çalışmasının tanıtımı.

1.6.2 Temel bilgiler

Bu bölüm çalışmada kullanılan kullanılan kaynaklar (ses işareti ve konuşma organları), analizler ve yöntemler hakkında bir arka plan niteliğindedir. Önce konuşma işareti, konuşma organlarının anatomisi ve konuşma fizyolojisi ele alınmış, konuşma işaretinde kişilik belirleyen öznitelikler anlatılmıştır. Sonra konuşma modeli ve bu modele dayanarak özniteliklerin nasıl elde edileceği üzerinde durulmuştur. Burada kulağın sesi algılamak farklı frekanslar arasındaki faz farkına "duyarsız olduğu kabul edilip", bu kulak modeline uygun bir biçimde özniteliklerin elde edileceği belirtilmiştir. Son olarak da SOM Yapay Sinir Ağı hakkında gereken bilgi sunulmuştur.

1.6.3 Öznitelik çıkarma

Önce konuşma işareti üzerinde uygulanan çerçeveleme ve pencereleme anlatılmıştır. Çerçeve uzunluğu olarak 256 örnek kabul edilmiştir. Bu çerçeveler TIMIT konuşmaları 8 kHz ile örneklendikleri için bu 32 ms uzunluğundadır. Spektral sızıntıyı kabul edilebilir boyutlarda tutmak için çerçeveler Hamming penceresiyle çarpılmıştır. Pencereleme dolayısıyla bastırılmış olan çerçeve kenarlarındaki bilginin tekrar vurgulanabilmesi için çerçeveler 10 ms aralıklarla üst üste bindirilmiştir. Bundan sonra özniteliklerin seçimi, yarım perde seslendirmesi vurgulanmış; iki kademeli öznitelik çıkarma işleminin LabVIEW ile nasıl gerçekleştirildiği (birinci geçiş, diferansiyel pitch dedektörü, ikinci geçiş vb sanal cihazları) anlatılmıştır. LabVIEW nisbeten yeni bir platform olduğu için sanal cihaz kavramı tanıtılmıştır. Sanal cihazların çalışması zaman zaman LabVIEW yapılarıyla C programlama dili arasında paralellikler kurularak anlatılmıştır.

1.6.4 Eğitim ve tanıma

Dördüncü bölümde her konuşmacı için kullanılan eğitim konuşmaları, eğitim ve tanıma kullanılan ayırık SOM'ların topografisi, eğitim parametreleri tanıtılmış, daha sonra tanıma algoritması verilmiştir. Bundan sonra glotal uyarım faktörünün belirlenmesi için yapılan

deneylerden ve bu deney sonuçlarından glotal uyarım faktörünü nasıl belirlediğimiz anlatılmıştır. Belirlenen glotal uyarım faktörü ($k = 0.03$) kullanılarak, hatalı tanıma görülen konuşmacıların SOM'ları istatistik olarak zenginleştirilip tanıma performansı 120 konuşmadan 118 (%98.3) olarak elde edilmiştir. En son olarak tanıma için LabVIEW ile gerçekleştirilen sanal cihaz anlatılmıştır

1.6.5 Tartışma ve öneriler

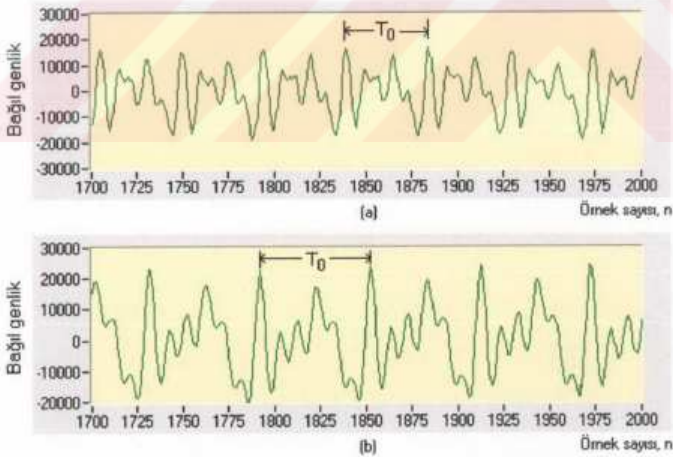
Bu çalışmadan elde edilen sonuçların tartışılması, ileriye dönük öneriler.

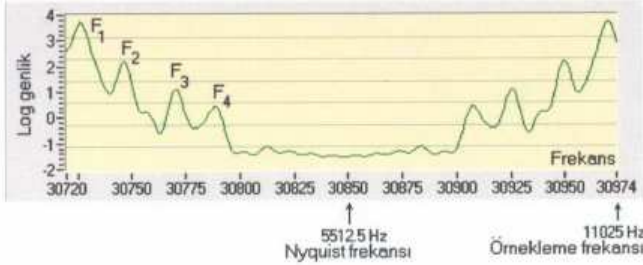


2 TEMEL BİLGİLER

2.1 Konuşma İşareti

Ses, çok karmaşık bir stokastik işarettir. İçinde, linguistik, cinsiyet, kişilik, fizyoloji, ruh hali, dil, yaş, ortam, lehçe, musiki gibi pek çok bilgiyi aynı anda bulundurabilir. Uzun yıllardır ses bilimcileri, bu işareten istenilen bilgiyi elde etmek, veya birden çok işareti belli bir bilgi itibarıyla sınıflandırmak için çalışmaktadırlar. Bu çalışmalar sonucunda pek çok faydalı sonuç elde edilmiş ve bazıları bugün günlük hayatta kullanılmaya başlamıştır (Clifford, 1991, Furui, 1994, Colombi, 1996). **Konuşma sentezi**, **konuşma tanıma** ve **konuşmacı tanıma** buna örnektir. Konuşma tanıma, ses araştırmalarının eski ve gelişmiş dallarından birisidir. Konuşma tanıma teknolojisi ile hava yolu rezervasyonları, sesle bilgisayar kumandası artık topluma kazandırılmış uygulamalardır. Konuşmacı tanıma





Şekil 2.2 /a/ sesinin frekans spektrumu. Örnekleme frekansı = 11025 Hz. Rezonans tepeleri (F_1 , F_2 , F_3 , F_4) formantları belirlenmektedir.

alanındaki uygulamalara örnek olarak VoicePass telefon kartı (VoiceID Systems, Inc.), Sprint FONCARD® (TI), gösterilebilir.

Şekil 2.1(a)'da zaman domeninde bir çocuk tarafından telaffuz edilen ünlü seslerden /a/ sesi görülmektedir. Bu işaretle, yukarıda saydığımız, linguistik (/a/ sesi), kişilik, cinsiyet, yaş v.s. bilgileri bulunmaktadır. Bu bilgilerden bir tanesini ele veren bir ipucu, **temel periyod**, şekilde T_0 yazılarak işaretlenmiştir. Temel periyodun tersi **temel frekans** veya **perdedir**. Temel frekans F_0 ile gösterilir. Sesin perdesi, tizliği ve pesliği ile alakalıdır; ve tam olarak temel frekansa eşittir. Burada temel periyodun 3.9 ms (256 Hz) olmasından, bu sesin bir çocuk sesi olduğu anlaşılır. Çünkü bu değer, yetişkin bir erkekte 150 Hz, yetişkin bir kadında ise 200 Hz civarındadır. Şekil 2.1(b), yetişkin bir erkek tarafından söylenen /a/ sesindeki temel periyod farkını belli etmektedir. Ses işaretinde, temel periyod gibi tanımaya yönelik olarak kullanılan değerlere **öznitelik** (feature) denir. Gerçekleştirmeye çalıştığımız tanıma için gerekli olan öznitelikleri, ses işaretinden ayırma işlemine **öznitelik çıkarma** (feature extraction) denir. Bir uygulama için gerekli olan öznitelik(ler) başka bir uygulama için gerekli olmayabilir. Konuşanı tanımak ve doğrulamak için F_0 veya T_0 bilgisi zaruridir; fakat konuşandan bağımsız konuşma tanımada gerekli değildir.

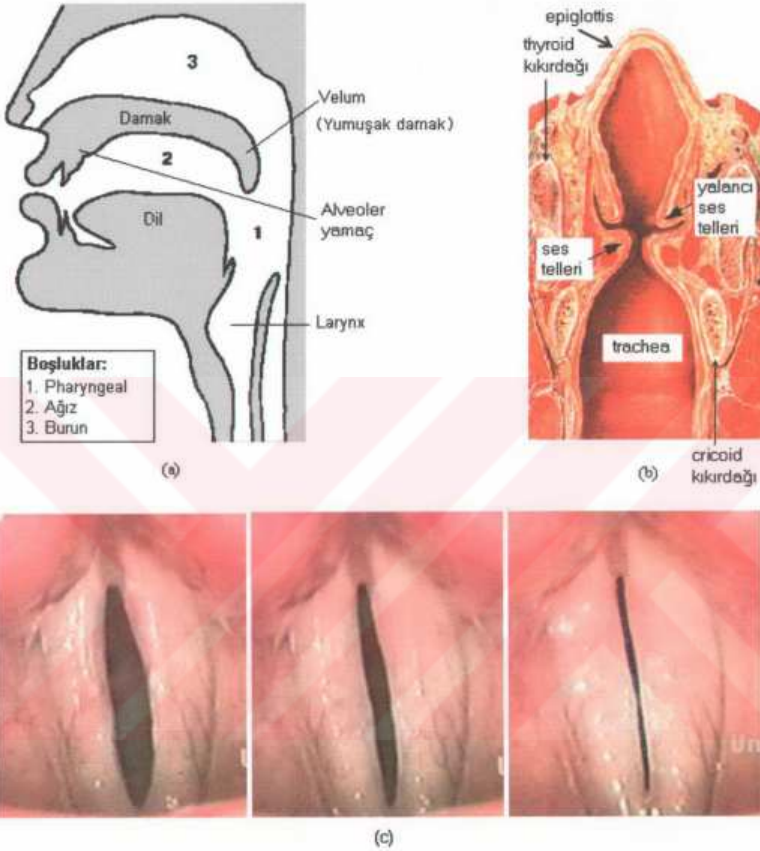
Ses işaretini zaman domeninde analiz etmek pek kolay ve öğretici değildir. Aşağıda, insanda ses üretim modeli ele alındığı zaman görüleceği gibi, sesin frekans domenindeki içeriği bize daha fazla bilgi ve ipucu vermektedir. Şekil 2.2'de /a/ sesinin çift taraflı formant spektrumunu görüyoruz. Artikülasyon kanalının rezonans frekanslarına **formant** denir. Ses kanalının kesit alanı, yer ve zamanla değiştiği için, zamanla değişen birden fazla

doğal rezonans frekansı, yani formant vardır. Formantlar $F_1, F_2, \dots$ şeklinde adlandırılır. Formantların şiddetlerini zamana bağlı olarak gösteren grafiklere **spektrogram** denir. Şekil 2.5(a), (c) ve (e) /a, e, i, i, o, ö, u, ü/ sesinin değişik söylenme şekillerine ait birer spektrogram örneğidir.

2.2 Ses Üretimi

İnsanda ses organları, boğazın alt tarafında ses telleri, bütün boğaz, yumuşak damak (velum), ağız boşluğu, dil, damak (palate), alveoler yamaç, dişler, dudaklar; geniz-burun boşluğu ve burun delikleridir (Şekil 2.3a). Konuşma sırasında, ses kanalının şekline yaptığı tartışmasız katkıdan dolayı, alt çene kemiği de ses organlarından kabul edilmektedir. Bu organların hepsi de konuşma üretimine (articulation) katkıda bulunur. Ses telleri, boğazın alt tarafında ve tracheanın üzerinde larinks (larynx) içinde yer alır [Şekil 2.3(b),(c)]. Ses telleri, önde thyroid kıkırdaklarından, arkada arytenoid kıkırdaklarına kadar uzayan kas ve mukoza zarından ibaret esnek bir banttır. Ses tellerine anatomik yapılarından dolayı ses kıvrımları (vocal folds) da denir. Ses kanalının uzunluğu ortalama olarak erkeklerde 17 cm, kadınlarda 14 cm, ve çocuklarda 10 cm; burun boşluğu da erkeklerde 12 cm civarındadır. Bu sebeple doğal olarak erkeklerin sesleri kadın ve çocuklarınkinden daha kalındır. Ses yolu yumuşak damakta ikiye ayrılır. Üst taraf burun boşluğu, burun mukozası, burun kılları ve burun deliklerinden dış havaya açılır. Alt yol ise ağız boşluğu ve bildiğimiz ağız içindeki organ ve yapılardan dudaklar yoluyla dış havaya açılır. Ağız boşluğunun kesit alanı tamamen kapalı durumdan, 20 cm²'nin üstüne çıkabilir. Burun boşluğu da yumuşak damak tarafından tamamen kapatılabilir. Tam açık halde yumuşak damak açıklığı yetişkin bir erkekte 5 cm²'ye ulaşabilir (Proakis vd., 1993). Artikülasyon organlarının fiziksel boyutları ve beyinden alınan konuşma kumandalarının temporal değişimleri, ses işaretinin kişisel özelliklerini belirler.

Linguistik temel ses parçacığına fonem adı verilir, /a/, /b/, /c/ şeklinde gösterilir. Ses organları esas itibarı ile iki türlü ses üretir: a) Ünlü (voiced), b) ünsüz (unvoiced). Ünlü ve ünsüz sesler zaman ve frekans domeninde kendilerine has görünüm arzederler. Ünlüler zaman domeninde yüksek genlikli ve periyodik; frekans domeninde ise formantları olan bir görünüme sahiptir. Ünsüzler zaman domeninde düşük genlikli bir görünüme sahiptir; frekans domeninde ise enerji yoğunluğu yüksek frekanslara doğru kaymış, geniş bantlı ve



Şekil 2.3 Artikülasyon organları. (a) Ses tellerinde üretilen uyarım, zamanla değişen bir filtreye uygulanır. Velumun fonksiyonu nasal kaviteyi açıp kapamaktır. Artikülasyon organlarının fiziksel boyutları ve beyinden alınan konuşma kumandalarının temporal varyasyonları, ses işaretinin kişisel özelliklerini belirler (Online Phonetics Course: <http://www.unil.ch/ling/phonetique/apil-eng.html>); (b) Larinksin kesiti (The Voice Center at Eastern Virginia Medical School, http://www.voice-center.com/voice_mecha.html); (c) Ses tellerinin açılıp kapanmalarının üstten görünüşü (UPMC Health System, University of Pittsburgh Voice Center <http://www.upmc.edu/upmcvoice/normal.htm>)

formantsız bir görünüm sergilerler. Şekil 2.5' teki seslerin hepsi de ünlü seslerdir. Ünlü seslerde ses tellerinin titreşimi periyodiktir. Ünlü sesleri çıkarırken parmağınızı gırtlak

üzerinde tutarak ses tellerinin titreşmesini hissedebilirsiniz. Ses telleri, merkezi sinir sisteminin kumandasıyla, üretilecek sesin temel frekansına göre gerilir ve aralarındaki boşluğu kapatırlar. Kaburga, karın kasları ve diyaframın sıkışmasıyla akciğerlerdeki hava basıncı yükselir. Basınç belli bir değerin üzerine çıkınca ses tellerini zorlayarak açar; akciğer tarafındaki basınç düşer ve ses telleri tekrar kasılarak kapanır. Sonra hava basıncı tekrar yükselir ve bu süreç tekrarlanır. Şekil 2.3(c)'de ses tellerinin açılıp kapanmaları görülmektedir. Ses tellerinin bu açılıp kapanmaları hava basıncını periyodik olarak modüle eder. Bu titreşimlerin frekansı merkezi sinir sisteminin ses tellerine verdiği emirle belirlenen kasılma şiddetine bağlıdır. Cümle içindeki vurgu, heyecan; müzik icra etmek ses tellerinin farklı şiddette kasılmasıyla olur. Herkesin sesi, larinksinin fiziksel büyüklüğüne bağlı olarak alt ve üst temel frekans sınırları ile sınırlıdır. Erkeklerde temel frekans genellikle 50-250 Hz, kadınlarda ise 120-500 Hz arasındadır (Proakis vd., 1993). Fısıltı ile konuşurken ses telleri, titreşecek kadar değil; akciğerlerden gelen havada türbülans oluşturacak kadar kısmen kapanırlar (Şekil 2.5).

Şekil 2.4, sineradyografi teknikleri ile yapılan bir konuşma kayıdır. Buraya kayıttan bazı tipik örnekler seçilerek alınmıştır. Artikülasyon organlarının konuşma sırasında aldıkları konum açıkça görülebilmektedir. Mesela üst sıranın ilk karesi /w/, en alt sıranın ikinci karesi /p/ sesidir. /p/ sesi çıkarılırken dudakların iyice kapanıp ses yolunu bloke ettiği dikkati çekmektedir. Diğer organların konumları da dikkate değerdir: ses kanalının kesit alanının değişimi ve nasal kavitenin velum tarafından kapatılması gibi.

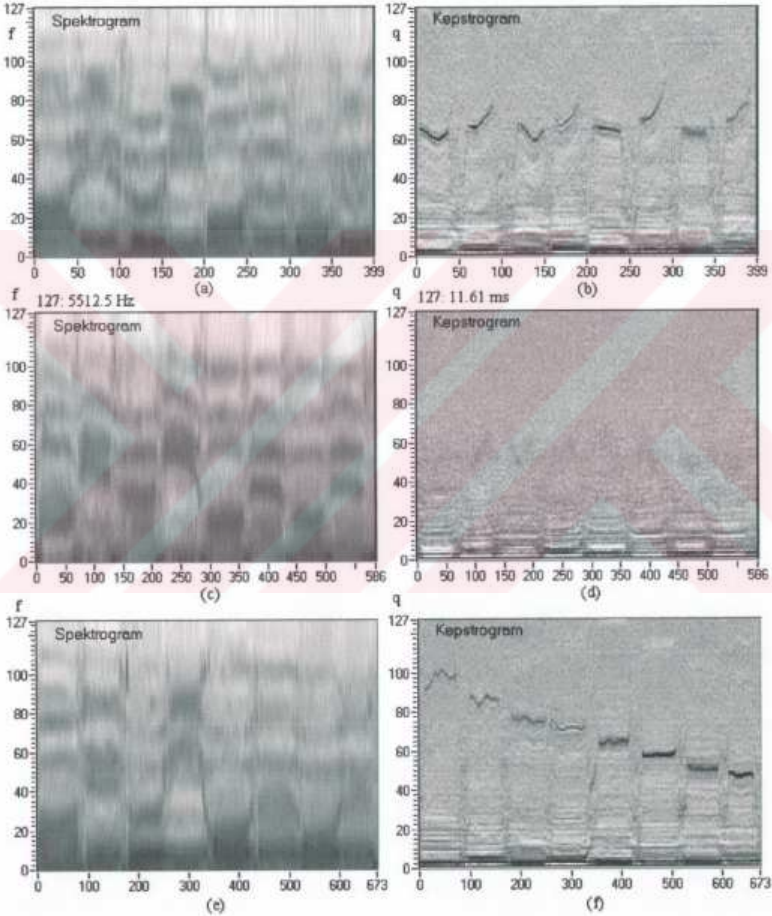
Ses telleri (kıvrımları) artikülasyon organlarına periyodik hava paketleri göndermektedir. Sistem teorisi açısından ses telleri bir uyarım kaynağı, artikülasyon organları da bu kaynakla uyarılan bir filtredir. Bu çalışmada uyarım kaynağına **glotal uyarım kaynağı** veya kısaca glotal uyarım, ses yolunun oluşturduğu filtreye de **artikülasyon filtresi** denilmiştir. Glotal uyarım, artikülasyon sırasında ses yolunun farklı noktalarda kesit değiştirmesiyle filtre edilir. Bazı sesler burun boşluğu kullanılarak üretilirler: /m/ ve /n/ seslerini çıkarırken yumuşak damak burun boşluğunu açar. İnsanların bu iki sesi söyleyememelerinden nezle olduklarını anlayabiliriz. Ses yolunun tıkanma noktasına bağlı olarak formantlar konuşma boyunca değişir.



Şekil 2.4 Konuşma sırasında konuşma organlarının konumu ve konuşma yolunun aldığı şekiller. */why did ken set the soggy net on top of his deck/* sözünün telaffuzu sırasında dil, diş, yumuşak damak ve dudakların konumu X-ışını enstantanelerinde açıkça görülebiliyor. Ayrıca yumuşak damak tarafından burun boşluğunun açılıp kapanması dikkat çekicidir. (Wenner-Gren Araştırma Laboratuvarı Nortull Hastanesi Stockholm, İsveç'te sineradyografik cihazlarla çekilen bir filmde. Dr. K. Stevens, ve Dr. S. Öhman, http://psyc.queensu.ca/faculty/munhall/x-ray/database/its_10_below.mov)

Ünsüz seslerde ses telleri titreşmez. Bu tür seslerde hava ses telleri tarafından engellenmeden üstteki artikülasyon organlarına ulaşır. Dilin, yumuşak damağın ve dudakların pozisyonu ve şekli ile oluşan dar kesitten geçerken, hava laminar akışını koruyamaz ve türbülans meydana gelir; örnek olarak */h/, /s/, /f/* seslerini çıkarırken bu daralma boğaz, diş, ve diş-dudak arasında olur. Bazen */k/, /t/, /p/, /ç/* seslerinde olduğu gibi hava yolu tamamen bloke edilir. Kısa bir müddet için ses kesilir. Oluşan engelin arkasındaki hava basıncı yükselir. Sonra engel birden kaldırılır. Patlama şeklinde */k/, /t/,*

/p/, /ç/ sesleri oluşur. Ünsüz sesler periyodik değildir. Dolayısıyla bir temel frekansları yoktur. Bu seslerin telaffuzu sırasında kişinin kendine özgü mahreçler kullanması, söz gelimi /s/ ve peltek /s/; /r/ sesinin telaffuzundaki güçlükler, /ll/, veya /y/ ile karıştırılması ses tanımda veya doğrulamada yararlanılabilecek ipuçları olabilir.

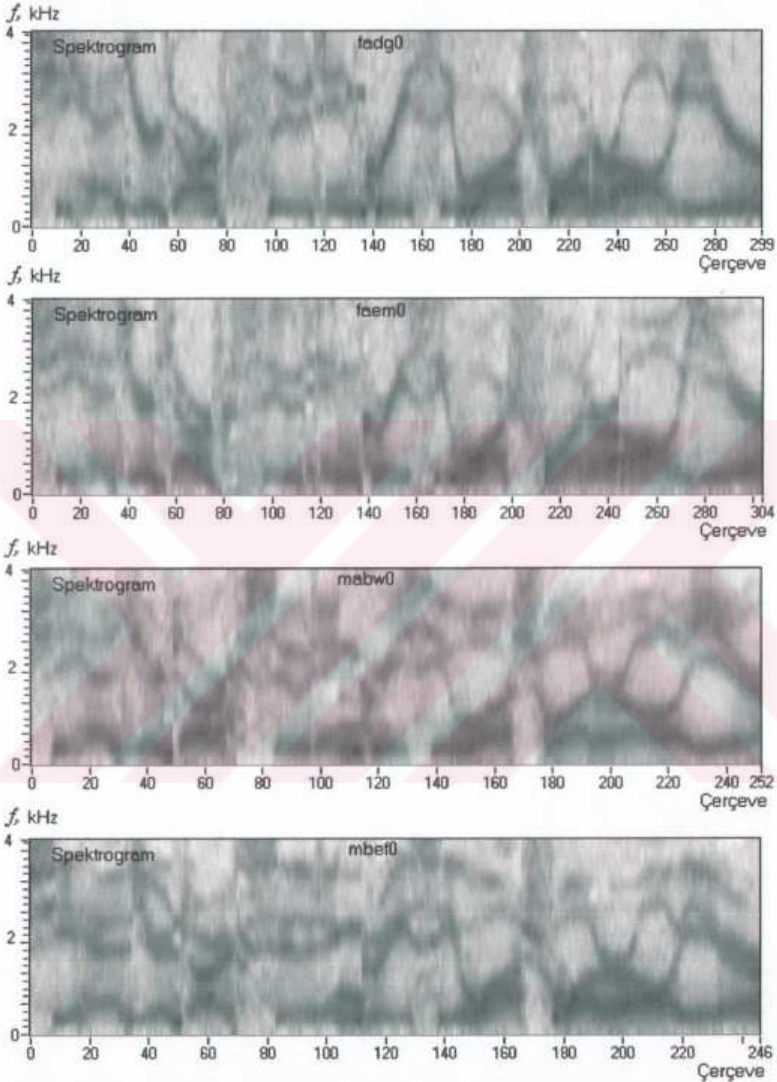


Şekil 2.5 /a,e,i,i,o,ö,u,u/ ses dizisinin üç ayrı şekilde söylenmesinden elde edilen spektrogram ve kepstrogramlar. Yatay eksen çerçeve sayısıdır. (a), (b) düz konuşma; (c), (d) fısıltı; (e) ve (f) do-re-mi-fa-sol-la-si-do notalarıyla okuma. Spektrogramların formant bilgilerinin hemen hemen aynı kaldığını, fakat temel periyod değerlerinin büyük bir farklılık arzettiğini görüyoruz. (Konuşan: Orhan Özhan)

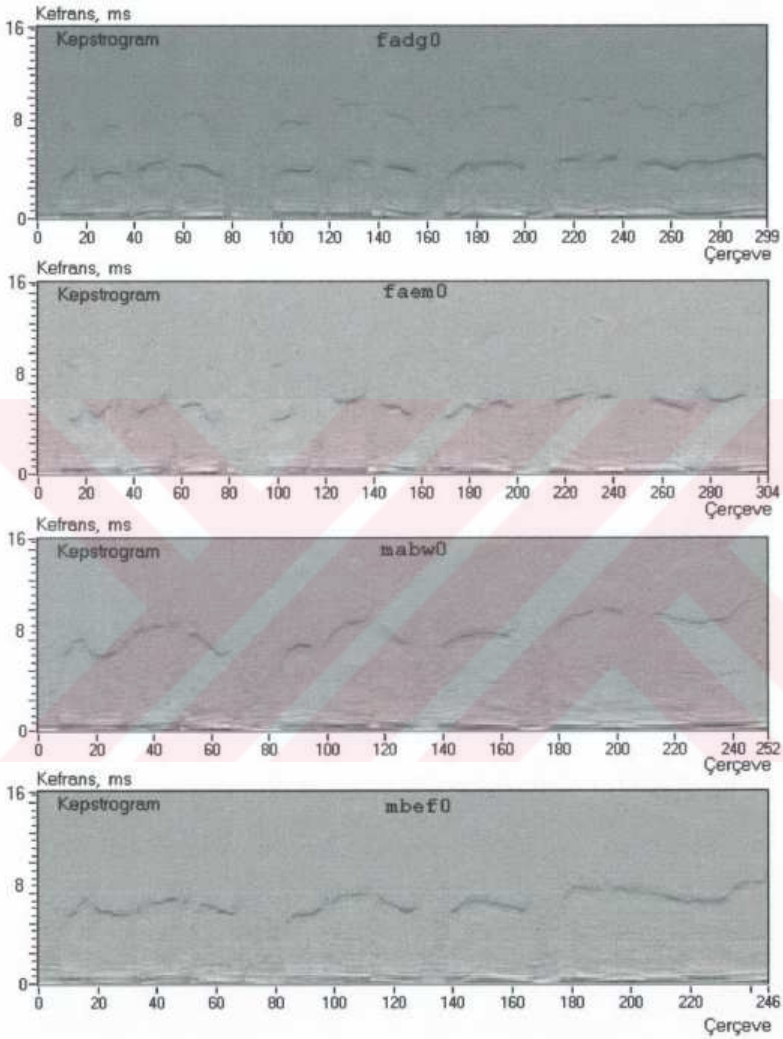
"Ses ve işitme organlarının limitasyonları sebebiyle konuşma işareti tipik olarak 7-8 kHz ile band genişliği ile sınırlandırılmıştır (Proakis vd., 1993)". Telefon endüstrisinde abone devreleri 300-3000 Hz arasına band sınırlıdır. Bu çalışmada kullanılan TIMIT konuşma tabanının örnekleme frekansı 8 kHz, dolayısıyla konuşmalar 4 kHz ile band sınırlıdır. TIMIT dışında kullandığımız konuşmalarda örnekleme frekansı 11025 Hz'dir. Ses istatistik olarak durağan olmayan bir işarettir. Bu özelliği sesle ilgili analiz ve tahmin yöntemlerini etkilemektedir. Bu konu hakkında daha fazla ayrıntıya girmeden şu kadarını söylemekle yetinebiliriz: çene kemiği ve dil gibi artikülasyon organlarının fiziki bir kütleyle sahip olmaları, bunları hareket ettiren kaslar, ve kaslara ulaşan sinirlerin polarizasyon ve depolarizasyon süreçlerinin bant-sınırlı olmaları ve artikülasyon organlarının aynı zamanda solunum ve yeme-içme fonksiyonları için de kullanılmaları sebebiyle bir fonemden başka bir foneme geçiş aniden değil tedricen olmaktadır. Bu gerçeği göz önünde bulundurarak, ses işaretini kısa zaman aralıklarında durağan kabul edebiliriz. Pratikte bu aralık, 10 ms - 40 ms arasında seçilir. Yetişkin bir erkeğin sesinin perdesini esas alırsak, analiz için 3 - 4 temel periyodu kapsayan bir aralık kullanılabilir. Şekil 2.1'deki erkek sesinin temel periyodu 1/150 s'dir. Dört periyod 26.6 ms değerindedir. TIMIT konuşma tabanı ile olan çalışmamızda ise 32 ms'lik aralık seçilmiştir.

Glotal uyarım ve artikülasyon filtresinin spektral bileşenlerinin zaman içinde gösterdikleri değişim, sesin sahibiyle ilgili ipuçları verir. Ses sahibi eğitim, heyecan, bölgesel vb. faktörlerle anatomik özelliklerini zaman içinde kendine özgü bir biçimde kullanır. Aynı konuşma değişik kişilerin "diliyle" farklı farklı söylenir. Şekil 2.6 ve Şekil 2.7 dört ayrı TIMIT konuşmacısının söylediği *sal* cümlesinin spektrogram ve kepstrogramlarını göstermektedir. Kepstrogramla konuşmanın keptrum bileşenlerinin zaman içindeki davranışlarını izleyebiliriz.

Şekil 2.6'da koyu izler formantlardır. 0 Hz'e en yakın formant F_1 , 8 kHz'e doğru F_2 , F_3 ... formantları görülmektedir. Kadın konuşmacılar *fadgθ* ve *faemθ*'da formant frekanslarının erkek konuşmacılar *mabwθ* ve *mbefθ*'in formant frekanslarından daha yüksek olmaları dikkati çekmektedir. Bu daha önce açıkladığımız üzere kadınlarda ses yolunun erkeklerden daha kısa olmasından kaynaklanmaktadır. Konuşmada ünsüz kısımlar F_1 formantının bulunmadığı kısımlardır. Örnek olarak spektrogramların baş tarafında /sh/ fonemini gösterebiliriz.



Şekil 2.6 Kişilerdeki söyleyiş farklarının spektrograma yansması. İki kadın ve iki erkek TIMIT konuşmacısı tarafından söylenen *"she had your dark suit in greasy wash water all year"* cümlesinin spektrogramında konuşma hızları, fazla ve eksik formantlar, formant değerlerindeki farklılıklar ve farklı formant izleri konuşanların farklı artikülasyon anatomilerine sahip olmalarından ve bunları zaman içinde farklı biçimde kullanmalarından kaynaklanmaktadır. F_1 , F_2 , F_3 , F_4 formantları ve ünsüz konuşma parçaları kolayca belli olmaktadır.



Şekil 2.7 Kişilerdeki söyleyiş farklarının kepstrograma yansımaları. İki kadın ve iki erkek TIMIT konuşmacısı tarafından söylenen "she had your dark suit in greasy wash water all year" cümlesinin kepstrogramında görülen koyu izler temel periyod T_0 'ı göstermektedir. İzlerin koptuğu yerlerde ünsüz sesler bulunmaktadır. Farklı T_0 izleri konuşanların farklı larinks anatomilerine sahip olmalarından ve bunları zaman içinde farklı biçimde kullanmalarından kaynaklanmaktadır.

Şekil 2.7'de aynı konuşmacıların söylediği /she had your dark suit in greasy wash water all year/ cümlesinin kepstrogramları görülmüyor. Şekilde koyu izler temel periyot T_0 'ın takip ettiği izlerdir. Bu izlerin koptuğu yerlerde ünsüz sesler bulunmaktadır. Ses kıvrımlarının titreşmediği bu yerlerde kepstrogram analizi glotal uyarım bulamamıştır. Kadın konuşmacılar *fadg0* ve *faem0*'da T_0 'ın erkek konuşmacılar *mabw0* ve *mbef0*'den daha küçük olmaları hemen göze çarpmaktadır. Yine *mabw0*'ın konuşmasında bir müzikallik varken *fadg0* konuşmasını daha monoton olarak sürdürmektedir.

Şekil 2.6 ve Şekil 2.7'den anlaşılacağı gibi T_0 (veya F_0) ve spektral parametrelerin zaman içindeki izleri, konuşmadan konuşanı tanımak için kullanılabilir. Bu parametrelerin temporal değişimleri konuşma tanımadaki giderilmesi gereken bir mahzur iken, konuşanı tanımadaki son derece önemli ve yardımcı özniteliklerdir. Konuşmayı yapan artikülasyon organlarıdır; larinks değildir. Larinks konuşmaya sadece "müzikalite" katar. Bu sebepten dolayı konuşma tanımadaki T_0 (veya F_0) istenmeyen unsurlardır. Konuşmacı tanımadaki ise durum bunun aksinedir. Kişinin konuşmasındaki müzikalite veya monotonluk onun kimliğini ele veren önemli bir bilgi kaynağıdır.

2.3 Doğrusal Konuşma Modeli ve Homomorfik Analiz

Rabiner ve Schafer (1978) konuşma işareti için bir doğrusal konuşma modeli önerdiler. Model Şekil 2.8'de amaçlarımıza uyarlanmış olarak yeniden çizilmiştir. Bu modelimiz bir işaret kaynağı ve bir filtreden oluşur. İşaret kaynağı ya periyodik bir darbe kaynağı veya bir gürültü kaynağıdır. Hem kaynak hem de filtre MSS'nin kontrolü altındadır.

$H(w, t)$ artikülasyon filtresinin basınç transfer fonksiyonudur. $x(t)$ girişi glotisten gelen periyodik uyarım veya gürültü kaynağıdır. /m/, /n/ sesleri çıkarılırken velum ağız ve burun boşluklarının meydana getirdiği $H_1(w, t)$ ve $H_2(w, t)$ filtrelerini birbirlerine paralel bağlar [Şekil 2.8(b)]. Türkçe konuşmalarda bu seslerin dışında $H(w, t)$ transfer fonksiyonu $H_1(w, t)$ 'den ibarettir. $H(w, t)$ notasyonundaki "t" parametresi bu filtrelerin zamanla değişen filtreler olduğunu vurgulamak içindir. Gerçekten Şekil 2.8(a)'da görüldüğü gibi MSS (Merkezi Sinir Sistemi) $H_1(w, t)$ 'nin karakteristiklerini konuşulan söze göre değiştirir.

Bu model uzun yıllar ses uygulamalarında kullanılan bir modeldir. Son zamanlarda bu model ciddi olarak sorgulanmakta ve yeni modeller aranmaktadır. İtirazlar filtrenin ve

kaynağın doğrusallığı, kaynakla filtrenin birbirinden bağımsızlıkları üzerinde yoğunlaşmaktadır (Fackrell, 1996). Buna rağmen bu model hala yaygın bir şekilde kullanılmaya devam etmektedir, ve bu çalışmanın amaçları için yeterlidir.

Bir $y(t)$ konuşma işareti içinde, artikülasyon ve glotal öz nitelikler konvolusyon halinde bulunur. Dolayısıyla konuşmacı tanıma için $x(t)$ ve $H(w,t)$ dekonvolusyon yolu ile $y(t)$ işaretinden elde edilmelidir. Bunu homomorfik analiz kullanarak yapabiliriz. Homomorfik analize cepstrum analizi de denir. $H(w,t)$ 'yi durağan kabul ettiğimiz analiz çerçevelerinde $H(w,t) = H(w)$ kabul edilebilir. Buna göre frekans domeninde

$$Y(w) = H(w)X(w) \quad (2.1)$$

yazılabilir.

Ünlü seslerde $x(t)$ periyodik hava paketleri olduğu için Fourier serisi açılımı

$$x(t) = \sum_{n=0}^{\infty} A_n \sin\left(\frac{2\pi n}{T_0} t + \theta_n\right) \quad (2.2)$$

ifadesiyle verilen harmonikler içerir. $y(t)$, artikülasyon filtresinin bu harmoniklere verdiği cevapların süperpozisyonudur. $y(t)$ konuşma işaretinin n -nci harmoniğinin faz açısı, (2.2)'deki θ_n 'in üzerine artikülasyon filtresinin faz açısı $\angle H(n\omega_0)$ eklenerek bulunur

($\omega_0 = 2\pi/T_0$):

$$y(t) = \sum_{n=0}^{\infty} A_n |H(n\omega_0)| \sin[n\omega_0 t + \theta_n + \angle H(n\omega_0)] \quad (2.3)$$

olur. Yani ağızdan çıkan konuşma işaretinde temel frekansın harmonikleri bulunur ve her harmoniğin fazı $\theta_n + \angle H(n\omega_0)$ 'a eşittir. Kulak, iki sinüzoidal işaretin toplamı olarak verilen

$$u(t) = A_1 \sin \omega_1 t + A_2 \sin (\omega_2 t + \phi) \quad (2.4)$$

işaretindeki ϕ faz açısına duyarlı değildir; sadece A_1 ve A_2 genliklerini duyar. (2.1) denklemi

$$|Y(w)| e^{-j\theta} = |H(w)||X(w)| e^{-j(\theta_k + \theta_s)} \quad (2.5)$$

şeklinde yazılırsa kulak $e^{-j(\theta_h + \theta_e)}$ ile verilen faz açısını duymayacak demektir. O zaman (2.1) ve (2.5)'in yalnız genlikleri kullanılarak

$$|Y(w)| = |H(w)||X(w)| \quad (2.6)$$

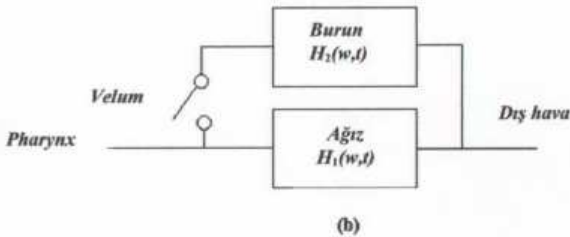
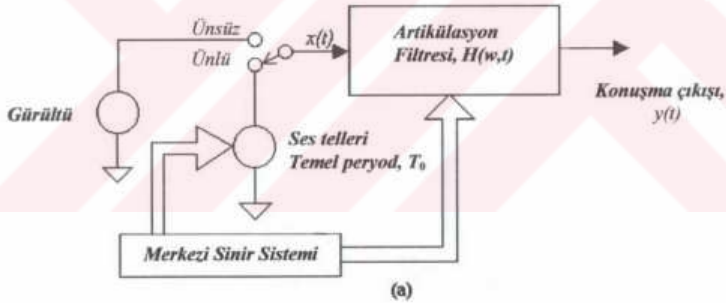
elde edilir. Bu ifadenin logaritması

$$\log |Y(w)| = \log |H(w)| + \log |X(w)| \quad (2.7)$$

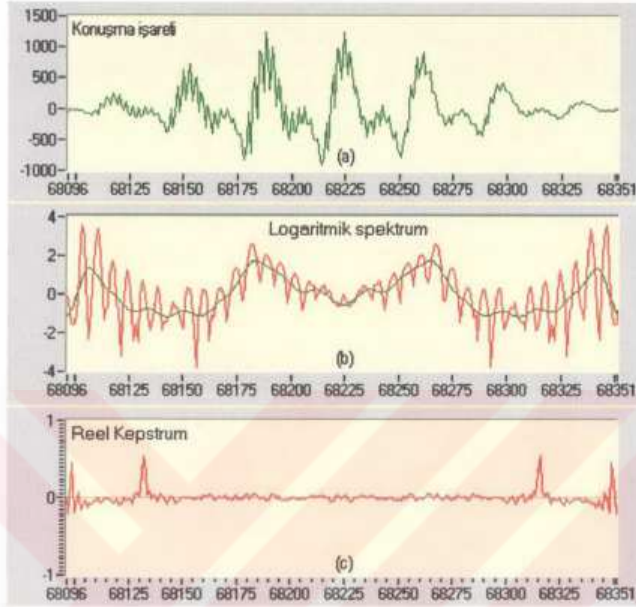
$H(w)$ ile $X(w)$ 'yı toplamsal hale dönüştürür. (2.7) haline getirilen konuşma işaretinden glotal uyarım ve artikülasyon filtresi ayrı ayrı elde edilebilir. (2.7) üzerine ters Fourier dönüşümü uygulanırsa

$$\mathcal{F}^{-1}[\log |Y(w)|] = \mathcal{F}^{-1}[\log |H(w)|] + \mathcal{F}^{-1}[\log |X(w)|] \quad (2.8)$$

olur. Bu domen kepstrum domeni olarak adlandırılır. Bu analiz de homomorfik analiz veya kepstrum analizi olarak tanınmaktadır (Bogert vd., 1963). Kepstrum domeninde glotal uyarım ve artikülasyon parametreleri birbirinden ayrılmıştır.



Şekil 2.8 Konuşma modeli. (a) Zamanla değişen artikülasyon filtresi, periyodik bir darbe kaynağı veya gürültü kaynağı ile sürülmektedir. Merkezi Sinir Sisteminin (MSS) kontrolü altında, darbe kaynağının periyodu, genliği ve artikülasyon filtresinin kutup yerleri zamana bağlı olarak değişir; (b) Ağız ve burun boşluğu filtreleri velum ile birbirlerine paralel bağlanır.



Şekil 2.9 Homomorfik analiz. (a) Zaman domeninde konuşma işareti; (b) Logaritmik spektrumda glotal uyarım ve artikülasyon filtresi toplamsal halde; (c) Kepstrum domeninde artikülasyon ve glotal uyarım ayrılmış durumda. Yatay eksen ayrık zamanda örnek sayıları, dikey eksenler de bağlı genliklerdir.

Örnek olarak Şekil 2.9'da denklem (2.6), (2.7) ve (2.8)'deki işlemler bir ses işareti üzerinde gösterilmiştir. Şekil 2.9(a) zaman domeninde bir konuşma işareti çerçevesidir. Şekil 2.9(b)'de yeşil eğri artikülasyon, kırmızı eğri ise "artikülasyon + glotal uyarımdır". Şekil 2.9(c) kepstrum domenidir. 68096'dan 68125'e kadar artikülasyon, 68125'den yukarıya glotal uyarımdır. 68130 civarındaki tepe glotal uyarımın genliği, tepenin pozisyonu ise T_0 'dır. Şekil 2.7'deki kepstrogramlar da homomorfik analizle elde edilmiştir.

Larinksten yukarıya doğru devam eden ses kanalı (artikülasyon kanalı), geleneksel olarak akustik borularla modellenmiştir. Bu model [Şekil 2.8(a)], uç uca eklenmiş kesit alanları değişen fakat uzunlukları aynı borulardan oluşur ve şu kabullere dayanır:

- Boruların duvarlarında enerji kaybı yoktur,
- En tiz sesin dalga boyu boruların boyundan daha uzundur,
- Ses dalgası borular içinde düzlemsel dalga olarak yayılır.

Akustik boru modelinden artikülasyon filtresinin davranışı kestirilebilir; ayrıca bu modelin parametreleri artikülasyon öznelikleri olarak kullanılabilir. Akustik boru modelinin ayrık zaman elektriksel karşılığı Şekil 2.10(b)'de gösterilmiştir. Farklı kesit alanları sebebiyle, bölümler arasında yansımalar meydana gelir. $x(i)$, $x(n)$, $y(i)$ ve $y(n)$ ses telleri uyarımı ve konuşma çıkışının sürekli-zaman ve ayrık-zaman notasyonlarıdır. Model "P" bölümden oluşmaktadır. A_i , "i" nci bölümün kesit alanı, K_i , "i" nci bölüm ile "i-1" nci bölüm arasındaki (akustik) yansıma katsayısı; " z^{-1} " ayrık-zaman birim gecikme operatörüdür. Atal ve Hanauer (1971), Markel ve Gray (1976) akustik modelden Şekil 2.10(b)'deki ayrık-zaman tüm-kutuplu latis (lattice) filtre modelini çıkarmışlardır. A_i ile K_i arasında

$$K_i = \frac{A_{i-1} - A_i}{A_{i-1} + A_i} \quad (2.9)$$

ve

$$\frac{A_{i-1}}{A_i} = \frac{1 + K_i}{1 - K_i} \quad (2.10)$$

bağıntıları vardır. Filtrenin transfer fonksiyonu ise

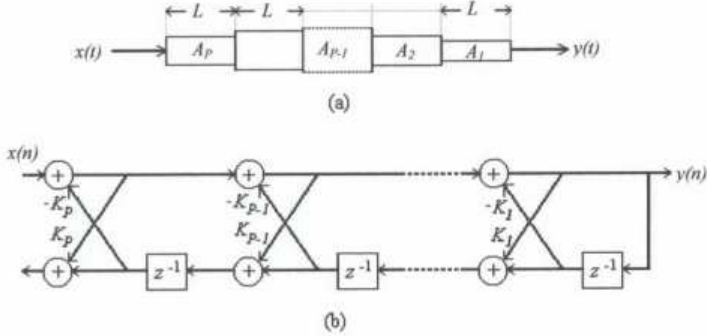
$$H(z) = \frac{1}{1 + \sum_{k=1}^p a(k)z^{-k}} \quad (2.11)$$

ifadesiyle verilir. Buradaki polinom katsayıları $a(k)$ 'ya tahmin katsayıları da denir. Yansıma katsayıları K_i ile $a(k)$ arasında rekursiv Levinson-Durbin bağıntısı vardır. (2.4)'deki transfer fonksiyonunu çarpanlarına ayırarak

$$H(z) = G \cdot \prod_{k=1}^p \frac{1}{z^{-1} - p_k^{-1}} \quad (2.12)$$

şeklinde yazabiliriz. Paydadaki p_k transfer fonksiyonunun k -nci kutbu, G de bir kazanç faktörüdür. Frekans bir kutbun yakınında iken rezonans tepeleri gözlenir. Şekil 2.2'de gösterilen formantlar bu davranışın ürünüdürler. Artikülasyon filtresi kararlı (stable) olduğu için, p_i kutupları z -düzleminde birim daire içinde bulunurlar. Kararlılık için filtre parametrelerinde

$$0 < |K_i| < 1 \quad (2.13)$$



Şekil 2.10 Artikülasyon filtresi modeli. (a) Akustik borular modeli. Konuşma işaretindeki rezonans frekansları (formantlar), değişik yarıçaplı boruların doğal rezonanslarına karşılık gelir; (b) Akustik modelin elektriksel eşdeğeri.

şartı aranır. $|K_i| = 1$ filtrenin osilasyon halidir ki, konuşma (artikülasyon) için böyle bir durum söz konusu olamaz. K_i ne kadar büyükse, rezonansın Q değeri o kadar yüksek olur. Kompleks kutup çiftlerinin sayısı, konuşma işaretindeki formant sayısına eşittir. $p_i = R_i e^{j\omega_i}$ şeklinde gösterilirse, frekans spektrumunda $\omega = \omega_i$ frekanslarında formantlar görürüz. Farklı boğaz anatomilerinden dolayı, farklı kişilerin aynı söze karşılık gelen akustik model parametreleri farklılık gösterir. Bu da tüm-kutup filtremizin kutup değerlerini, yani formant değerlerini farklı kılar. Fakat farklı kişiler tarafından söylenen aynı sözün aynı anlama gelebilmesi için formant genlik oranlarının ve formant frekans oranlarının sabit kalması gerekir. Bu oranlar linguistik içeriği belirler; R_i ve ω_i 'nin aldığı değerler ise kişiliği belirler.

2.4 SOM Yapay Sinir Ağı

Yapay sinir ağları fikri insan beyninin çalışmasından esinlenmiştir. Beynin çalışması üzerindeki öncü çalışmalar 1911'e kadar uzanır. Ramón y Cajal (1911) beynin nöron denilen yapısal birimlerden oluştuğu fikrini ortaya attı. Nöronlar, teknolojinin şu anda ulaştığı nokta itibarıyla silisyum lojik devrelerinden 10^6 defa daha yavaştır. Örnek olarak Texas Instruments (1998) tarafından üretilen ALB16244 BiCMOS kapısının hızı

maksimum 2 ns'dir. Bir nöron için ise iletim hızı 10^{-3} s mertebelerindedir. Fakat milyarlarla ifade edilen sayıları ve bunun binlerce katı kadar ara bağlantıları sayesinde silisyum çiplerine göre yavaş çalışmalarını kat kat telafi ederler. Sheperd ve Koch (1990) kortekste 10 milyar nöron olduğunu, nöronlar arasında da 60 trilyon sinaps (bağlantı) bulunduğunu tahmin etmektedir. Bu kadar çok sayıda nöron ve sinapslar bir paralel hesaplama mimarisi teşkil eder. Bu paralellik sayesinde en gelişmiş bilgisayarları kısındıracak bir performans sergilerler. Yine beynin enerji verimliliği 10^{-16} joule/işlem/s olarak bildirilmiştir; bugün en iyi bilgisayarlarda bu rakam 10^{-6} joule/işlem/s dir (Faggin, 1991).

Yapay sinir ağları çağının McCulloch ve Pitts ile başladığı söylenebilir (Haykin, 1994). McCulloch ve Pitts (1943) sinir ağlarının mantıksal matematikini ortaya koyan bir makale yazdılar. Bu makale o zaman büyük bir heyecan ve ilgi uyandırdı. Von Neuman bile ENIAC bilgisayarının arkasından yaptığı EDVAC bilgisayarında, McCulloch-Pitts modelinin ideal sinir elemanlarından çıkardığı "anahtar-gecikme" elemanlarını kullandı. Bundan sonra 1949 yılında Donald Hebb *The Organization of Behavior* isimli kitabını çıkardı. Hebb (1949) ilk defa sinaptik değişiklikler için fizyolojik öğrenme prensiplerini ortaya koydu, organizma yeni fonksiyonel görevler öğrenirken beynin sinir bağlantılarında yenilenmeler olduğunu ve böyle değişikliklerin yeni oluşumlara yol açtığını önerdi:

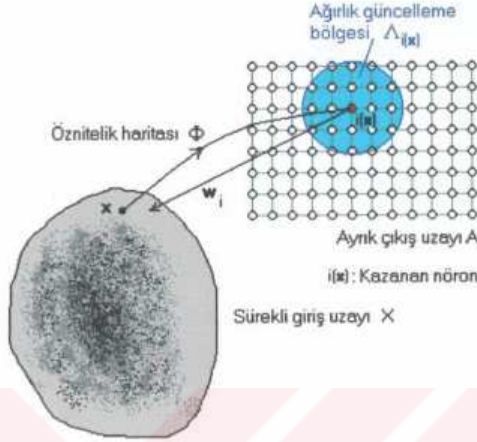
"Bir A hücresinin aksonu bir B hücrelerini uyaracak kadar (ona) yakınsa ve tekrar tekrar, tutarlı bir şekilde B hücrelerini uyarmada rol alırsa, hücrelerin birinde veya her ikisinde A'nın B'yi uyaran hücreler içinde etkinliğini arttıracak bir büyüme süreci veya metabolik değişimler meydana gelir (The Organization of Behavior s.62: Hebb, 1949)."

Hebb'in kitabı psikologlar üzerinde büyük bir etki gösterdi, fakat mühendislik camiasında maalesef hemen tanınmadı (Haykin, 1994). Hebb'in getirdiği öğrenme kuralı çok önemli bir kilometre taşıdır. Bu çalışmamızda kullanılan SOM algoritması Hebb öğrenme kuralının yetersizliklerinin kaldırılması ile geliştirilen "rekabet öğrenme (competitive learning) kuralına" dayanır. Rosenblatt (1958) perceptron (algılayıcı) çalışmasında örüntü tanımayla yönelik yeni bir yaklaşım getirdi. Daha sonra kendisinin ortaya koyduğu *perceptron yakınsama teoremini* ispatladı (Rosenblatt, 1960). Algılayıcıları kullanarak rakam sınıflayan sinir ağı yaptı. Doğrusal olarak ayrılabilir sınıfları ayırmada başarılı olan algılayıcılar o kadar ilgi uyandırdı ki o zamanlar her problemin yapay sinir ağları ile çözülebileceği düşüncesi topluma hakim oldu. Fakat Minsky ve Pappert (1969) algılayıcıların çözemedikleri problemlerin varlığını gayet açık ve seçik olarak ortaya koyan

bir kitap yazdılar . Mesela tek katmanlı bir algılayıcı AND, OR gibi lojik işlemleri ile ayrılabilen verileri sınıflayabildiği halde, XOR lojik işlemi ile ayrılan verileri sınıflayamıyordu. Minsky ve Papert argümanlarını genelleyerek, tek katmanlı algılayıcılarla çözilemeyen problemlere çok katmanlı algılayıcılarla çözüm aramanın boşuna olduğunu bildirdiler. Bu çalışma yapay sinir ağları çalışmaları üzerinde soğuk düş etkisi yaptı. Bir kaç araştırmacı dışında hemen herkes bu sahayı terketti; ajanslar bu çalışmaları finanse etmekten vaz geçtiler. 70'li yıllar yapay sinir ağları çalışmalarının uyuklama yıllarıdır. Buna rağmen 70'li yıllarda rekabet öğrenmesine dayanan kendi kendini düzenleyen haritalar üzerinde önemli çalışmalar ortaya çıktı. von der Malsburg (1973) bilgisayar simülasyonlarıyla ilk defa kendi kendine düzenlenmeyi gösterdi. Beyinde topolojik olarak düzenlenmiş haritalardan yola çıkarak Willshaw ve von der Malsburg (1976) kendi kendini düzenleyen haritaların oluşumları üzerine ilk makaleyi yazdılar. Fukushima da (1975) kendi kendini düzenleme üzerinde ilk çalışanlardandır. Kohonen (1982) bir veya iki boyutlu sinir ağları ile kendi kendini düzenlemeyi anlattı. Kohonen'in modeli Willshaw ve von der Malsburg modelinden bazı noktalarda ayrılıyordu. Bu çalışmada Hebb öğrenmesine nonlinear bir unutma faktörü katarak, Hebb öğrenmesinin sinaptik ağırlıklarda sebep olduğu eksponansiyel büyümeyi ve dolayısıyla doymayı önlemiş oldu (Kohonen, 1988a). Çeşitli araştırmacılar duyu organlarının algıladığı bilgilerin beyinde topolojik olarak düzenlenmiş hesap haritaları halinde depolandıklarını bulmuşlardır. Kaas vd. (1983) dokunma bilgilerinin, Hubel ve Wiesel (1962, 1977) görme bilgilerinin, Suga (1985) da ses bilgilerinin kortekste değişik bölgelerde topolojik olarak düzenlenerek depolandığını bildirmişlerdir. Beyindeki topografik haritaların gelişiminde rekabet öğrenmesinin rol oynadığına dair önemli kanıtlar bulunmakatadır (Durbin vd., 1989). Beyindeki topografik haritalarla ilgili olarak Kohonen (1990)

"Bir çıkış nöronunun uzaysal konumu giriş bilgilerinin özel bir domenine veya belli bir özneliliğe tekabül eder."

demiştir. Kohonen modeli nörobiyolojik olayların ayrıntılarını açıklamaktan çok beyinde oluşan hesap haritalarının lüzumlu özelliklerini yakalamayı, bunu da kolay bir hesaplama ile yapmayı hedeflemiştir. Modelin nörobiyolojik uygulanabilirliği (Kohonen, 1993)'te tartışılmıştır. Başka bir açıdan bakışla kendi kendini düzenleme bir vektör kuantalama işlemidir. Bir kodlayıcı ve bir kod çözücünden meydana gelen bir modelle vektör kuantalama yaklaşımıyla Kohonen modelini çıkarmak mümkündür (Luttrell, 1989, 1991).



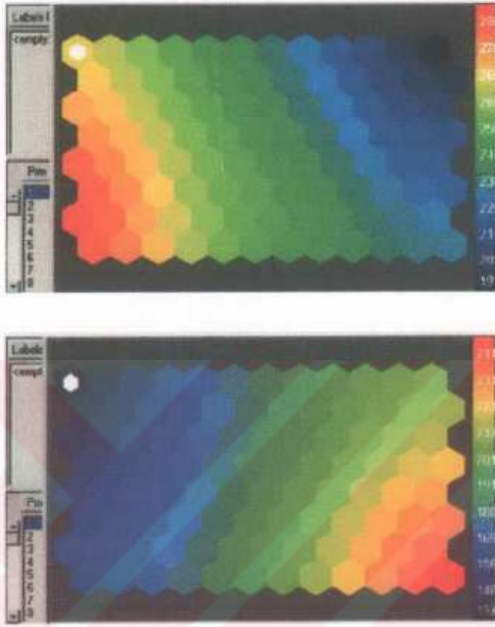
Şekil 2.11 Öznitelik haritası. Sürekli giriş uzayı X , ayrık çıkış uzayı A 'ya haritalanır (Simon Haykin: *Neural Networks – A Comprehensive Foundation*, 1994'ten uyarlanmıştır).

Bu tezin motivasyonu, Kohonen modelini kullanarak beyindeki öznitelik haritalarına veya olasılık yoğunluk dağılımlarına benzer konuşma haritaları oluşturup, bilinmeyen konuşmayı bu haritalarla eşleştirerek konuşmacıyı teşhis etmektir.

SOM yapay sinir ağı sürekli bir giriş uzayı X 'in ayrık bir çıkış uzayı A 'ya algoritmik olarak haritalanmış şeklidir. Bu haritalar bir süpervizöre ihtiyaç duymadan kendi kendilerini düzenledikleri için Self Organizing Map (SOM) veya Self Organizing Feature Map (SOFM) adlarıyla tanınır.

SOM algoritması veya Kohonen algoritması adı verilen haritalama işlemi sonunda 1- veya 2-boyutlu olarak oluşan harita sürekli giriş uzayının bir temsidir. Şekil 2.11'de 2-boyutlu bir harita gösterilmiştir. " X uzayındaki bir x vektörü öznitelik haritası Φ ile A uzayındaki bir düğüme haritalanır." A uzayındaki bu düğüme **kazanan nöron** denir ve $i(x)$ sembolü ile gösterilir.

A uzayındaki kazanan nöronun, x vektörü ile aynı boyutta bir w ağırlık vektörü vardır. w ağırlıklarına konunun fizyolojik tarihçesi sebebiyle **sinaptik ağırlık vektörleri** de denir. $x \in X$ yaklaşık olarak $w \in A$ ile temsil edilir. Aşağıda verilen üç özellik sebebiyle sürekli X



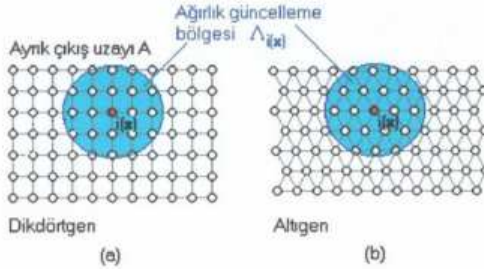
Şekil 2.12 İki konuşmacının temel frekanslarının SOM haritaları. (a) *fadg0*, (b) *faem0* (SOM haritalarını göstermek için NENET programı kullanılmıştır).

uzayı yerine ayrık A uzayını kullanmak isteriz (Haykin, 1994):

ÖZELLİK 1. Öznitelik haritası giriş uzayının bir yaklaşımıdır (approximation)

A çıkış uzayında $\{w_j | j = 1, 2, \dots, N\}$ sinaptik ağırlık vektörleri kümesi ile temsil edilen SOM X giriş uzayı için iyi bir yaklaşım oluşturur. Bu algoritmanın ana hedefi çok sayıda $x \in X$ giriş vektörünü daha az sayıda $w_j \in A$ vektörleri bularak saklamak veya temsil etmektir. Örnek olarak Şekil 2.11'deki giriş uzayı bir konuşmacıya ait konuşma öznitelik vektörleri olsun. Giriş uzayının vektör sayısının konuşmacıdan konuşmacıya değişeceği açıktır. Şekildeki 12×8 büyüklüğündeki SOM bu konuşma uzayının bir yaklaşımıdır.

ÖZELLİK 2. Öznitelik haritasında topolojik düzenlenme



Şekil 2.13 SOM yapay sinir ağı komşuluk topolojileri. (a) Dikdörtgen, (b) Altgen.

SOM algoritması ile hesaplanmış öznitelik haritası Φ , topolojik olarak düzene girer. Sonuçta haritadaki bir nöronun uzaysal konumu giriş uzayındaki bir öznitelige karşılık olur. Aşağıda görüleceği gibi bu, kazanan nöron ve çevresindeki nöronların giriş vektörüne doğru hareket ettirilmelerinin bir sonucudur. Bu sebeple Hertz vd (1991) öznitelik haritası Φ 'nin A çıkış uzayında 1- veya 2-boyutlu topolojiye sahip, düğümlerinin ağırlıkları giriş uzayının koordinatları olan esnek bir ağı olarak düşünülebileceğini söylemişlerdir. Böylece algoritmanın genel amacı:

"Öznitelik haritası Φ , $\mathbf{x} \in X$ giriş vektörlerinin önemli özelliklerini sadakatle koruyacak bir şekilde X giriş uzayını, sinaptik ağırlık vektörü şeklinde $\mathbf{w}_j$ prototipleri ile yaklaşık olarak temsil etmektir (Haykin, 1994)."

ÖZELLİK 3. Yoğunluk bakımından denklik

Öznitelik haritası Φ giriş uzayının istatistik değişimlerini yansıtır. $\mathbf{x} \in X$ vektörlerinin giriş uzayından büyük bir olasılıkla çekildiği bölgeler A çıkış uzayında daha büyük domenlere dönüşür; dolayısıyla $\mathbf{x}$ 'in daha düşük olasılıkla çekildiği bölgelere göre çıkış uzayına daha yüksek çözünürlükle haritalanmış olurlar (Haykin, 1994).

Şekil 2.12 iki konuşmacıya ait öznitelik haritalarında F_0 'ın dağılımını göstermektedir. Yukarıda sayılan özellikler bilhassa SOM'lardaki düzen dikkat çekicidir. Haritanın sağ tarafında görülen skala iki dağılımın konuşmacı özelliklerini nasıl yansıttığına dair bir fikir vermektedir. Şekildeki her altgen hücre bir nöron veya SOM haritasının bir düğümüdür. Harita altgen topolojiye sahiptir. Her nöron bir sinaptik ağırlık vektörüne sahiptir. Şekil 2.12'de bu vektörlerin F_0 katmanı gösterilmiştir. Diğer katmanlar formant değerleri,

formant genişlikleri, veya kepstrum katsayıları gibi öznitelik vektörünün diğer bileşenleri olabilir.

Haritalarda dikdörtgen veya altıgen komşuluk topolojileri kullanılabilir. Görsellik açısından altıgen topoloji daha çok tercih edilir. Şekil 2.13 her iki komşuluk topolojisini göstermektedir. Bu çalışmada Şekil 2.12'deki gibi altıgen topoloji kullanılmıştır. Giriş uzayından gelen vektörler paralel olarak bütün nöronlara bağlıdır. SOM rekabet öğrenme kuralına dayanarak öğrenir. Rekabet öğrenme kuralında bir $\mathbf{x}$ giriş vektörü verildiğinde haritada nöronlar arasında bir rekabet (yarış) meydana gelir. Her nöron $\mathbf{x}$ ile kendi sinaptik ağırlık vektörünü karşılaştırır. Bu rekabetten sadece $\mathbf{x}$ 'e en yakın olan nöron galip çıkar. Galip gelen nöron Şekil 2.11 ve Şekil 2.13'te $i(\mathbf{x})$ ile gösterilmiştir. $i(\mathbf{x})$ 'in çevresinde $\Lambda_{i(\mathbf{x})}$ ile gösterilen bölgedeki bütün nöronların ağırlıkları öğrenme miktarı kadar $\mathbf{x}$ 'e doğru hareket ettirilir (Şekil 2.4). Bu işlem giriş uzayındaki bütün vektörler için defalarca tekrarlanır. $\Lambda_{i(\mathbf{x})}$ bölgesi ötelemeler arttıkça yavaş yavaş küçültülür. Bu ötelemelerin sonunda öznitelik haritası Φ oluşur ve yukarıda sayılan üç özelliğe sahip olur.

Öznitelik haritalarını elde etmek için kullanılan SOM algoritmasının sonuç şeklini - fizyolojik veya istatistik çıkarılımını atlayarak - aşağıdaki adımlarla verebiliriz (Kohonen 1988c, 1990):

1. *Başlangıç.* Başlangıçta $\mathbf{w}(0)$ ağırlık vektörleri için rastgele değerler seçin. N haritadaki nöron sayısını gösterirse, seçimdeki tek kısıtlama $j \neq k$ için $\mathbf{w}_j \neq \mathbf{w}_k$ olmasıdır ($j, k = 1, 2, \dots, N$).
2. *Örnekleme.* Giriş dağılımından belli bir olasılıkla bir $\mathbf{x}$ vektörü seçin.
3. *Benzerlik Arama.* n -anında minimum Euklid uzaklığı şartı

$$i(\mathbf{x}) = \arg \min \|\mathbf{x}(n) - \mathbf{w}_j(n)\|, \quad (j = 1, 2, \dots, N) \quad (2.14)$$

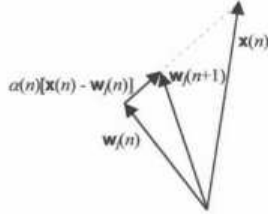
ifadesini kullanarak en uygun (kazanan vektör) $i(\mathbf{x})$ 'i bulun.

4. *Güncelleme.* Güncelleme bölgesindeki bütün nöronların sinaptik ağırlıklarını

$$\mathbf{w}_{j(n+1)} = \begin{cases} \mathbf{w}_{j(n)} + \alpha(n)[\mathbf{x}(n) - \mathbf{w}_{j(n)}], & j \in \Lambda_{i(\mathbf{x})} \\ \mathbf{w}_{j(n)} & \end{cases} \quad (2.15)$$

güncelleme formülünü kullanarak güncelleyin.

5. *Devam.* Öz nitelik haritasında dikkate değer bir değişme görülmedikçe kadar 2.nci adımdan işleme devam edin.



Şekil 2.14 Öz nitelik haritasında ağırlıkların güncellenmesi.

Ağırlıkların (2.15) ile verilen güncellenmesi Şekil 2.14'te geometrik olarak gösterilmiştir. $\alpha(n)$ n -nci adımdaki öğrenme hızı katsayısıdır. Kazanan nöronun $\Lambda_{i(x)}$ güncelleme bölgesi içindeki nöronların ya hepsi $\alpha(n)$ kadar giriş vektörüne hareket ettirilirlir veya $i(x)$ nöronuna uzaklıklarının Gaus fonksiyonu ile hareket ettirilirlir. Bütün nöronların aynı hızla güncellenmelerine "kabarcıklı güncelleme" (bubble updating), diğer güncellemeye ise "Gauss güncellemesi" (Gaussian updating) denir. Gaus güncellemesi durumunda (2.15)'in

$$\mathbf{w}_j(n+1) = \begin{cases} \mathbf{w}_j(n) + \alpha(n)\beta_j(n)[\mathbf{x}(n) - \mathbf{w}_j(n)], & j \in \Lambda_{i(x)} \\ \mathbf{w}_j(n) \end{cases} \quad (2.16)$$

şeklinde ayarlanması gerekir.

Burada

$$\beta_j(n) = e^{-\frac{d^2_j(n)}{R^2}} \quad (2.17)$$

$d_j(n)$, n -nci adımda nöronla kazanan nöron arasındaki uzaklık ve R , $\Lambda_{i(x)}$ bölgesinin yarıçapıdır.

Hem $\alpha(n)$ hem R eğitim ilerledikçe küçültülür. $\alpha(n)$ tipik olarak 0.05'ten başlayıp 0'a kadar azaltılabilir. R başlangıçta bütün haritayı kaplayacak kadar büyük alınır, eğitimin sonuna doğru 1'e yaklaştırılır. SOM_PAK (Kohonen vd., 1995) yukarıdaki öğrenme sürecinin iki

aşamalı olarak yapılmasını tavsiye etmektedir. Birinci aşamada $\alpha(0) = 0.05$, $R = uzun_kenar$ alınabilir ve $\mathbf{W}_f(0)$ rastgele seçilir. *uzun_kenar* haritanın uzun kenarında bulunan nöron sayısıdır. Öteleme sayısı 2000 dolayında olabilir. Bu aşamada öznetelik haritası kabaca düzene girer. İkinci aşamada $\alpha(0) = 0.03$, $R = uzun_kenar/3$, $\mathbf{W}_f(0)$ birinci aşamadan elde edilen değerler ve öteleme sayısı 15000 civarında olabilir. Bu aşamada öznetelik haritası son düzenlemesini kazanır.

Bu çalışmada öznetelik haritaları SOM_PAK yazılımı ile eğitilmiştir (EK2) ve iki aşamalı eğitim kullanarak topolojisi altıgen komşuluk olan 12×8 büyüğünde haritalar elde edilmiştir. Oluşan öznetelik haritaları Şekil 2.12'deki haritalara benzer. Ancak bu çalışmada yoğun olarak kullanılan LabVIEW öznetelik haritalarını Şekil 2.12'deki gibi altıgen komşuluğuyla gösterememektedir. LabVIEW'da öznetelik haritaları Şekil 4.2'deki gibi dikdörtgen komşulukla gösterilebilmektedir.

3 ÖZNETELİK ÇIKARMA

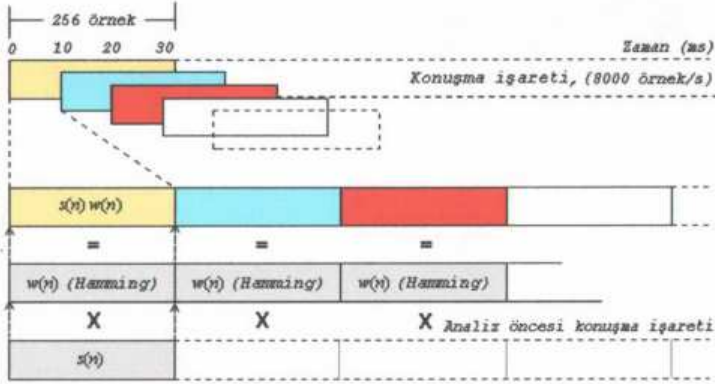
3.1 Giriş

Başarılı bir sınıflama için sınıfları belirleyen özniteliklerin seçilmesi büyük bir önem taşır. Ünlü seslerde olduğu gibi bazı ünsüz seslerde de kimlik bilgisi bulunabilir. Düşük enerji ve yüksek frekans muhtevasından dolayı ve kimlik belirleyen ünsüzlerin sayıca az olmaları sebebiyle, konuşanı teşhis etmek için konuşma işaretinin sadece ünlü kısımlarını kullandık. Konuşma işaretinin ünsüz kısımlarını atıp, sınıflama için ünlü kısımlarını dosya halinde sakladık. Hem eğitim kümesi, hem de test kümesi öznitelik çıkarma işlemi için bir ön işlemden geçmek zorundadır. Bu bölümde daha önce sözü edilen statik özniteliklerin çıkarılmasında kullandığımız analiz yöntemimiz anlatılacaktır. Dinamik öznitelikler SOM eğitimi sırasında YSA tarafından dolaylı olarak öğrenilecektir.

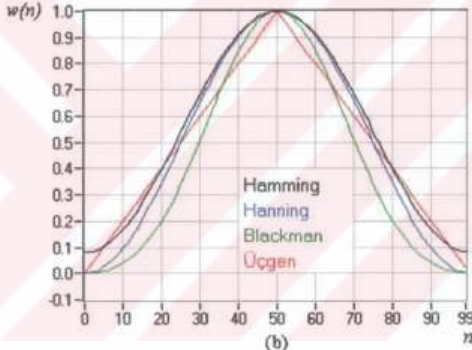
Bu analiz çalışmamız sırasında karşılaştığımız ve literatürde pek üzerinde durulmayan bir fenomene vurguda bulunacağız: Kadın seslerinin analizi sırasında temel periyod değerinin bazı bölgelerde aniden iki katına çıktığı görülmüştür*. Bu olaya **Yarım Perde Seslendirmesi**, YPS ismini verdik. İngilizcede **Pitch Halving** denilen bu olay, doğrusal ses üretim modelimizin geçerli olmadığını düşündüğümüz bir olgudur. Bölüm 3.4.1'de ele alacağımız bu olgu temel periyod çıkarımında yanıltıcı olabilir. Bu yüzden temel periyod çıkarımına (pitch detection) özel bir dikkat sarfettik. Buna rağmen yine de YPS sebebiyle iki misli değerlere sahip temel periyod kayıtları öznitelik dosyalarına girebilir. Bu yüzden 2-geçişli bir analiz sistemi geliştirdik. Birinci geçişte yakalanamayan YPS ikinci geçişte yakalanıp kayıttan çıkarılmaktadır.

Analizlerimizde TIMIT konuşma veri tabanını kullandık. TIMIT'de konuşmacılar yazılı metinden okuyarak konuşmaktadırlar. Konuşmalar 8 kHz ile örneklenmiştir. Ayrıca TIMIT oluşturulurken **"antialias"** ve 60 Hz – çentik filtre kullanıldığı için ve kayıtlar Texas

* Kullandığımız 32 ms uzunluktaki 256 örnekli analiz çerçevesi sadece kadın ve çocuk konuşmacılarda bu olayı gözlememize imkan tanımaktadır. Daha uzun çerçevelerde erkek seslerinde de YPS olayı ile karşılaşabiliriz.



(a)



(b)

Şekil 3.1 (a) Çerçeveleme ve pencereleme işlemleri, (b) Yaygın olarak kullanılan bazı pencereler.

Instruments tarafından stüdyo şartlarında yapıldığından işaret üzerinde bu işlemleri yapmamıza gerek kalmamıştır.

3.2 Çerçeveleme ve Pencereleme

Ses işareti kısa zaman dilimlerinde durağan kabul edilebilir. Ses biliminde bu dilim 10 ms ila 40 ms arasında seçilmektedir. Bu zaman dilimine **çerçeve** (frame) adı verilir. Analizler bu çerçeve içinde kalan işaret üzerinde yapılır. Bizim çerçevelerimiz 256 örnekten oluşur.

TABLO 3.1 BAZI PENCERELERİN MATEMATİKSEL İFADELERİ

Pencere	Pencere fonksiyonu
Dikdörtgen	$w(n) = 1$
Hamming	$w(n) = 0.54 - 0.46 \cos (2\pi n / N)$
Hanning	$w(n) = 0.5 [1 - \cos (2\pi n / N)]$
Blackman	$w(n) = 0.42 - 0.50 \cos (2\pi n / N) + 0.08 \cos (4\pi n / N)$
Bartlett (Üçgen)	$w(n) = 1 - (2n - N) / N $

TIMIT'in 8 kHz örnekleme hızı ile çerçeve uzunluğumuz 32 ms'dir. Konuşma işaretinin her 10 ms'sinde 32 ms'lik çerçeveler elde edildi. 10ms, 8 kHz örnekleme hızında ve 256 örnekli çerçevede 80 örneğe karşılık gelir. Dikdörtgen pencere dışındaki pencereler, kenarlara düşen bilgiye nazaran pencere ortasına düşen bilgiye daha fazla vurgu yapmaktadırlar. Pencereleri 10 ms aralıklarla üst üste bindirerek, öznelilik çıkarırken pencerenin kenarlarında kalan önemli bilginin kaybolmaması sağlanmış oldu (Şekil 3.1).

Pencereleme, $s(n)$ konuşma işareti ile $w(n)$ pencere fonksiyonunun çarpımı olarak düşünülebilir. Pencereleme sonunda elde edilen işarete $x(n)$ dersek, m 'nci çerçeve için

$$\begin{aligned} x(n) &= s(n)w(n-80m), \quad m = 0, 1, \dots \\ w(n) &= 0, n \notin [0, 255] \end{aligned} \quad (3.1)$$

yazabiliriz. $w(n-80m)$ 10'ar ms aralıklı pencere fonksiyonlarıdır. Ayrık zaman domenindeki bu çarpım, frekans domenine işaretleme pencere fonksiyonunun konvolusyonu olarak yansır:

$$X(\omega) = S(\omega) * W(\omega) \quad (3.2)$$

Bu sebeple işaretin gerçek frekans spektrumunu gözleyemeyiz. $w(n)$ pencere fonksiyonunun seçimi $X(\omega)$ üzerinde büyük bir etkiye sahiptir. Örnek olarak beş pencere alıp, pencere karakteristiklerini ve işaret üzerindeki etkilerini anlatıp, analizimiz için uygun olanını göstereceğiz.

TABLO 3.1'de beş tane pencere ve bunların ayırık zaman fonksiyonları verilmiştir. N çerçeve içindeki örnek sayısıdır. Fonksiyonların tanım aralığı $n \in [0, N-1]$ dir. Bu aralık dışında, yani $n \notin [0, N-1]$ için $w(n) = 0$ dir. Çerçeve içindeki işaret olduğu gibi alınırsa dikdörtgen pencere kullanılmış olur. Şekil 3.1(b)'de dikdörtgen, Hamming, Hanning, Blackman ve Bartlett (üçgen) pencereleri ayırık zaman domeninde gösterilmiştir.

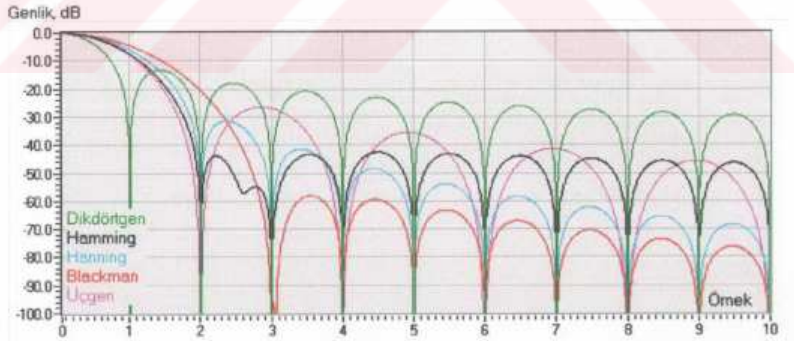
Şekil 3.2'de bu pencerelerin logaritmik genlik frekans cevapları karşılaştırma amacıyla üst üste çizilmiştir. Dikdörtgen pencerenin Fourier dönüşümü

$$W(\omega) = e^{-j\omega N/2} \frac{\sin(\omega N/2)}{\sin(\omega/2)} \quad (3.3)$$

denklemleri ile verilir. Burada N çerçeve uzunluğudur. Ayırık Fourier Dönüşümü bu sürekli spektrumu N ayırık noktada örnekler. Bu örnekleme frekansları $\omega = 2\pi k/N$, $k = 0, 1, \dots, N-1$ olarak alındığında (3.3)'den

$$|W(k)| = \left| \frac{\sin(k\pi)}{\sin(k\pi/N)} \right| = \begin{cases} N, & k = 0 \\ 0, & k = 1, \dots, N-1 \end{cases} \quad (3.4)$$

bulunur, yani **dikdörtgen pencere spektrumu, örnekleme noktalarında sıfırdan geçer.**



Şekil 3.2 Dikdörtgen, Hamming, Hanning, Blackman ve Bartlett pencerelerinin frekans genlik karakteristikleri. Yatay eksen spektrum örnekleme noktalarıdır. Dikdörtgen pencere en dar ana loba dolayısıyla en yüksek çözünürlüğe, buna karşılık -13 dB ile en yüksek ikinci loba sahiptir. Hamming penceresi birinci ve ikinci loblara nazaran Hanning penceresinden daha yüksek çözünürlüğe ve daha düşük ikinci loba etkisine sahiptir.

TABLO 3.2 PENCERELERİN SPEKTRAL ÖZELLİKLERİ

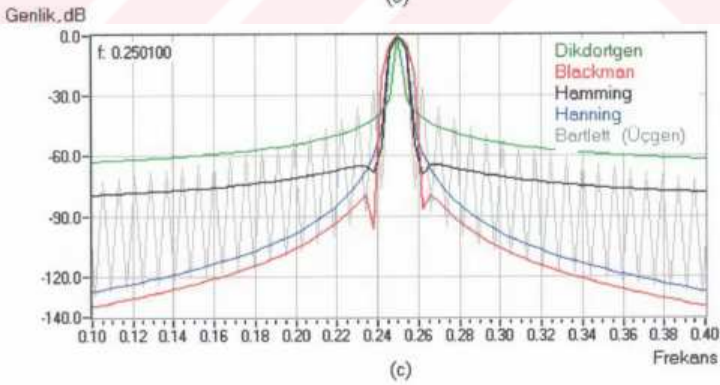
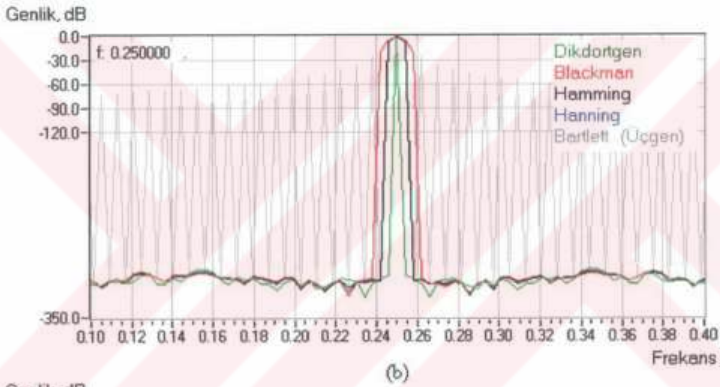
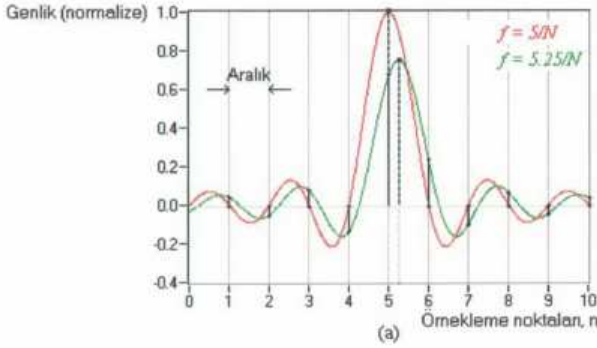
Pencere	-3 dB Ana Lob Genişliği (aralık)	-6 dB Ana Lob Genişliği (aralık)	Maksimum Yan Lob Seviyesi (dB)	Yan Lob Düşüş Hızı (dB/dekad)
Dikdörtgen	0.88	1.21	-13	20
Hamming	1.30	1.81	-43	20
Hanning	1.44	2.00	-32	60
Blackman	1.68	2.35	-58	60
Bartlett	1.28	1.78	-27	40

(3.3)'de ω 'yı $2\pi f = 2\pi F / F_s$ şeklinde yazarsak,

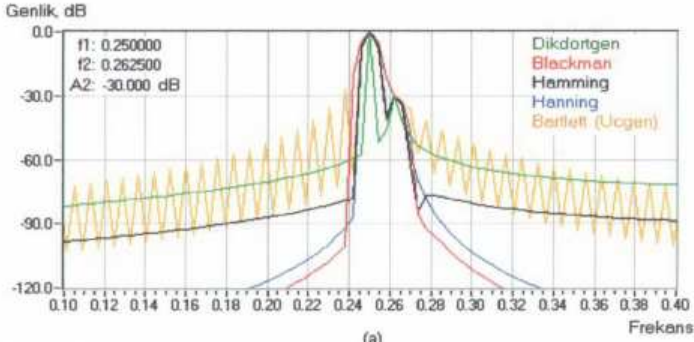
$$|W(F)| = \left| \frac{\sin(\pi N F / F_s)}{\sin(\pi F / F_s)} \right| \quad (3.5)$$

elde edilir. Burada F frekans (Hz), f normalize frekans, F_s de örnekleme hızıdır (örnek/s). Çerçevesi işaretin Fourier dönüşümü $X(F)$, konuşma işaretinin dikdörtgen pencere ile konvolusyona olup frekans bileşenleri etrafında "sinc" fonksiyonuna benzeyen bir yayılma gösterir. (3.4)'den dolayı, NF/F_s tam sayı ise $|W(k)| = 0$ olur; aksi halde $|W(k)|$ sıfırdan farklıdır. Bir frekans bileşeninin sağ ve solundaki spektrum örnekleme noktalarını göz önüne alalım. k , örnekleme noktalarını gösteren bir tam sayı olsun ve $k-1 < NF/F_s < k$ şartını sağlasın. Bu durumda $W(k-1)$ ve $W(k)$ sıfıra eşit olmazlar. Bu olaya **spektral sızıntı** denir. Şekil 3.3 (a)'da kırmızı eğri sızıntısız, yeşil eğri sızıntılı durumu göstermektedir.

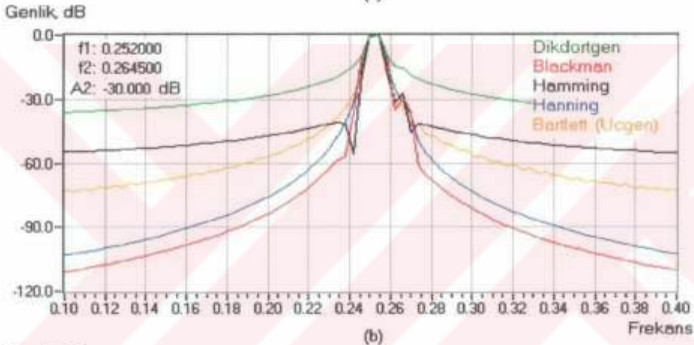
Pencere spektrumunun -3 dB ve -6 dB ana lob genişlikleri, yan lob seviyeleri ve işaret frekansının spektrum örnekleme noktalarına olan uzaklığı sızıntı miktarını belirler [Şekil 3.3(b),(c), TABLO 3.2]. TABLO 3.2'de ikinci ve üçüncü sütundaki "**aralık**" iki spektral örnek noktası arasındaki aralıktır.



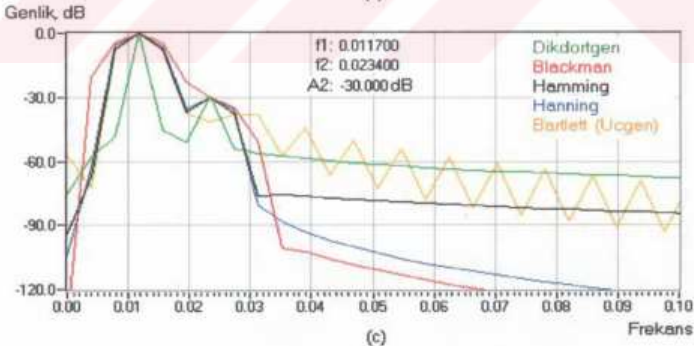
Şekil 3.3 Çeşitli pencerelerde spektral sızıntı. (a) İşaret frekansı spektrumun örnekleme noktalarıyla çakışmadığında spektral sızıntı görülür (yeşil eğri, $f = 5.25/N$). $f = 5/N$ için spektral sızıntı görülmez (kırmızı eğri). (b) $f = 0.25$ ve (c) $f = 0.2501$ için pencerelerin işaret spektrumunu nasıl etkiledikleri görülmektedir.



(a)



(b)



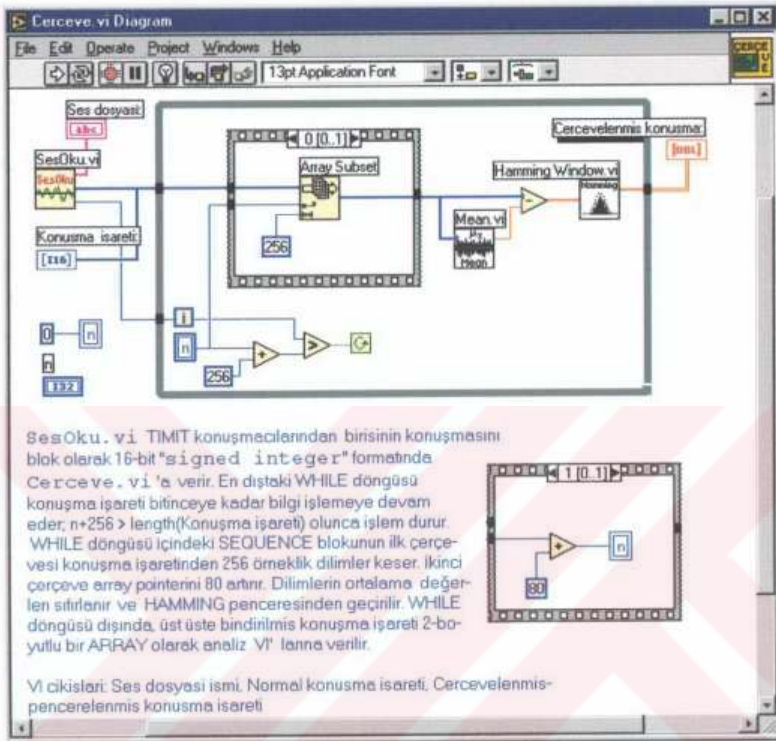
(c)

Şekil 3.4 Çeşitli pencerelerle işaret çözünürlüğü. (a) 2000 Hz ve 2100 Hz, (b) 2016 Hz ve 2116 Hz frekansındaki sinüzoidal tonların pencere çıkışlarındaki spektrumları. Tonlar arasında 100Hz frekans farkı, 30 dB seviye farkı vardır. 100Hz kalın bir erkek sesinin temel frekansındır. (c) Logaritmik genlik spektrumunda 3/256 ve 6/256 frekansındaki iki sinüzoidal tonun keşifleri.

Bu pencerelerden birini seçerken konuşma işaretinin özellikleri göz önüne alınmalıdır. Temel frekans erkeklerde 100 Hz değerini alabilmektedir. Dolayısıyla zaman domenindeki penceremiz 100 Hz ve harmoniklerini ayırabilecek çözünürlüğe sahip olmalıdır. Pencereleri test etmek için iki sinüzoidal ton kullanabiliriz. İki tonun frekansları arasında 100 Hz fark olsun. İkinci tonun seviyesi de birinci tonun seviyesinden 30 dB düşük olsun. Şekil 3.4(a) 2000 Hz ve 2100 Hz için, Şekil 3.4 (b) ise 2016 Hz ve 2116 Hz için pencere performanslarını göstermektedir. 256 örnek ve 8000 Hz örnekleme frekansında iki spektral örnek noktası arası 31.25 Hz dir. Maksimum spektral sızıntı elde edebilmek için Şekil 3.4(b)'de frekansları bu aralığın yarısı kadar kaydırdık. Şekil 3.4(a)'da 2000 Hz ($f = 0.25$) frekanslı ton için Bartlett penceresi dışındaki pencerelerden spektral sızıntı görülmektedir. Sızıntı 2100 Hz'de görülmektedir. Dikdörtgen pencere ve Hamming penceresi bu iki tonu rahatça ayırabilmektedirler. Şekil 3.4(b) de ise her iki tondan sızıntı görülmekte olup sadece Hamming ve Blackman pencereleri -30 dB 2116 Hz'lik tonu 0 dB 2016 Hz'den ayırabilmektedir. Bartlett penceresi en kötü performansa sahip bulunmaktadır. Diğer pencerelerin spektrumları $\Delta f = F_s / 256$ aralıklarla sıfırdan geçerken, Bartlett penceresinin spektrumu $2\Delta f$ aralıklarla sıfırdan geçmektedir (Şekil 3.2). Zaman domeninde en uygun performans veren Hamming penceresini kullanmayı uygun gördük.

Logaritmik genlik spektrumunda en düşük frekans içeriği formant zarfında görülür. En düşük frekans için tipik olarak $3/256$ değeri alınabilir. Şekil 3.4(c)'de $f = 3/256$ (0 dB) ve $f = 6/256$ (-30 dB)'lik iki sinüzoidal tonun keprstrumu görülmektedir. Burada en iyi performansı dikdörtgen pencere vermektedir. Bu yüzden logaritmik genlik spektrumundan keprstrum elde ederken ters FFT işleminden önce bir pencere kullanmadık; yani aslında dikdörtgen pencere kullanmış olduk.

Şekil 3.5 buraya kadar anlatılanların LabVIEW platformunda gerçekleştirilişini göstermektedir. LabVIEW, C ve PASCAL gibi ardışıl komut işleme (sequential instruction execution) esasına göre değil, bilgi akışı (data flow execution) esasına göre çalışır. Bilgi akışlı çalışmada bir blokun çalışabilmesi için ona bağlanan bütün bilgilerin hazır olması gerekmektedir. Program akış kontrolü yazılı bir kod sırasıyla belirlenmez. Şekil 3.5'deki WHILE döngüsünün başlayabilmesi için SesOku.vi'dan Konuşmaİşareti ve KonuşmaUzunluğu'nun WHILE döngüsüne ulaşması gereklidir. WHILE döngüsü içindeki SEQUENCE bloğu işlem sırasını zorla belirlemek istediğimiz durumlarda



Şekil 3.5 Çerçeveleme ve pencerelleme işlemlerinin LabVIEW platformunda gerçekleştirilmesi.

kullanılır. "0" numaralı blok, "1" numaralı bloktan önce işlem görür: Önce uzun konuşma işaretinden 256 örnekli bir çerçeve kesilir, sonra "n" 80 artırılır.

SEQUENCE bloklarının işlemi bittikten sonra Mean.vi tarafından çerçevenin ortalama değeri bulunur ve bütün çerçeve elemanlarından çıkarılır. Bu 0-ortalama 256 örnekli dizin üzerine Hamming.vi tarafından Hamming penceresi uygulanır. WHILE döngüsü çıkışında konuşma bilgisi 2-boyutlu bir dizin olarak bir sonraki işleme verilir.

Cercele ve.vi'ye eşdeğer C kodu TABLO 3.3'te verilmiştir:

TABLO 3.3 C KODU ile ÇERÇEVELEME ve PENCERELEME

```

double   KonusmaIsareti[], Cerceve[256];
double   Orta;
int       n = 0;                      /* çerçeve başı */
int       i;
int       uzunluk(double array[]); /* array uzunlugunu bul */

while (n < (uzunluk(KonusmaIsareti) - 256))
{
    /* Konusma işaretinden 256 örnekli bir dilim al.
    Çerçeveye kopyala. Aynı zamanda dilimin ortalamasını bul */

    Orta = 0.0;
    for (i=0; i < 256; i++) {
        Cerceve[i] = KonusmaIsareti[n+i];
        Orta += (Cerceve[i] - Orta)/(i+1);
    }

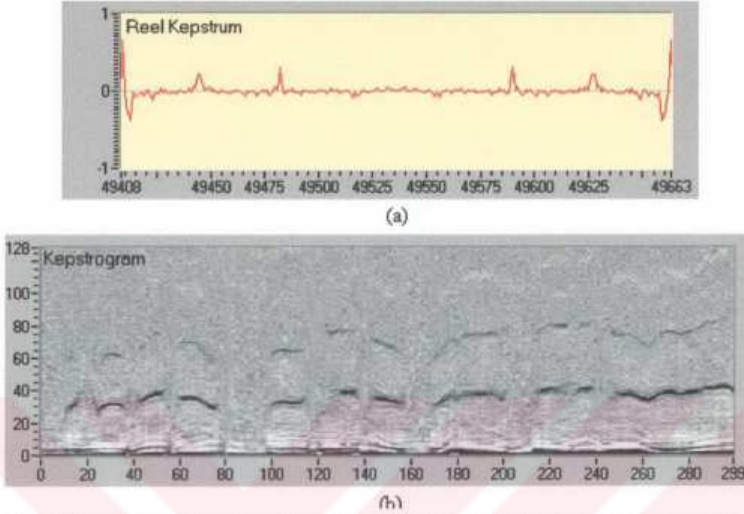
    /* Çerçevenin ortalamasını sıfırla ve Hamming penceresiyle çarp */
    for (i=0; i < 256; i++) {
        Cerceve[i] -= Orta;
        Cerceve[i] *= 0.54-0.46*cos(2.0*pi*i/256.0);
    }
    n += 80;      /* Bir sonraki çerçeveyi 10 ms öteden başlat */
}

```

3.3 Öznitelik Uzayı

Kepstrum analizi sırasında artikülasyon ve glottal uyarım karakteristikleri birbirlerinden ayrılmış olarak elde edilir. Şekil 3.6 tek çerçevede ve konuşma süresince kepstrum katsayılarını göstermektedir. Kepstrumun aşağı kefrans bölgesine düşen 20 katsayı artikülasyon filtresini temsil eden özniteliklerdir. 49445 numaralı örnek civarındaki tepe ise temel periyod T_0 'dır. Konuşmacaya ait kim bilgisi hem glottal uyarımda hem de larinks anatomisinde olduğu için tanımda bu katsayılar ve T_0 kullanılmalıdır. Bu yüzden analiz boyunca konuşmanın ünlü bölümleri alınıp, T_0 bilgisi bulundurmeyan ünsüz kısımlar atılmıştır. Ancak ilk kepstrum katsayısı logaritmik genlik spektrumunun DC değeri (yani artikülasyon+glottal uyarımın çerçeve içindeki ortalaması olduğu için) olduğu için öznitelik vektörüne dahil edilmesine gerek yoktur.

Öznitelik vektörü aşağı kepstrum bileşenlerinin aldığı değerleri ve temel periyod bilgisini içerir. Konuşmacılar sabit bir konuşma metnine bağımlı kalarak tanınmak istenirse, T_0 'ın hem görüldüğü yerin hem de tepe değerinin öznitelik olarak kullanılması gerekir. Çünkü



Şekil 3.6 Öznitelik olarak kullanılan reel kepstrum katsayıları. (a) 1-19 arasındaki kepstrum bileşenleri ve temel periyod T_0 (sivri tepe) çalışmamızda öznitelikler olarak kullanılmıştır. (b) Konuşma süresince kepstrum bileşenlerinin değişimini gösteren kepstrogram. Yatay eksenlerde rakamlar ms olarak ayrıık kefrans (a) ve çerçeve sayısını temsil etmektedir (b).

konuşma boyunca intonasyonun yeri ve şiddeti kişilik belirleyici bir dinamik öznitelik olabilir. İntonasyon ise T_0 'ın hem yeri hem de değeriyle belirlenir. Bizim hedefimiz metinden bağımsız olarak konuşmacıyı tanımak olduğu için konuşmacıların okudukları değişken metinlerde intonasyonu öznitelik olarak kullanamayız. Ayrıca Bölüm 3.4.1'de görüleceği gibi, eğer T_0 'ın yanında YPS de varsa T_0 'ın değeri hiç bir anlam ifade etmez. Bu yüzden sadece T_0 'ın kefrans değeri özniteliklere dahil edilmiştir.

Üzerinde durulması gereken ikinci bir nokta da glottal uyarım öznitelikleri ile artikülasyon özniteliklerinin kişilik belirlemedeki nisbi ağırlıklarıdır. Bu sorunun psikolojik karşılığı şudur: "Biz kişilerin seslerini daha çok tizlik ve peslikle mi, yoksa hançere özellikleri ile mi ayırdederiz?" Kepstrum katsayılarının tepe değerleri ve varyanslarına göre, temel periyod değerleri ve varyansları dominanttır. Eğer T_0 değerini olduğu gibi kullanırsak iki öznitelik vektörünün arasındaki uzaklığı hesaplamak temel periyod diğer kepsral özniteliklerin etkisini gölgeleyebilir. Bu yüzden T_0 değerini doğrudan doğruya kullanma yerine kT_0 ($0 < k \leq 1$) şeklinde bir skala faktörü ile çarparak kullanmak daha doğru olacaktır. k

arttıkça temel periyodun tanımadaki ağırlığı artar. $k = 0$ olduğu zaman glottal uyarımın tanımadaki belirleyici etkisi tamamen ortadan kalkar. Sırf larinks karakteristikleri ile konuşmacıyı tanımanın fazla başarılı olmayacağını düşündüğümüz için $k = 0$ alınmamalıdır. En iyi konuşmacı tanıma için k ne olmalıdır? Bu yukarıda yöneltilen sorunun bir başka şeklidir. k 'ya değişik değerler vererek elde edilecek en doğru tanıma skoru, k 'nın değerini belirleyecektir.

Böyle bir k skala faktörünü gerekli kılan diğer bir etken de iki vektör arasındaki uzaklık kavramıdır. $\mathbf{X}$ ve $\mathbf{Y}$ karşılaştırmak istediğimiz iki öznitelik vektörü olsun. Karşılaştırma için Euklid uzaklığını kullandığımızda uzaklık, vektörler farkının normu olarak verilir:

$$d(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\| = \left[\sum_{i=0}^{19} (x_i - y_i)^2 \right]^{1/2} \quad (3.6)$$

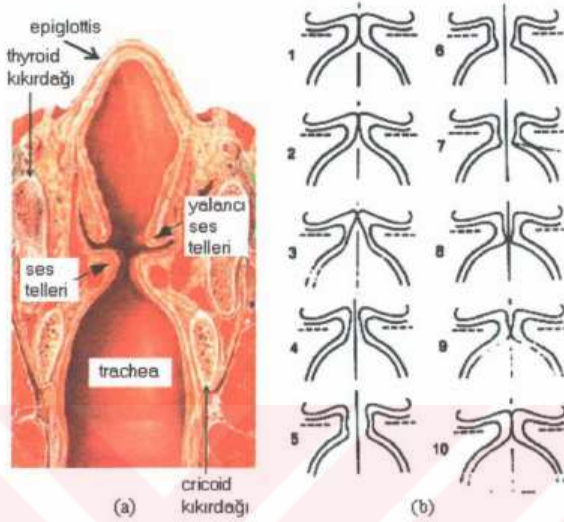
x_0, y_0 temel periyodu, $x_i = c_i^x$, $y_i = c_i^y$ ($i = 1, \dots, 19$) kepstrum katsayılarını gösterebilir. x_0 , y_0 'ın birimi saniye, x_i, y_i 'nin birimi ise volt'tur. (3.6)'daki toplamın anlamlı olabilmesi için x_0, y_0 yerine kx_0 ve ky_0 kullanılmalıdır. O zaman k skalasının birimi volt/saniye olur ve (3.6) volt olarak elde edilir. Buna göre (3.6)

$$d(\mathbf{X}, \mathbf{Y}) = \left[k^2 (T_{0x} - T_{0y})^2 + \sum_{i=1}^{19} (x_i - y_i)^2 \right]^{1/2} \quad (3.7)$$

olur. Sonuç olarak öznitelik vektörümüz $[kT_0, c_1, \dots, c_{19}]$ olarak gösterilebilir. Burada c_1 ikinci kepstrum katsayısıdır. c_0 logaritmik güç spektrumunun DC değeri olduğu için vektöre dahil edilmemiştir.

3.4 Analiz

Eğitim ve tanıma işlemlerinden önce konuşma işaretini analiz etmek gereklidir. Analizden elde edilen öznitelik vektörleri eğitim ve tanıma amacıyla kullanılabilir. Bölüm 3.1'de anlatıldığı gibi bu çalışmada iki kademelili bir analiz geliştirildi. Birinci kademede elde edilen sonuçların üzerinden ikinci kademede bir daha geçildiği için bu kademelere 1.nci Geçiş ve 2.nci Geçiş isimleri verildi. 1.nci Geçiş'te konuşma işaretinin ünlü kısımları alınıp öznitelik vektörleri çıkarılır. 2.nci Geçiş'te YPS veya gürültü sebebiyle elde edilen anormal T_0 değerleri (outliers) veri setinden ayıklanır.

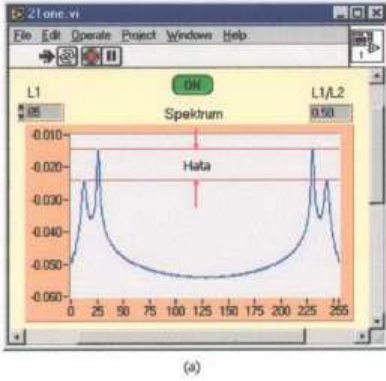


Şekil 3.7 Glotal uyarım. (a) Larinksin ara kesiti ve ses tellerinin anatomisi; (b) Ses tellerinin titreşmemekanzması (The Voice Center: Anatomy of the Larynx. <http://www.unil.ch/ling/phonetique/api1-eng.html>)

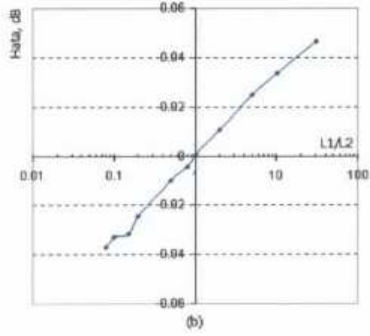
Analiz yaparken Bölüm 3.1'de değinilen YPS'ye dikkat etmek gereklidir. Şekil 3.6(a)'da 49445.nci örnekleme noktasındaki tepe Normal Perde Seslendirmesi (NPS), 49485 noktasındaki tepe ise YPS'dir. Şekil 3.6(b)'de alttaki koyu izler NPS ile üretilen temel peryod, üst tarafta iki kat uzaklıktaki açık renkli iz ise YPS'dir. YPS aynı çerçevede NPS ile beraber, veya tek başına görülebilir. NPS ile beraber görüldüğünde seviyesi NPS seviyesinden büyük veya küçük olabilir. YPS'yi dikkate almadan T_0 dedeksiyonu hatalı öznitelik vektörü üretebilir. Onun için YPS üzerinde kısaca durmakta fayda görüyoruz.

3.4.1 Yarım perde seslendirme problemi

YPS'nin, Bölüm 2'de anlatılan doğrusal konuşma üretim modelinin geçerliliğini kaybettiği bir olgu olduğunu düşünüyoruz. Doğrusal konuşma üretim modeli glotal uyarımla artikülasyon filtresinin bağımsız olduklarını varsayar. Halbuki büyük bir olasılıkla akustik



(a)



(b)

Şekil 3.8 Aynı çerçeve içinde peşpeşe gelen iki sinüzoidal tonun güç spektrumu. (a) Spektrumu elde eden sanal cihaz (2Tone.vi). Yatay eksen ayrık frekans, dikey eksen dB cinsindendir; (b) Hatanın $L1/L2$ ile değişimi.

modelin arakesitlerinden geriye dönen yansımalar bu varsayımı geçersiz kılmaktadır. Kaynakla filtre arasındaki bu bağlantı lineer sistem analizinin öngördüğü konvolusyon ilişkisini bozar. Bu yüzden NPS'nin görülmeyip sadece YPS'nin görüldüğü çerçeveler öznetelik değerlendirme-sine alınmamıştır.

Temel Bilgiler Bölümünde ses üretim mekanizması anlatılmıştı. Şekil 3.7'de ses telleri ve bunların titreşimi görülmektedir. Şekildeki yalancı ses telleri ses üretimine katkıda bulunmazlar. **Glotal uyarımda sadece gerçek ses telleri rol almaktadır.** Bir başka deyişle ses üreten kaynak bir tanedir. Öyleyse çerçevelerimizde hem NPS hem de YPS aynı anda nasıl görülebilmektedir? Bu sorunun cevabı Fourier analizi ve çözünürlük kavramlarında yatmaktadır.

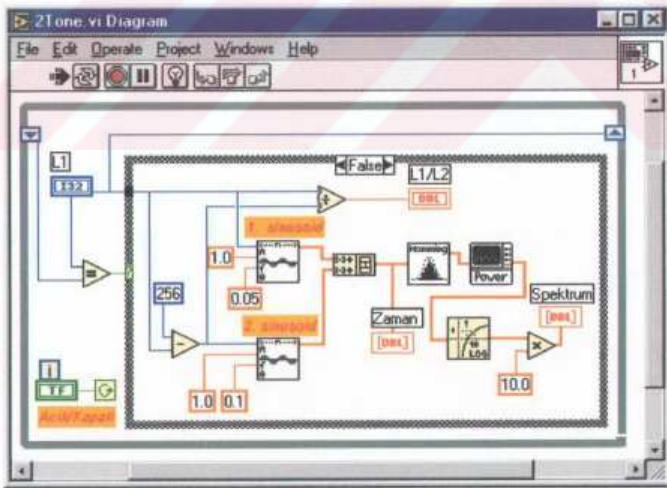
Yukarıda yöneltilen soruyu cevaplamak için 256 örnekli çerçevemizde peşpeşe iki sinüzoidal ton bulunduğunu düşünelim. Birinci ton L_1 süresince, ikinci ton da $L_2 = 256 - L_1$ süresince uygulanır:

$$x(n) = \begin{cases} \sin(2\pi f_1 n) & f_1 = 0.05, 0 \leq n \leq L_1 \\ \sin(2\pi f_2 n) & f_2 = 0.10, L_1 \leq n \leq 256 \\ L_2 = 256 - L_1 \end{cases} \quad (3.8)$$

olsun.

(3.8) ile verilen $x(n)$ 'in Fourier dönüşümünün açık yazılışı oldukça karışık ve uzun bir ifadedir. Bizim için bu dönüşümün açık yazılışından çok ifade ettiği anlam önemlidir. LabVIEW ile kolaylıkla bu dönüşümü yapabiliriz. Bu amaçla 2Tone.vi isimli bir sanal cihaz hazırladık. Şekil 3.8(a) ve Şekil 3.9 2Tone.vi'nin panosunu ve diyagramını göstermektedir. Bölüm 3.2'de anlatılan spektral sızıntı sebebiyle 2Tone.vi'da Hamming penceresi kullanılmıştır.

2Tone.vi'nin elde ettiği güç spektrumuna bakarak - eğer kaynağın ardışıklığı bilinmezse - bu iki ardışıl sinüzoidal ton, iki toplamsal ton olarak yorumlanabilir. Larinksin anatomisinden iki tonun toplamsal olamayacaklarını biliyoruz. O halde Şekil 3.6'da $2T_0$ 'da görülen ton T_0 ile ardışıldır ve YPS'den kaynaklanmaktadır. 2Tone.vi'da seviyeler dB cinsinden bulunur. Eşit genliklerine rağmen iki tonun spektral tepeleri arasında seviye farkı bulunmaktadır. Bu fark "hata"yı oluşturur. Şekil 3.8(b)'de hatanın logaritmik skalada lineer olduğu görülmektedir. Hata 0.03 dB/dekad hızıyla değişmektedir. Kısa Zamanlı Fourier



Şekil 3.9 2Tone sanal cihazının diyagramı. Ön panodaki anahtar ON durumunda olduğu müddetçe WHILE döngüsü devam eder. L1'e yeni bir değer atanınca FALSE CASE icra edilir. L1'in eski değeri WHILE döngüsünün duvarlarındaki REGISTER'da saklanmaktadır.

Dönüşümünün zaman çözünürlüğü kullanılan çerçevenin uzunluğu kadardır (32 ms). Bu sebeple L_1 ve L_2 'yi ölçemeyiz. Dolayısıyla L_1/L_2 'yi bilmeden, ardışıl iki sinüzoidin seviyelerinin spektrum grafiğinden hatasız elde edilemeyeceği anlaşılmaktadır. Bu yüzden de temel periyodun genliği öznitelik vektörüne sokulmamıştır.

3.4.2 Birinci geçiş

Daha önce LabVIEW sanal cihazlarına örnek verilmişti (Şekil 3.5, Şekil 3.8, Şekil 3.9). Pascal, C, BASIC gibi dillerdeki "program" nosyonunun yerini LabVIEW'da Virtual Instrument (sanal cihaz) kavramı almıştır. Virtual Instrument kısaca VI diye isimlendirilir ve "viay" diye okunur. Yukarıda anılan dillerdeki function, procedure, subroutine gibi kavramlar yerine SubVI (sabviay) kavramı getirilmiştir. Program akış kontrolleri (for-next, while döngüleri), dallanma kontrolleri (if, switch, case) vb. programlama elemanlarının karşılıkları da LabVIEW'da vardır. Burada LabVIEW'nun ayrıntılarına girmek konumuzun dışına çıkmak olacaktır. Bunun yerine, sanal cihazların çalışması anlatılırken ve ihtiyaç doğdukça LabVIEW'nun ilgili özellikleri ele alınacaktır.

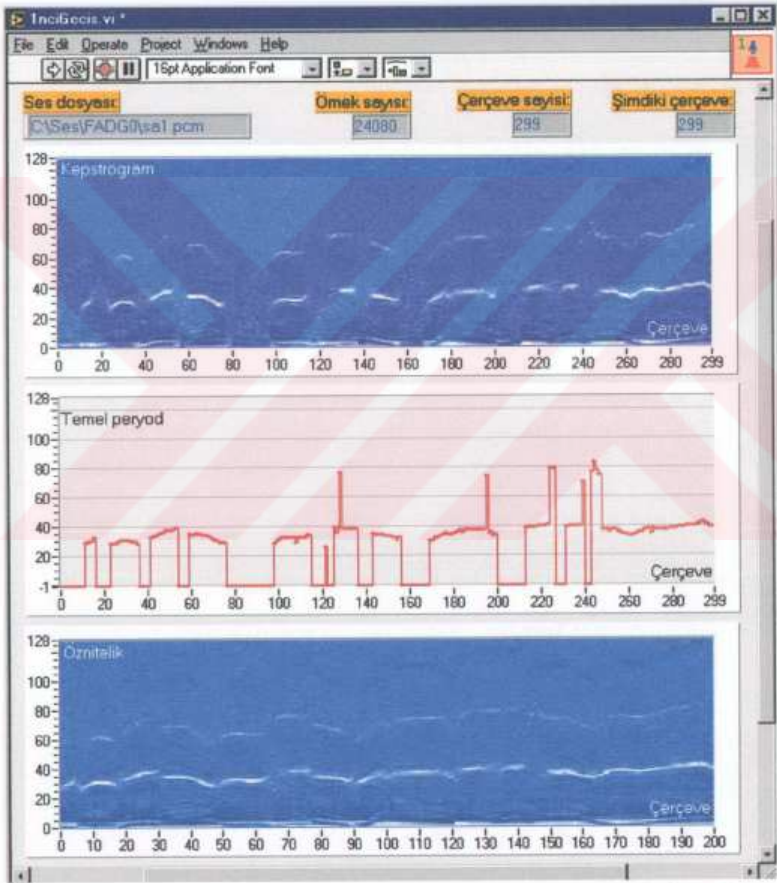
Birinci geçiş analizi `1nciGecis.vi` sanal cihazıyla yapılır (Şekil 3.10). Bu sanal cihazda yapılan işlemler ve analizler:

1. Çerçeveleme ve pencereleme,
2. Kepstrum analizi (Homomorfik analiz),
3. "Thresholding" ve temel periyod dedeksiyonu, ünsüz seslerin atılması,
4. Öznitelik vektörlü oluşturup ikinci geçişte kullanmak üzere dosyalamak ve
5. Kullanıcıya görsel temel periyod ve kepstrogram imkanı sunmak

şeklinde sıralanabilir.

Şekil 3.10'da `1nciGecis.vi`'in panosunda bir örnek analiz gösterilmiştir. Sanal cihazın en üst sırasında ses dosyasının ismi, örnek sayısı, çerçeve sayısı ve analiz edilen çerçevenin numarası gibi bilgiler bulunur. Çerçeve sayısını ve kepstrogram görüntüsünü yorumlarken çerçevelerin 10'ar ms (80 örnek) aralıklarla üst üste bindiklerini tekrar

hatırlatmak isteriz. *fadg0\sal.pcm* cümlesi 24080 örnekten oluşmaktadır ve *lnciGecis.vi* çerçeve sayısını 299 olarak vermektedir. Halbuki üst üste binmeyen 256 örnekli çerçevelerle bu konuşmada ancak 94 çerçeve bulunabilirdi. Bu sebeple kepstrogram ve temel periyod diyagramları gerçeğe nazaran 3.2 kat daha uzundurlar (Bölüm 3.2).



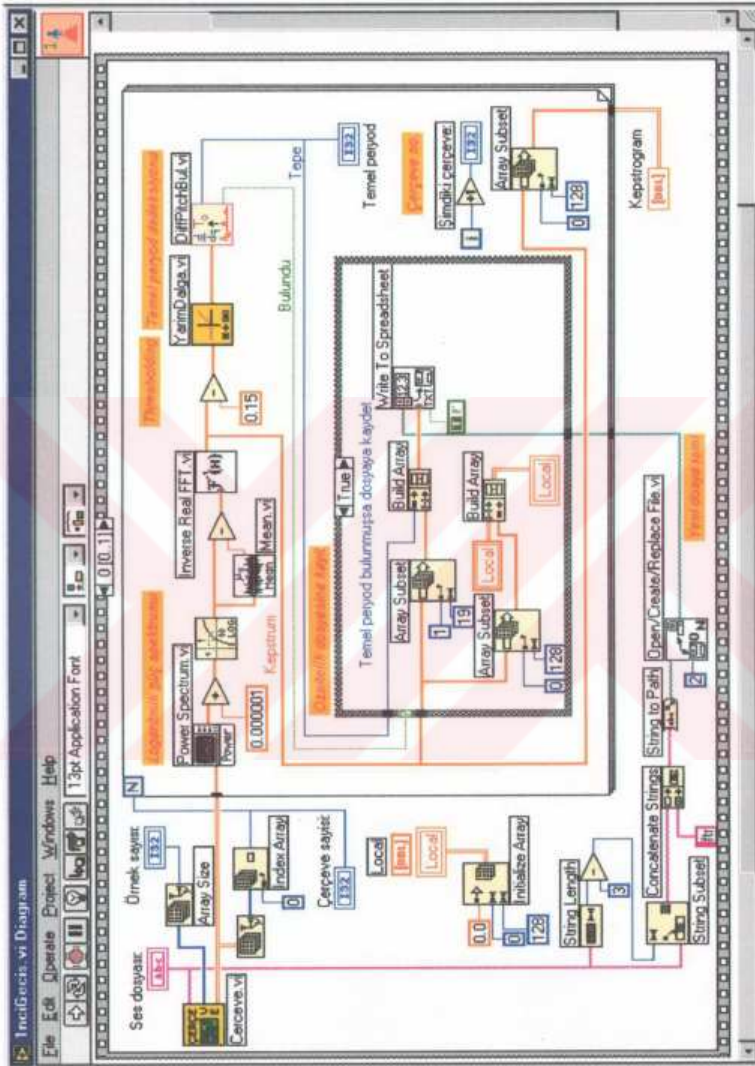
Şekil 3.10 *lnciGecis.vi* panosu. En üstte konuşma işaretinin kepstrogramı, onun altında temel periyod dedektörünün çıkışındaki temel periyod T_0 , en altta ise ünsüz kısımlar atıldıktan sonra kepstrogram görülüyor. Temel periyod ve en alt 19 kepstrum katsayısı *2nciGecis.vi* tarafından okunmak üzere dosyalanır.

Şekil 3.10'da en üstte homomorfik analizle elde edilen kepstrogram bütün konuşmayı kapsamaktadır. Kefrans ekseninde 20-40 arasında seyreden beyaz iz temel periyod T_0 'dır. $2T_0$ 'da seyreden soluk beyaz iz ise YPS'den kaynaklanmaktadır. T_0 izinin kesintiye uğradığı yerlerde ünsüz sesler bulunmaktadır. Ortadaki şekilde temel periyod dedektörü ile elde edilen temel periyod izi görülmüyor. Burada da ünsüz kısımlar ve YPS belirgin olarak gözlenmektedir. En alta, ünsüz kısımlar atıldıktan sonra geriye kalan konuşma kepstrogramı verilmiştir. Bu kepstrogramdaki 1-19 kefrans değerleri arasındaki katsayılar öznitelik vektörüne dahil edilerek dosyalanır.

`InciGecis.vi`, konuşma işaretini çerçevelenmiş, Hamming penceresi ile çarpılmış, 10 ms aralıklarla üst üste bindirilmiş $m \times 256$ büyüklüğünde iki boyutlu bir dizin halinde `Cerceve.vi`'dan alır. m konuşmanın içerdiği çerçeve sayısıdır. Şekil 3.11'de `Cerceve.vi`'in çıkışındaki 1-boyutlu dizine bağlı `ArraySize` fonksiyonu konuşma işaretinin örnek sayısını, 2-boyutlu dizine bağlı `ArraySize` fonksiyonu `IndexArray` fonksiyonu ile beraber çerçeve sayısı m 'i bulur. Çerçeve sayısı m `for` döngüsünün "N" terminaline bağlanır. `for` döngüsü içindeki "i" değişkeni döngü sayacıdır. Bu şekilde bütün çerçeveler `for` döngüsü içindeki işlemlerden geçmiş olur. Bunları C sintaksına "yaklaşık olarak"

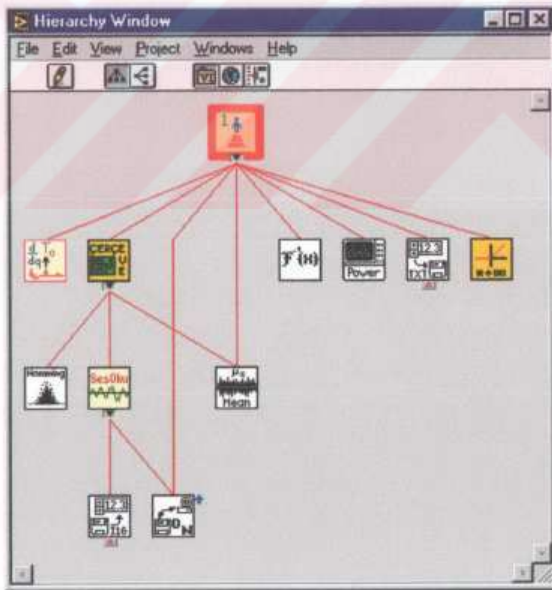
```
for(i=0; i < m /* ÇerçeveSayısı */; i++)
{
    KepstrumAnalizi(); /* Homomorfik analiz */
    Thresholding();
    T0 = TemelPeryodDedeksiyonu();
    if (T0 != -1)
    {
        /* T0 ile 19 kepstrogram katsayısını birleştir..*/
        /* ..ve öznitelik dosyasına kaydet*/
        SimdikiCerceve = i+1;
    }
}
```

şeklinde tercüme edebiliriz. "Yaklaşık olarak"tan kasdımız LabVIEW'nun yukarıdaki analizi C sintaksına paralel bir sıra takip ederek yapmadığını vurgulamak içindir. Şekil 3.11'de sonu `vi` ile biten ikonlar `subvi`'lardır (`Cerceve.vi`, `YarımDalga.vi` gibi). Bunun dışındaki ikonlar hazır LabVIEW fonksiyonlarıdır (`Initialize Array`, `String Subset` gibi). Parametre geçişi (parameter passing) LabVIEW'da ikonları birleştiren teller ile sağlanır.



Şekil 3.11. InciGecis.vi diyagramı. Bu geçişte sadece ünlü kısımların öznetelik vektörleri elde edilir.

Şekil 3.12'de InciGecis.vi'nin SubVI hiyerarşisi görülmüyor. Diğer dillerdeki fonksiyon çağırma eşiği olarak LabVIEW SubVI çağırır. VI diyagramına bir VI yerleştirerek çağırma işlemi gerçekleşir. VI'nın giriş terminal(ler)ine tellerle bağlantı yapılarak VI'ya parametre geçişi sağlanır. InciGecis.vi, DiffPitchBul, Cerceve, InverseRealFFT, Mean, PowerSpectrum, ResetArrElem, WriteToSpreadsheet ve YarımDalga VI'larını; Cerceve.vi SesOku, Hamming, ve Mean VI'larını; SesOku.vi da ReadFromI16File ve OpenFile+ VI'larını çağırır. TABLO 3.4'te InciGecis.vi'nin kullandığı subvi'ların listesi verilmiştir. En sol sütunda bu subvi'ların Şekil 3.11'de kullanılan ikonları yer almıştır. Tarafımızdan hazırlanan subvi'ların açıklamaları aşağıda gelecektir. LabVIEW tarafından sağlanan VI'lar genel maksatlı VI'lar olup yerine getirdikleri fonksiyonlar isimlerine yansıtılmıştır. Bunlar hakkında daha detaylı bilgi "LabVIEW User's Manual'da bulunabilir. InciGecis.vi ile yapılan işlem ve analizler Bölüm 3.4.2'nin girişinde (sayfa 56) anlatılmıştı. Aşağıda bu işlemleri birer birer ele alacağız.



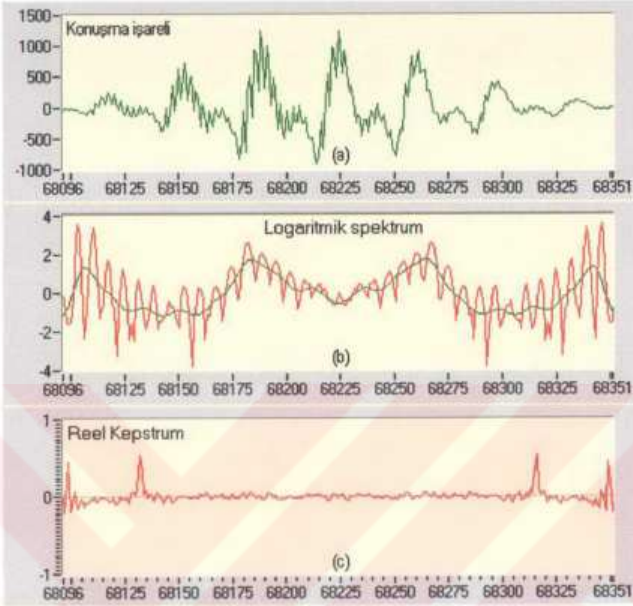
Şekil 3.12 InciGecis.vi'nin bağımlı olduğu SubVI hiyerarşisi. SubVI'lar diğer dillerdeki subroutine veya alt fonksiyonların karşılığıdır.

TABLO 4.4 InciGecis.vi'DA KULLANILAN SubVI'LAR

İKON	SubVI Adı	Fonksiyonu	Orijini
	SesOku.vi	*.pcm ses dosyasını okur	Yazar
	Cerçeve.vi	Çerçeveleme, pencereleme, dosya diyalogu	Yazar
	Yarımdalga.vi	Yarım dalga doğrultucusu	Yazar
	DiffPitchBul.vi	Türevsel temel periyod bulucusu	Yazar
	Power Spectrum.vi	1-Boyutlu bir dizinin güç spektrumu	LabVIEW
	Inverse Real FFT.vi	1-Boyutlu bir kompleks dizinin ters Fourier dönüşümü	LabVIEW
	Mean.vi	1-Boyutlu bir dizinin ortalama değeri	LabVIEW
	Open/Create/Replace File.vi	Dosya açma, yeni dosya oluşturma, eski dosyayı yenisiyle değiştirir	LabVIEW
	ReadFromI16File	Dosyadan 1-Boyutlu veya 2-Boyutlu signed integer okur.	LabVIEW
	Write To Spreadsheet.vi	1-Boyutlu veya 2-Boyutlu tek-presizyonlu bir dizini bir text dosyasına yazar.	LabVIEW

3.4.2.1 Çerçeveleme ve pencereleme

Çerçeveleme ve pencereleme Bölüm 3.2'de ele alınıp incelenmiş ve Şekil 3.5'te Cerçeve.vi'nin diyagramı verilmişti. Cerçeve.vi, SesOku.vi'dan 1-boyutlu dizin halinde konuşma işaretini alır. 10 ms aralıklarla üst üste binen 256 örnekli çerçeveler üretir. Bu çerçevelerin ortalama değerlerini sıfırlar Hamming penceresinden geçirip, 2-boyutlu bir dizin olarak InciGecis'e verir (TABLO 3.3). InciGecis'e dosya ismi ve 1-boyutlu konuşma işareti de verilir.



Şekil 3.14 Homomorfik analiz (kepstrum analizi). (a) Hamming penceresinden geçirilmiş konuşma işareti; (b) çift taraflı logaritmik spektrum; (c) çift taraflı reel kepstrum. (Formantlar başka bir ω ile elde edilip grafik olarak bu şekle monte edildi).

glottal kaynakla artikülasyon filtresini toplamsal hale getirerek birbirinden ayırır¹. (3.10)'nun ters Fourier dönüşümü

$$\mathcal{F}^{-1}[\log |S(\omega)|] = \mathcal{F}^{-1}[\log |E(\omega)|] + \mathcal{F}^{-1}[\log |H(\omega)|] \quad (3.11)$$

ile kefrans domeninde glottal ve artikülasyon karakteristikleri birbirlerinden akuple olurlar. Bu dekonvolüsyon işlemi homomorfik veya kepstrum analizi diye adlandırılır. Şekil 3.11'da for döngüsünün sol üst köşesinde bulunan LabVIEW bloğu (3.10) ve (3.11) işlemlerini yerine getirir. Yalnız burada logaritma işleminden önce $|S(\omega)|$ terimine 10^{-6} eklenerek $|S(\omega)| = 0$ durumundan kaçınılmıştır. İkinci olarak logaritmik genlik spektrumundan ortalama değeri çıkarılarak logaritmik spektrumun DC bileşeni

¹ Konuşma ve konuşmacı tanıma uygulamalarında $E(\omega)$ ve $H(\omega)$ 'nın genliği ile ilgileniriz. Bu yüzden faz bilgisi, $\angle S(\omega)$ ifadesine yer verilmedi.

sıfırlanmıştır. Bu bileşen glotal+artikülasyon karakteristiklerinin ortalamasıdır. Kaynakla filtrenin çerçeve boyunca ortalamasının bizce hiç bir anlamı yoktur (Bölüm 3.3).

Şekil 3.14 `1nciGecis.vi`'la yapılan homomorfik analize bir örnektir. Şekil 3.14(a)'da 256 örnekli, Haamming Penceresinden geçmiş konuşma işaretimiz görülüyor. Hamming penceresinden dolayı çerçevenin kenarlarına doğru işaret genliğinin zayıfladığı göze çarpmaktadır. Şekil 3.14(b) ve (c)'de (3.10) ve (3.11) denklemlerinin gerçekleşmesi görülüyor. Şekil 3.14(b)'de logaritmik spektrumun (kırmızı çizgi) yüksek frekansları atılarak elde edilen F_1 ve F_2 formantları görülmektedir (yeşil çizgi).

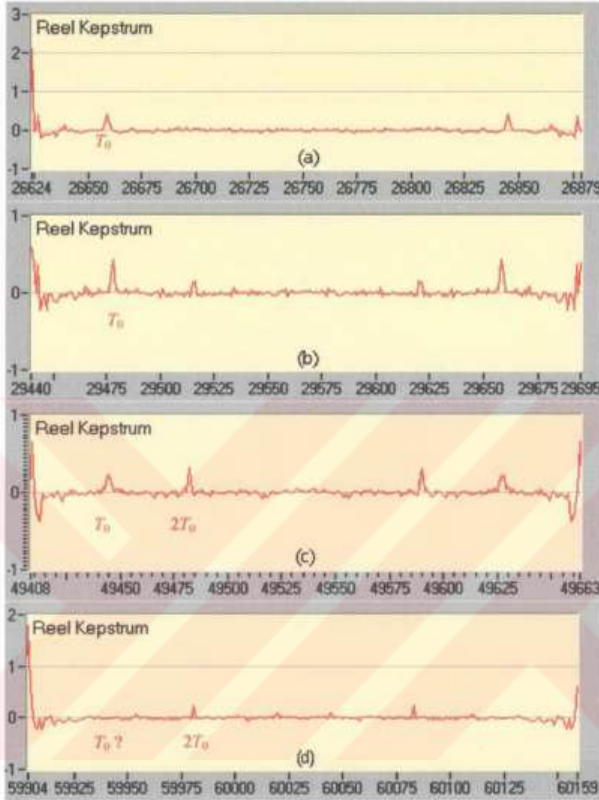
Şekil 3.14(c)'de ise öznelitik vektörlerimize temel teşkil eden (çift taraflı) kepstrum görülüyor. Bundan sonraki aşamalarda kepstrum bileşenlerinden öznelitik vektörünü oluşturacağız.

3.4.2.3 Temel periyod dedeksiyonu

Temel periyod dedeksiyonu veya temel frekans dedeksiyonu (pitch detection) çok temel ve önemli, bir o kadar da zor ve kaypak bir işlemdir. Otokorelasyon ve sıfırdan-geçiş (zero-crossing) teknikleri temel periyod bulmakta kullanılabilir. Bizim çalışmamızda kepstrum tekniği kullanılmış ve bu teknik ıslah edilmiştir.

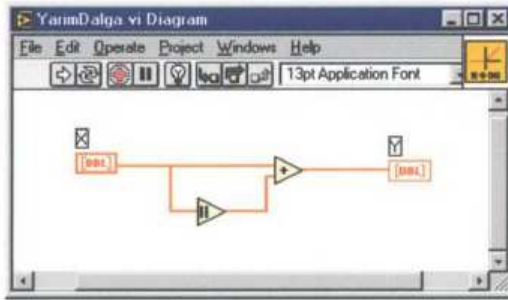
Kepstrum tekniğiyle temel periyod bulmanın iki zorluğu vardır: 1) T_0 tepesinin değişken yüksekliği, 2) bazen YPS ile gelen $2T_0$ tepesinin T_0 'dan daha yüksek olması (Şekil 3.15). Eğer eşik dedektörü kullanılacaksa eşğin çerçeveden çerçeveye adaptif olarak ayarlanması gerekir. Aksi takdirde sabit bir eşik kullanıldığı zaman düşük genlikli T_0 tepelerinin yakalanmadığı veya Şekil 3. 15(c) durumunda T_0 yerine $2T_0$ 'ın yakalandığı görülmüştür. Eşik dedektörünün bir mahzuru da yayvan T_0 tepelerinde tam tepe noktasının kefrans değerini değil, tepe şeklinin eşği geçtiği noktayı T_0 olarak bildirmesidir. Bu mahzurları ortadan kaldırmak için diferansiyel temel periyod dedektörü geliştirdik.

Diferansiyel temel periyod dedektöründen önce bütün kepstrum bileşenlerinden bir sabit (0.15) çıkarılır, yarım dalga doğrultucusundan geçirilir. Şekil 3.16'da LabVIEW ile yapılan yarım dalga doğrultucusu (`YarımDalga.vi`) görülüyor. Sonra $x'(n) = x(n) - x(n-1)$ ile verilen basit bir türev alıcı kefrans eksenini üzerinde $20 \leq n \leq 128$ için $x'(n)$ 'in işaretini izler.



Şekil 3.15 Kepstrum analizinde T_0 ve $2T_0$ 'ın değişik kombinasyonları. *fadg0\sal* cümlesinde (a) yalnız T_0 , (b) T_0 ve $2T_0$ beraber, T_0 daha büyük, (c) T_0 ve $2T_0$ beraber, T_0 daha küçük, (d) yalnız $2T_0$ görülüyor.

$x'(n)$ 'in işareti sıfırla başlar, bir tepeye rastgeldiği zaman önce pozitif sonra negatif olur. $x'(n)$ 'in pozitiften negatife döndüğü n değerinin 1 eksiği T_0 olarak kabul edilir ve tepe arama işlemine son verilir. Böylece Şekil 3.15(c)'de olduğu gibi $2T_0$ 'da T_0 'daki tepeden daha yüksek bir tepe olsa bile dikkate alınmaz. Çerçeve içinde bir tepe bulunamazsa T_0 'a -1 değeri atanır. $T_0 = -1$ değeri bir sonraki işlemde konuşmanın ünsüz kısımlarını atmak için kullanılır.



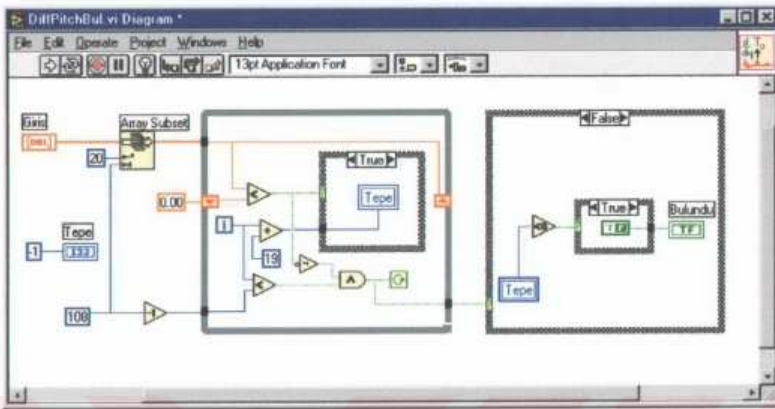
Şekil 3.16 LabVIEW ile gerçekleştirilmiş yarım dalgı doğrultucusu. X ve Y giriş ve çıkış dizinleridir. Mutlak değeri operatörü giriş olarak skalar veya vektörel kemiyetler alabilir. Transfer fonksiyonu $Y = X + |X|$.

Şekil 3.17'de DiffPitchBul görülmektedir. Sol üst köşedeki ArraySubset fonksiyonu giriş dizinin 20'den 128'e kadar olan bileşenlerini alır. WHILE döngüsü türev negatif oluncaya kadar devam eder. Türev işlemi için $x(n) - x(n-1)$ fark denklemi kullanılır. $x(n-1)$, WHILE döngüsünün kenarlarındaki (turuncu renkli (^) ve (v) ile gösterilen) shift register'de tutulur. Tepe, ilk negatif eğimin bulunduğu n değerinin bir öncesinde bulunur. Bu yüzden $Tepe = n + 20$ değil $Tepe = n + 19$ 'dur. Eğer geçerli bir tepe bulunamazsa $Tepe = -1$ olarak kalır. Sağda görülen CASE yapısı, WHILE döngüsünden gelen $Tepe$ değerine bakarak Bulundu değişkenine TRUE veya FALSE değerlerini atar. ($Tepe < 0$) ise $Bulundu = FALSE$ olur. Bu işlemleri C dilinde aşağıdaki kodla özetleyebiliriz:

```
int      i, Tepe = -1, Bulundu;
double   c[128], son;

/* Thresholding ve yarım dalgı doğrultucusu */
for(i = 0; i < 128; i++) {
    c[i] = c[i] - 0.15;
    c[i] = c[i] + abs(c[i]);
}

/* DiffPitchBul: diferansiyel temel periyod dedektörü */
son = 0.0; Tepe = -1; i = 0;
while(i < 108) {
    if (c[i] < son) { Tepe = i + 19; break; }
    son = c[i++]; /* son = c[i-1] */
}
if (Tepe == -1) Bulundu = 0; else Bulundu = 1;
...
```



Şekil 3.17 DiffPitchBul.vi diyagramı. 20 ile 128 arasındaki kepsturm bileşenleri içinde basit bir türev işlemiyle T_0 bulunur.

3.4.2.4 Öznitelik vektörü oluşturulması ve dosyalama

IncıGecis.vi'nin en son yaptığı iş konuşmanın ünlü kısımlarının öznitelik vektörlerini oluşturup dosyalamak ve bu kısımların spektrogramını göstermektir. Bu işler Şekil 3.11'deki for döngüsünün sol altında görülen CASE yapısı tarafından gerçekleştirilir. DiffPitchBul'un "Bulundu" çıkışı "doğru" ise "True CASE" yapısı aktif olur, BuildArray fonksiyonu T_0 ile 19 kepsturm bileşenini 20-boyutlu bir vektör halinde birleştirir, WriteToSpreadsheet fonksiyonu da bu vektörü dosyalar. "Bulundu" çıkışı "yanlış" ise "False CASE" yapısı çağrılır; "False CASE" boş bırakıldığı için ünsüz konuşma kısımları atılmış olur.

Dosya ismi konuşma dosyasının isminden dosya isminin uzantısı ".fir" yapılarak elde edilir. Şekil 3.10'un sol alt köşesinde for döngüsünün dışında görülen string fonksiyonları (StringLength, StringSubset, ConcatenateStrings, StringToPath) bu işlemi gerçekleştirirler. Open/Create/ReplaceFile.vi da yeni dosyayı açar ve WriteToSpreadsheet.vi'a verir.

3.4.3 İkinci geçiş

T_0 ve $2T_0$ tepeleri birlikte bulundukları zaman birinci geçiş analizi T_0 'ı bulabilir. Ancak Şekil 3. 15(d)'de görüldüğü gibi sadece $2T_0$ 'da tepe bulunuyorsa bu değer öznelilik olarak dosyaya kaydedilir. Bu da kadın konuşmacılarda konuşmanın bazı kısımlarının erkek konuşması olarak kaydedilmesi demektir. Böyle bir durum Kohonen ağıının eğitiminde öznelilik haritalarının hatalı oluşmasına, sınıflama esnasında da bilinmeyen konuşmanın yanlış sınıfa atanmasına sebep olur.

İkinci geçiş analizi, konuşma boyunca T_0 'ın μ_{T_0} ortalama değerini ve σ_{T_0} standard sapmasını hesaplar.

$$|T_0 - \mu_{T_0}| \leq 1.75\sigma_{T_0} \quad (3.12)$$

Şartını sağlayan temel periyot değerleri Kohonen ağıının eğitiminde kullanılmak üzere tutulur. (3.12)'yi sağlamayan T_0 değerleri atılır. Bölüm 3.3'te tartışıldığı üzere (3.12)'yi sağlayan T_0 değerleri, öznelilik dosyasına bir k sabitiyle çarpılarak kaydedilir. 2nciGecis sanal cihazı (Şekil 3.11), konuşma boyunca T_0 'ı, T_0 'ın ortalama değerini ve standard sapmasını, ayrıca k sabitini de aynı pano üzerinde göstermektedir. k 'nın değeri sürgülü kumanda ile $0 \leq k \leq 1$ şeklinde değiştirilir.

2nciGecis sanal cihazının diyagramı Şekil 3.19'da verilmiştir. Bu sanal cihaz üç ayrı SEQUENCE yapısından meydana oluşmuştur. 0 numaralı SEQUENCE yapısı içindeki PitchStat.vi 1nciGecis sanal cihazının yazdığı bir konuşma öznelilik dosyasını açar, T_0 'ın ortalama değerini ve standart sapmasını hesaplar. Bir sonraki bloklara dosya yolunu, 2-boyutlu dizin halinde öznelilikleri, μ_{T_0} ve $1.75\sigma_{T_0}$ değerlerini verir.

1 numaralı SEQUENCE (3.12) şartını sağlayan çerçevelerde $T_0 \leftarrow kT_0$ değişikliği yaparak 2-boyutlu yeni bir öznelilik matrisi oluşturur. Bu şartı sağlamayan çerçeveler matrise dahil edilmez.

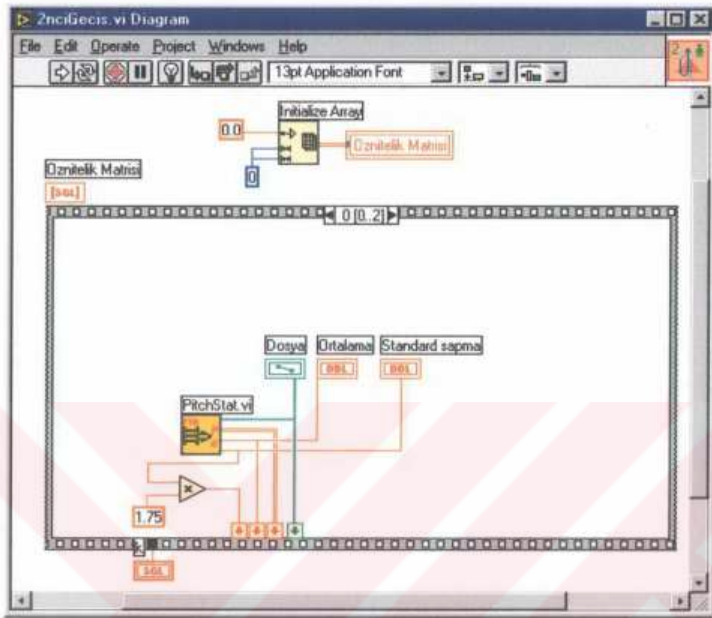
2 numaralı SEQUENCE öznelilik matrisini spektrogram olarak sunar ve "ftr" uzantısıyla dosyalar.



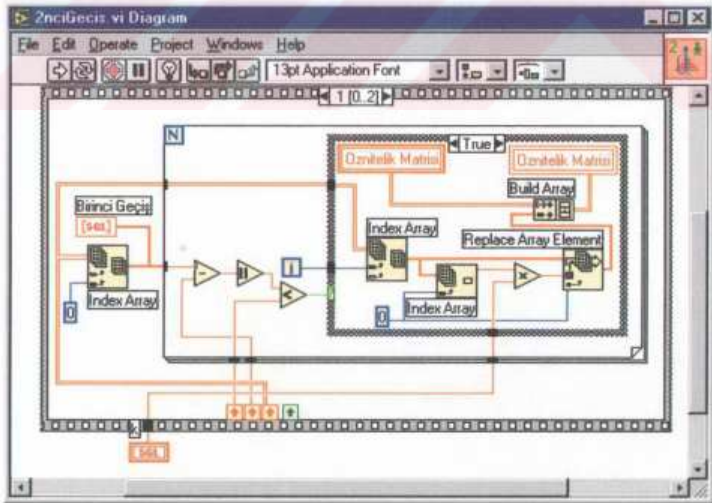
Şekil 3.18 2nciGecis sanal cihazının panosu. Üst görüntü 1nciGecis sanal cihazının çıkışıdır. İkinci geçişte YPS'den kaynaklanan $2T_0$ tepelerinin ayıklandıkları görülmektedir.

2nciGecis sanal cihazının panosunda birinci geçiş ve ikinci geçiş sonuçları karşılaştırmak amacıyla alt alta gösterilmiştir. İkinci geçişte YPS kaynaklı $2T_0$ kayıtlarının silindiği görülmektedir.

Bu bölümde elde edilen öznelilikler dördüncü bölümde "Ayrık SOM Yapay Sinir Ağlarının" eğitiminde ve sınıflamada kullanılacaktır. Yine en iyi konuşmacı tanıma için glotal uyarım faktörü k 'nın sahip olması gereken değer, k 'ya farklı değerler verilerek yapılacak deneylerden sonra belirlenecektir.

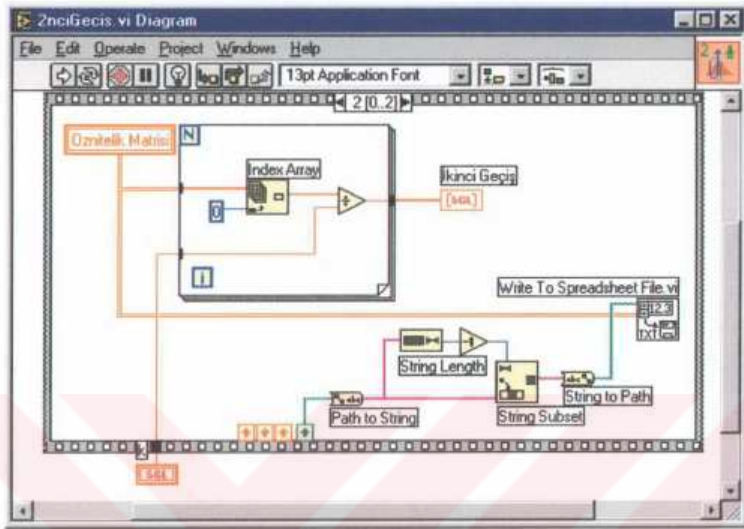


(a)



(b)

Şekil 3.19 (a), (b) ZnciGecis.vi diyagramı.



Şekil 3.19 (c) 2nciGecis.vi diyagramı (devam).

4 EĞİTİM ve TANIMA

4.1 Giriş

Bu bölümde iki şeyin gerçekleştirilmesi hedeflenmiştir: 1) SOM yapay sinir ağının konuşmacı özneliklerini öğrenebildiğini ve buna dayanarak bir konuşmanın kimin tarafından söylendiğini tahmin edilebileceğimizi göstermek, 2) Tanımaya esas olan glotal ve artikülasyon parametreleri arasında optimum bir tanıma için bulunması gereken "karışım oranını" belirlemek.

Bölüm 2'de SOM yapay sinir ağının beynin korteksindeki ilgili öğrenme merkezlerinin çalışmasını esas aldığı söylenmişti. Bu bölümde her konuşmacı için işitme korteksinde ayrılmış bir bölge olduğu ve bir konuşmacıyı dinlerken **ona** ait kortikal bölgenin sinir hücrelerinde öğrenme sürecinin olduğu kabul edilmiştir. Bu kabul, adı geçen öğrenme merkezlerinin fiziki olarak birbirleriyle örtüşmedikleri anlamına gelmektedir. Bu korteks bölgelerine karşılık olmak üzere oniki TIMIT konuşmacısı için oniki kod defteri oluşturulacaktır. Bu kod defterleri, kortikal öğrenme bölgelerinin YSA karşılığıdır. **Her konuşmacı birbirinden tamamen farklı seslerle eğitilecek**; eğitimden sonra eğitim kümesi içinden veya dışından bilinmeyen bir konuşmanın kime ait olduğu tahmin edilecektir.

Bölüm 3.3'te konuşma işaretinden kişiliği teşhis ederken, hem glotal hem de artikülasyon filtresinin özelliklerinin kullanıldığı söylenmişti. Bu bağlamda 20 elemanlı bir öznelik vektörü tanımlanmıştı. Burada cevaplanması gereken soru şudur: *"Acaba kimin konuştuğunu anlama sürecinde beynimiz glotal uyarımı ne oranda, artikülasyon filtresi parametrelerini ne oranda kullanmaktadır? Buna paralel olarak bizim geliştirdiğimiz tanıma sistemi, bu iki bilgi kaynağı arasında nasıl bir denge sağlamalıdır?"* İşte öznelik vektörünün kT_0 elemanındaki k skala faktörü bu sorudan doğmuştur. Bu bölümde bu k faktörü belirlenecektir.

TABLO 4.1 KONUŞMACILARA AİT EĞİTİM KONUŞMALARI

Konuşmacı	tn.dat: Eğitim konuşması içeriği	Çerçeve sayısı
fadq0	<i>sx199, sx379, sx109, sx289, sx19</i>	835
faem0	<i>sx312, sx42, sx222, sx402, sx132</i>	696
faks0	<i>sx43, sx223, sx403, sx133, sx313</i>	751
fbas0	<i>sx307, sx37, sx217, sx397, sx127</i>	487
mabc0	<i>sx331, sx61, sx241, sx421, sx151</i>	336
mabw0	<i>sx134, sx314, sx44, sx224, sx404</i>	609
madc0	<i>sx287, sx17, sx197, sx377, sx107</i>	650
maeb0	<i>sx90, sx270, sx450, sx180, sx360</i>	702
maeo0	<i>sx246, sx426, sx156, sx336, sx66</i>	768
maf m0	<i>sx39, sx219, sx399, sx129, sx309</i>	711
mahh0	<i>sx214, sx394, sx124, sx304, sx34</i>	1273
mbef0	<i>sx201, sx381, sx111, sx291, sx21</i>	482

Eğitimde kullanılan konuşmalar bazı konuşmacılar için son derece başarılı tanıma dereceleri sağlarken, bazı konuşmalarda sınıflamada sürekli hata yapıldığı görülmüştür. Bunun anlamı gayet açıktır: hata yapılan bu konuşmaların içerdiği bilgi eğitim esnasında yapay sinir ağı tarafından öğrenilememiştir; bu konuşmacılara ait eğitim konuşmaları istatistik bilgi bakımından noksandır. Bu olumsuzluğu gidermek için, *k* faktörü belirlendikten sonra, hata yapılan konuşmalar eğitim kümesine dahil edilerek hata yapılan konuşmacıların kod defterleri yeniden eğitime tabi tutulmuştur. Böylece hata oranı %5'ten %1.7'ye düşürülmüştür. Bu takviyeli eğitim **istatistik zenginleştirme** başlığı altında ele alınacaktır.

Bölüm 4.6'da tanıma amacıyla geliştirdiğimiz *Tanı . vi*'den bahsedilecektir.

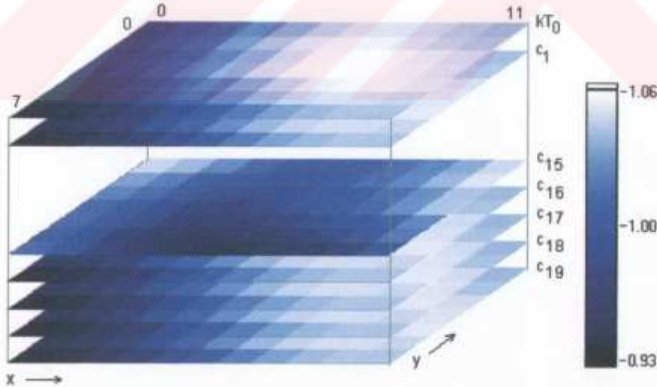
4.2 Eğitim

Konuşmacılarımız TIMIT konuşma tabanından seçilen dört kadın ve sekiz erkekten oluşmaktadır. Eğitim konuşmaları her konuşmacının *sx* cümlelerinin birleşiminden oluşur.

TABLO 4.1'de her konuşmacıya ait xx cümleleri verilmiştir. Bu cümlelerin birleşimi txn diye isimlendirilmiştir. EK1'de konuşmacılarla ilgili ayrıntılı bilgi verilmiştir.

Her konuşmacı için tanımaya esas olacak bir kod defteri gereklidir. Kod defterleri değişik algoritmalarla elde edilebilir: LBG (Linde-Buzo-Gray) algoritması, Lloyd veya K-ortalamalar algoritması gibi. Elde edilen kod defteri eğitim verilerinin vektör kuantalamasıdır. Bizim kod defterlerimiz biyolojik motivasyonlara dayanarak Kohonen yapay sinir ağı ile elde edilmiştir. Elde edilen kod defteri SOM haritasıdır.

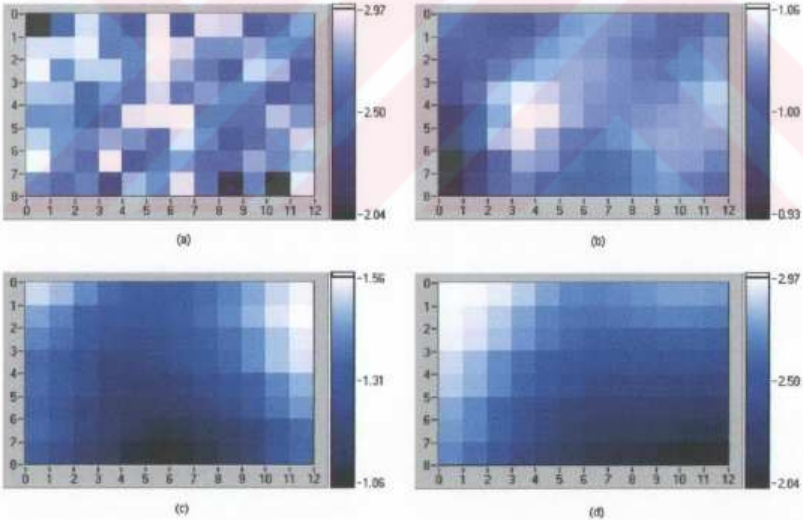
Çalışmamızda kod defterinin boyutları 8×12 olarak alınmıştır. Böylece her konuşmacı için 96 nöronlu (dügümlü) bir Kohonen yapay sinir ağı kullanılmıştır. Öznitelik vektörlerimizin 20 elemanlı oldukları da göz önüne alındığı zaman SOM haritası Şekil 4.1'deki gibi canlandırılabilir. LabVIEW ile altıgen komşuluğu grafik olarak gösteremediğimiz için kod defterini dikdörtgenler prizması şeklinde görüntüledik. Şekilde her nöron 20 elemanlı bir vektördür. En üst kt_0 katmanındır, daha sonra $c_1, \dots, c_{19}$ kepstrum katsayılarından her biri için bir katman gelmektedir.



Şekil 4.1 Eğitimden sonra bir konuşmacı için oluşturulan SOM haritası. SOM haritası konuşmacıyı tanıırken kod defteri olarak kullanılır. SOM algoritması konuşmacının statik ve dinamik özelliklerine göre haritanın nöron değerlerini bir düzene sokar. Nöronların renk tonları sahip oldukları değerlere tekabül etmektedir.

SOM haritasını elde etmek için Helsinki Teknoloji Üniversitesi SOM Programlama Grubu'nun dağıttığı SOM_PAK yazılımını kullandık. SOM_PAK ile ilgili ayrıntılar EK2'de verilmiştir. Haritamız ve eğitime ait teknik özellikler şunlardır:

1. Harita büyüklüğü: 8 x 12 nöron
2. Başlangıç vektörleri: rastgele
3. Komşuluk: altıgen
4. Birinci öteleme sayısı: 2000
5. Birinci öteleme başlama yarıçapı: 10 nöron (Gauss)
6. Birinci öteleme öğrenme hızı: 0.05
7. İkinci öteleme sayısı: 15000
8. İkinci öteleme başlama yarıçapı: 3 nöron (Gauss)
9. İkinci öteleme öğrenme hızı: 0.03



Şekil 4.2 Bazı SOM haritalarının kt_0 katmanları. (a) Eğitmeden önce nöronlara rastgele değerler atanmış, (b) *fadg0*, (c) *faem0* ve (d) *mabc0* konuşmacılarının eğitim sonrası kt_0 katmanları. Konuşmacılar arasındaki dağılım farklılıkları konuşmaların statik ve dinamik özellikleri ile ilgilidir.

Bütün konuşmacılarda trn.dat dosyaları eğitim konuşmalarını içerir. vsom.exe programı SOM eğitimi için trn.dat dosyalarını okur (EK2). MS-DOS ortamında cereyan eden bu eğitimin sonunda her konuşmacı için bir cod dosyası oluşur. cod dosyaları Şekil 4.1 ve Şekil 4.2'de görülen SOM haritalarından başka bir şey değildir. SOM eğitimi için yazılan somtrain.bat dosyasında ilk iki konuşmacı (*fadg0* ve *faem0*)'ın eğitimi ile ilgili satırlar şöyledir:

```
randinit -xdim 12 -ydim 8 -din e:\ses\fadg0\trn.dat -
cout fadg0.cod -topol hexa -neigh gaussian -rand
123
vsom -buffer 250 -din e:\ses\fadg0\trn.dat -cin
fadg0.cod -cout fadg0.cod -rlen 2000 -alpha 0.05 -
radius 10
vsom -buffer 250 -din e:\ses\fadg0\trn.dat -cin
fadg0.cod -cout fadg0.cod -rlen 15000 -alpha 0.03 -
radius 3
randinit -xdim 12 -ydim 8 -din e:\ses\faem0\trn.dat -
cout faem0.cod -topol hexa -neigh gaussian -rand
123
vsom -buffer 250 -din e:\ses\faem0\trn.dat -cin
faem0.cod -cout faem0.cod -rlen 2000 -alpha 0.05 -
radius 10
vsom -buffer 250 -din e:\ses\faem0\trn.dat -cin
faem0.cod -cout faem0.cod -rlen 15000 -alpha 0.03 -
radius 3
```

Yukarıdaki kumandalarda Sayfa 68'de belirtilen teknik özellikler görülebilir. Birinci satır fadg0.cod için hazırlık, ikinci ve üçüncü satırlar SOM eğitiminin "kaba" ve "ince" ayarlarını sağlamaktadır. Diğer konuşmacılar için de durum aynıdır. **Eğitim sonunda SOM'lar sahiplerine ait özneliliklerin istatistik dağılımını öğrenmiş olurlar.**

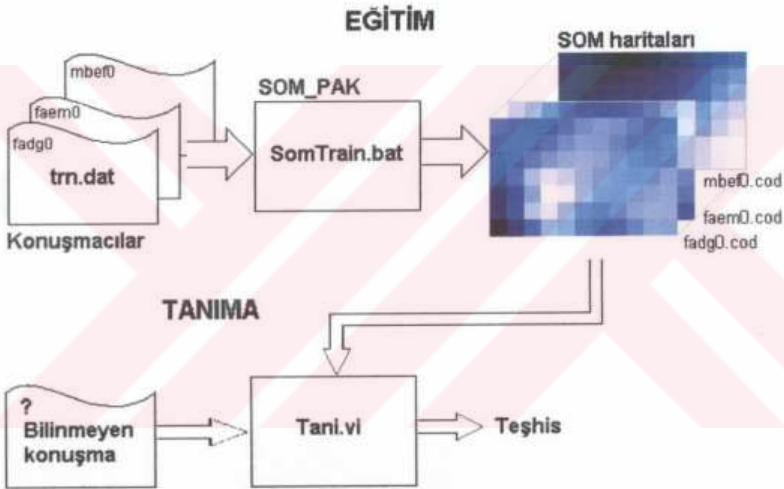
Şekil 4.3 eğitim işlemini göstermektedir. Şekilde SOM haritaları kolaylık amacıyla sadece tek katmanlı olarak çizilmiştir

4.3 Tanıma Algoritması

Şekil 4.3'te görüldüğü gibi bilinmeyen bir konuşmanın sahibini tahmin etmek için eğitimle elde edilen SOM haritaları kullanılır. KNN sınıflama algoritması SOM haritalarını kullanarak bilinmeyen konuşmayı sınıflar. Bunun için Tani.vi isimli bir LabVIEW sanal cihazı geliştirilmiştir. Tanıma felsefemiz şudur: **Bilinmeyen konuşmanın istatistiği**

SOM'larda temsil edilen konuşma istatistiklerinden hangisine en fazla uyarsa bu konuşmayı yapan muhtemelen o SOM'un sahibidir. Burada muhtemelen deyiimiyle, tanımaya bir olasılık dahilinde varıldığını vurgulamak istiyoruz.

Yukarıdaki paragrafta bahsedilen "uyarsa" ifadesi bilinmeyen konuşma ile SOM'lar arasında bir çeşit karşılaştırma işlemi yapıldığını anlatmaktadır. Bu karşılaştırmanın matematik karşılığı uzaklık hesaplamadır. Literatürde pek çok uzaklık fonksiyonu tarif edilmiştir. Ancak biz (3.6) ile tanımlanan Euklid uzaklığını kullandık.



Şekil 4.3 Eğitim ve tanıma işlemleri. SomTrain bütün konuşmacıların SOM haritalarını elde eder. Tani.vi sanal cihazı SOM haritalarına yakınlıklarını kullanarak bilinmeyen bir konuşmanın sahibini teşhis eder.

Tanıma işleminde her konuşmacı için puantaj (skor) tutulur. Tanıma işlemine başlandığında konuşmacıların puanları sıfırlanır. Tani.vi bilinmeyen konuşmayı çerçeve çerçeve 12 adet SOM ile karşılaştırır: Önce ilk çerçevenin, fadg0.cod SOM'unun 96 kod vektörüne uzaklıkları hesaplanır. Bunların minimumu o çerçevenin fadg0.cod SOM'una uzaklığı olarak kabul edilir. Sonra birinci çerçevenin diğer SOM'lara (faem0.cod,..., mbe0.cod)'a uzaklıkları hesaplanır. Birinci çerçeve hangi SOM'a en

yakın çıkarsa o konuşmacının puanı bir arttırılır. Sonra diğer çerçevelerde aynı işlem tekrar edilir. Son çerçeveden sonra maksimum puan alan konuşmacı o konuşmayı yapan kişi olarak teşhis edilir.

Konuşmacıların puanları $n[j]$, $j = 0,1,...,11$ ile gösterilirse, tanımaya başlarken $n[j] = 0$ yapılır. Bilinmeyen konuşma $f_0, f_1,..., f_M$ çerçevelerini içersin,

C_j ($j = 0,1,...,11$) kod defterleri, C_{jk} ($k = 0,1,...,95$) da j-nci kod defterinin k-ncı kod vektörü olsun. Yukarıdaki paragrafta anlatılan karar verme işlemi f_i çerçevesi için

$$m_i = \arg \min_j d(f_i, C_{jk}) \quad j = 0,1,...,11; k = 0,1,...,95 \quad (4.1)$$

ile ifade edilebilir. m karar verilen kod defterinin indisi; $d(f_i, C_{jk})$ ise f_i ile C_{jk} arasındaki (Euklid) uzaklığıdır. Konuşmacı indisi m böylece bulununca m-nci konuşmacının puanı bir arttırılır:

$$n[m_i] = n[m_i] + 1 \quad (4.2)$$

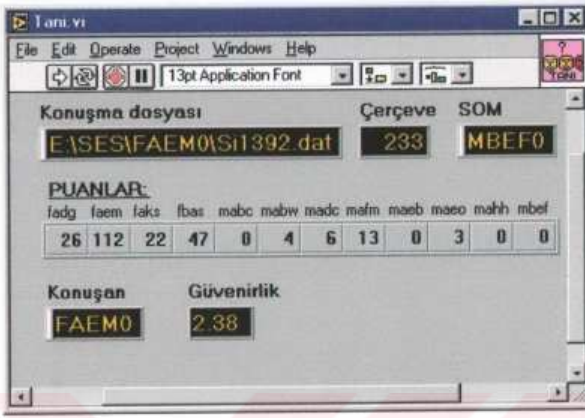
Bütün çerçeveler bitinceye kadar (4.1) ve (4.2) uygulanır. Son çerçeveden sonra bilinmeyen konuşma

$$s = \arg \max_m n[m] \quad m = 0,1,...,11 \quad (4.3)$$

ifadesiyle sınıflandırılır.

Örnek olarak Şekil 4.4'te bir sesin *Tani.vi* sanal cihazı ile sınıflandırılması görülmektedir Her konuşmacının aldığı puanlar sanal cihaz üzerinde gösterilmiştir. *si1392* konuşmasını, toplam 233 çerçeveden 112'sini kazanan *faemf*ın yaptığına karar verilmiştir.

Şekildeki *güvenirlilik oranı* (confidence ratio) en büyük puanın kendisinden sonra gelen puana bölümü ile elde edilir: $112/47 = 2.38$. Güvenirlilik oranı Bölüm 4.4'te kullanılacaktır.



Şekil 4.4 Tani.vi ile konuşmacı tanıma. En üstte tanımak istenen konuşma ve toplam çerçeve sayısı, ortada konuşmacıların aldıkları puanlar, en altta ise teşhis edilen konuşmacının ismi görüldüyor. En fazla puan alan *faem0* 2.38'lik bir güvenilirlik oranı ile konuşmacı ilan edilmiştir.

4.4 Glotal Uyarım Katsayısının Belirlenmesi

Bölüm 3.3'te $[kT_0, c_1, \dots, c_{10}]$ 'un öznitelik vektörü olarak seçildiği belirtilmişti. Burada k bir kazanç faktörü olarak glotal uyarımın miktarını değiştirir. Şüphesiz eğitimle oluşan **SOM haritaları ve buna bağlı olarak tanıma performansı k değerine bağlılık gösterir.** Bölüm 4.1'de sorduğumuz "En iyi tanıma için glotal uyarım faktörü ne olmalıdır?" sorusuna bu bölümde cevap vereceğiz.

M çerçeveli bir S konuşması için bir G güvenilirlik oranı tanımlayabiliriz. A ve B en yüksek puan alan iki konuşmacı ve A'nın puanı B'nin puanından yüksekse, güvenilirlik oranı S konuşmasını A ve B'nin yapma olasılıklarının oranıdır:

$$G = \frac{n[A|S]}{n[B|S]} = \frac{n[A|S]/M}{n[B|S]/M} = \frac{P(A|S)}{P(B|S)} \quad (4.4)$$

$n[A|S]$ ve $n[B|S]$, S konuşması için A ve B'nin puanları; $P(A|S)$ ve $P(B|S)$ de S konuşması verildiğinde A ve B'nin şartlı olasılıklarıdır. **G ne kadar yüksek olursa tanıma o kadar güvenlidir** denir. $G = 1$ iki konuşmacının da aynı puanı almaları halidir ki burada kararsızlık söz konusudur. Eğer $G < 1$ ise o zaman yanlış sınıflama durumu ortaya çıkar.

TABLO 4.2 GÜVENİRLİK FAKTÖRÜNÜN SKALA İLE DEĞİŞİMİ

(Tablonun tamamı için EK3'e bakınız)

KONUŞMALAR	SKALA FAKTÖRÜ, k									
	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80	
FADGO\Sa1	1.02	1.32	2.02	2.19	1.94	2.33	1.62	2.06	2.12	
FADGO\Sa2	1.41	4.28	4.52	2.8	3.32	2.39	1.95	2.34	1.7	
FADGO\S11279	2.26	3.56	2.19	8.44	11.29	8.89	3.94	3.3	3.14	
FAEMO\S1762	1.18	1.29	2.22	2.17	2.06		1.5	1.63	1.58	
MARCO\S11620	3.17	15						13	28	
MARCO\S12041	2	2	2.56	3.25	2.5	3.83	4.8	4.36	5.56	
MADCO\Sa1	1.87	2.46	2.29	1.05	1.02	1.1	1.22	1.1	1.17	
MADCO\Sa2	7.67	7.64	6.81	2.83	2.7	3.28	3.61	3.28	2.29	
MADCO\Sx287	15	17.43	9.67	10.09	9.5	8.64	7.93	6.62	6.12	
MAERO\Sa1	4.85	4.78	4.94	3.9	5.44	5.21	4.18	4.35	2.64	
MAERO\Sa2	2.13	2.18	1.84	1.17	1.69	1.39	1.45	1.64	2.25	
MAERO\S11411	3.2	4.43	3.61	3.35	3	2.47	3.02	2.58	2.11	
MAERO\S12250	5.36	5.83	7.3	5.67	3.75	4.44	5.77	6.27	3.67	
MAERO\S1990	1.12	1.26	1.19	1.14	1.17	1.13	1.05	1.08	1	
MAERO\Sx180	9.67	13	19.6	9.2	11.5	8.27	11.75	8.7	7.33	
MAEOO\Sa1	1.13	1.27	1.88	1	1.08	1.07	1.14	1.56	1.59	
MAEOO\Sa2	1.49	2	3.4	1.18	1.75	1.63	1.55	1.05	1.5	
MAEOO\S11326	1.33	1.34	1.54	1.16	1.07	1.3	1.33	1.15	1.33	
MAEOO\S11655	2.11	1.23	1.13	1.33	1.31	1.25	1.27	1.92	1.75	
MAEOO\S11956	1.04	2	1.24					1.14		
MAEOO\Sx156	5.54	5.46	5.17	4.91	4.25	3.27	3.29	2.92	2.5	
MAEOO\Sx246	4.22	2	1.53	1.56	1.75	1	1.35	1.62	2	
MAFMO\S11569	1.25	3.18	1.95	1.67	1.4	1.14	1.11	1.25	1.03	
MAHHO\Sa1	1.67	2	3.32	4.54	4.33	5.09	5.32	3.06	2.26	
MAHHO\Sa2	1.31	1.28	1.53	1.69	1.5	1.1	1.15	1.14	1.14	
MAHHO\S11294	1.14	1.25	2.31	2.34	2.32	2.08	1.87	1.38	1.44	
MAHHO\S11664	1	2	2.67		1.03	1.21	1	1.17	1.28	
MAHHO\S11924	1.82	1.22	1.09				1.15	1	1.13	
MAHHO\Sx124	2.95	3.88	2.85	2.75	2.5	1.55	2.43	3.53	3.81	
MAHHO\Sx214	2.41	5.35	4.81	4.95	4.26	3.41	2.89	2.23	2.69	
MBEFO\Sa1	9.09	9.78	3.25	1.67	1.83	1.8	1.26	1.16	1.05	
MBEFO\Sa2	2.17	2.5	3.13	3.92	2.94	6.8	5	3.71	2.7	
MBEFO\S11281	3.36	2.58	2.24	2.25	1.87	1.24	1.28	1.06	1.03	
MBEFO\S11911	2.32	3.06	1.04	1.17	1.03	1	1.1	1.27	1.35	
MBEFO\S1651	5.53	4.09	3.75	1.53	1.55	1.08		1.3	1.19	
MBEFO\Sx111	4.57	5	7	10.33	7	3.75	2.56	2.2	1.58	
Hatalar	7	6	6	9	9	9	9	8	9	

12 TIMIT konuşmacısının bütün cümleleri ile k 'nın değişik değerleri için tanıma deneyleri yaptık (EK3). Güvenirlik oranı G her cümle için, k 'nın değişik değerlerine karşı TABLO E3-1 'de gösterilmiştir. TABLO E3-1'den alınan bazı sonuç örnekleri TABLO 4.2'de verilmiştir. Tabloda sarıya boyanan hücreler hatalı tanımları gösterir. Tablonun en alt satırında her k değeri için kaç hata yapıldığı görülmektedir. Bu deneylerin bazısında güvenirlilik faktörü " ∞ " olmakta bazılarında ise 1'e yaklaşmaktadır. Hatalı hücrelerdeki 1'den büyük G değerleri geçersizdir. Minimum tanıma hatası %5 (120 konuşmada 6 konuşma) olarak gerçekleşmiştir.

TABLO E3-1'deki değerlerden en iyi tanımayı gerçekleştiren k değerini bulmak için basit bir performans fonksiyonu geliştirmek faydalı olacaktır. Bu fonksiyonda kullanılmak üzere, verilen bir S konuşması ve k değeri için bir a ağırlığı tarif edelim:

$$a(k, S) = \begin{cases} -1 & \text{hatalı karar} \\ 2^{-d} & d = 0, 1, 2 \\ 0 & \text{diğer durumlarda} \end{cases} \quad (4.5)$$

(4.5) ifadesinde d tablodaki bir hücre ile $G_{\max}$ hücresi arasındaki uzaklıktır. S konuşması için k 'nın artan değerleriyle tabloda yatay olarak ilerlerken belli bir k değeri için G maksimum olmaktadır. $G_{\max}$ değerine en yakın olan hücrenin uzaklığı $d = 1$, yakınlıkta ondan sonra gelen hücrenin uzaklığı $d = 2$ 'dir. Tablodaki sarı hücrelerde $a(k, S)$ 'ye -1 atanarak performans fonksiyonunda hatalı sınıflamaya "ceza" unsuru getirilmiştir. Daha uzak hücrelerin performansa etkileri olmadıklarını kabulle $a(k, S)$ değerleri 0 alınmıştır. Örnek olarak $sx199$ konuşmasının G değerlerini alalım:

k	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80
$fadgl/sx199$	2.51	3.67	6.21	4.74	4.35	2.9	3.11	2.79	2.26
$a(k, sx199)$	0	0	1.00	0.50	0.25	0	0	0	0

Burada $k = 0.03$ olduğunda $G = G_{\max} = 6.21$ olmaktadır. Bu hücrede güvenilirlik $G_{\max}$ 'a eşit olduğu için $d = 0$ ve $a(0.03, sx199) = 1$ olur. Maksimuma en yakın hücreler $k=0.05$ ve $k=0.06$ değerlerinde görülür. Bu hücreler için $G(0.05) = 4.74$, $G(0.06) = 4.35$ olarak okunmaktadır. Dolayısıyla $a(0.05, sx199) = 0.5$ ve $a(0.06, sx199) = 0.25$ olur.

$a(k, S)$ böylece tarif edildikten sonra şimdi k için bir performans fonksiyonu tarif edebiliriz:

$$p(k) = \sum_i a(k, S_i) \quad (4.6)$$

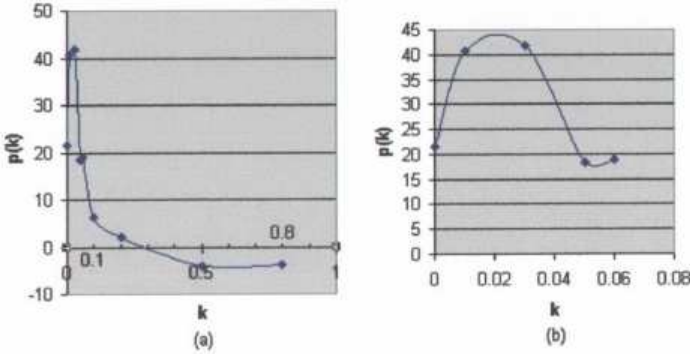
Denklem (4.6)'daki " i " konuşma indeksidir. Toplam tablodaki bütün konuşmaları kapsar. Böylece her k sütunu için bir performans değeri elde edilmiş olur. Aranan k faktörü (4.6) ifadesini maksimum yapan değerdir:

$$k_{\text{en iyi}} = \arg \max_k [p(k)] \quad (4.7)$$

TABLO 4.3 değişik k değerlerine için elde edilen performans değerlerini vermektedir. Tablonun en alt satırında $p(k)$ değerlerinden en yüksek tanıma performansının $k = 0.03$ civarında gerçekleştiği anlaşılmaktadır. $p(k)$ 'nin grafiği Şekil 4.5'te gösterilmiştir.

TABLO 4.3. TANIMA PERFORMANSININ SKALA FAKTÖRÜ ile DEĞİŞİMİ

	SKALA FAKTÖRÜ, k									
KONUŞMALAR	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80	
FADGO\sa1	-1.00			0.50		1.00				0.25
FADGO\sa2		0.50	1.00		0.25					
FADGO\si1279				0.25	1.00	0.50				
FAEMO\si1762			1.00	0.50	0.25	-1.00				
MAHCO\si1620		0.25	1.00	1.00	1.00	1.00	1.00			0.50
MAHCO\si2041							0.50	0.25	1.00	
MADCO\sa1	0.25	1.00	0.50		-1.00					
MADCO\sa2	1.00	0.50	0.25							
MADCO\si1367	1.00	0.50	0.25							
MAEBO\sa1			0.25		1.00	0.50				
MAEBO\sa2	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	
MAEBO\si1411		1.00	0.50	0.25						0.50
MAEBO\si2250		0.25	1.00							
MAEBO\si1990		1.00	0.50		0.25		-1.00	-1.00	-1.00	
MAEBO\sx180		0.50	1.00				0.25			
MAEBO\sa1	-1.00		1.00	-1.00			-1.00	0.25	0.50	
MAEBO\sa2		0.50	1.00		0.25					
MAEBO\si1326	0.25	0.50	1.00	-1.00	-1.00	-1.00	-1.00			0.25
MAEBO\si1655	-1.00	-1.00	-1.00	-1.00	-1.00		0.25	1.00	0.50	
MAEBO\si1956	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	
MAEBO\sx156	1.00	0.50	0.25							
MAEBO\sx246	1.00	0.50			0.25	-1.00	-1.00	-1.00	-1.00	
MAFMO\si1569		1.00	0.50	0.25						-1.00
MAHHO\sa1				0.25		0.50	1.00			
MAHHO\sa2	-1.00	-1.00	-1.00	-1.00	-1.00	0.50	1.00	-1.00	-1.00	
MAHHO\si1294			0.25	1.00	0.50					
MAHHO\si664		-1.00	-1.00	-1.00	-1.00	-1.00		-1.00	-1.00	
MAHHO\si1924	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	
MAHHO\sx124		1.00							0.25	0.50
MAHHO\sx214		1.00	0.25	0.50						
MBEFO\sa1	0.50	1.00	0.25							
MBEFO\sa2				0.25		1.00	0.50			
MBEFO\si1201	1.00	0.50		0.25		-1.00	-1.00			
MBEFO\si1911	0.50	1.00	0.25	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	
MBEFO\si651	1.00	0.50	0.25							
MBEFO\si111		0.25	0.50	1.00	0.50					
PERFORMANS, $p(k)$	21.50	40.75	41.75	18.50	19.00	6.25	2.25	-4.00	-3.75	



Şekil 4.5 Performansın k faktörü ile değişimi. (a) bütün k değerleri, (b) $0 \leq k \leq 0.6$. $p(k)$ 'nin genel davranışının belirginleşmesi için k değerlerinin arasında interpolasyon uygulanmıştır.

Şekil 4.5(a), k 'nin küçük değerleri için Şekil 4.5(b)'de açılmıştır. Burada $p(k)$ 'nin maksimal davranışı daha açık olarak görülebilmektedir. Deneyler arasındaki davranış interpolasyonla elde edilmiştir. Sonuç olarak en iyi tanıma için öznitelik vektörümüzün

$$[0.03 \ T_0, c_1, \dots, c_{19}] \quad (4.8)$$

olarak seçilmesi gerekmektedir.

4.5 İstatistik Zenginleştirme

Dikkat edilirse TABLO 4.2'de k faktörü ne olursa olsun *maeb0\sa2*, *maeo0\si1956* ve *mahh0\si1924* konuşmaları hatalı sınıflandırılmaktadır. *mahh0\si664* dokuz deneyden yedisinde hatalı sınıflandırılmakta, ikisinde ise kararsızlık durumu görülmektedir. Buradan *maeb0.cod*, *maeo0.cod* ve *mahh0.cod* kod defterlerinin, yani SOM'ların, eğitim sırasında kendi konuşmacıları hakkında yeterince bilgilendirilemediklerini söyleyebiliriz. Yine *mbef0\si1911* konuşmasında $k = 0.03$ için kararsızlığa çok yakın bir teşhis elde edilmiştir; olayı insanlarla kıyaslırsak kararda bir tereddüd olduğunu söyleyebiliriz.

SOM'ların gerçekten verilen verilerden "öğrendiklerini" göstermek üzere hatalı sınıflandırılan bu konuşmalar kendi sahiplerinin *trn.dat* dosyalarına eklendi; $k = 0.03$ alınarak dört tane SOM yeniden eğitildiler ve tanıma deneyi bütün konuşmacılar için

TABLO 4.2 GÜVENİRLİK FAKTÖRÜNÜN SKALA ile DEĞİŞİMİ

(Tablonun tamamı için EK3'e bakınız)

KONUŞMALAR	SKALA FAKTÖRÜ, <i>k</i>									
	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80	
FADG0\sa1	1.02	1.32	2.02	2.19	1.94	2.23	1.62	2.06	2.12	
FADG0\sa2	1.41	4.28	4.52	2.8	3.32	2.39	1.95	2.34	1.7	
FADG0\si1279	2.26	3.56	2.19	8.44	11.29	8.89	3.94	3.3	3.14	
FAEM0\si1762	1.18	1.28	2.22	2.17	2.06		1.5	1.63	1.58	
MABCO\si1620	3.17	15	00	00	00	00	00	13	28	
MABCO\si2041	2	2	2.56	3.25	2.5	3.83	4.8	4.36	5.56	
MADCO\sa1	1.87	2.46	2.29	1.05	1.02	1.1	1.22	1.1	1.37	
MADCO\sa2	7.67	7.64	6.81	2.83	2.7	3.28	3.61	3.28	2.29	
MADCO\sa287	15	17.43	9.67	10.09	9.5	8.64	7.93	6.62	6.12	
MAEBO\sa1	4.85	4.78	4.94	3.9	5.44	5.21	4.18	4.35	2.64	
MAEBO\sa2	2.13	2.18	1.84	1.17	1.69	1.39	1.45	1.64	2.25	
MAEBO\si1411	3.2	4.43	3.61	3.35	3	2.47	3.02	2.58	2.11	
MAEBO\si2250	5.36	5.83	7.3	5.67	3.75	4.44	5.77	6.27	3.67	
MAEBO\si990	1.12	1.26	1.19	1.14	1.17	1.13	1.05	1.00	1	
MAEBO\sa180	9.67	13	19.6	9.2	11.5	8.27	11.75	8.7	7.33	
MAEO0\sa1	1.13	1.27	1.88	1	1.08	1.07	1.14	1.56	1.59	
MAEO0\sa2	1.49	2	3.4	1.18	1.75	1.63	1.55	1.05	1.5	
MAEO0\si1326	1.33	1.34	1.54	1.16	1.07	1.3	1.33	1.15	1.33	
MAEO0\si1655	2.11	1.23	1.13	1.33	1.31	1.25	1.27	1.92	1.75	
MAEO0\si1956	1.04	2	1.24					1.14		
MAEO0\sa156	5.54	5.46	5.17	4.91	4.25	3.27	3.29	2.92	2.5	
MAEO0\sa246	4.22	2	1.53	1.56	1.75	1	1.35	1.62	2	
MAFMO\si1565	1.25	3.18	1.95	1.67	1.4	1.14	1.11	1.25	1.03	
MAHNO\sa1	1.67	2	3.32	4.54	4.33	5.09	5.32	3.06	2.26	
MAHNO\sa2	1.31	1.28	1.53	1.69	1.5	1.1	1.15	1.14	1.14	
MAHNO\si1294	1.14	1.25	2.31	2.34	2.32	2.08	1.87	1.38	1.44	
MAHNO\si664	1	2	2.67		1.03	1.21	1	1.17	1.28	
MAHNO\si1924	1.82	1.22	1.09				1.15	1	1.13	
MAHNO\sa124	2.95	3.88	2.85	2.75	2.5	1.55	2.43	3.53	3.81	
MAHNO\sa214	2.41	5.35	4.81	4.95	4.26	3.41	2.89	2.23	2.69	
MBEFO\sa1	9.09	9.78	3.25	1.67	1.83	1.8	1.26	1.16	1.05	
MBEFO\sa2	2.17	2.5	3.13	3.92	2.94	6.8	5	3.71	2.7	
MBEFO\si1281	3.36	2.58	2.24	2.25	1.87	1.24	1.28	1.06	1.03	
MBEFO\si1911	2.32	3.06	1.04	1.17	1.03	1	1.1	1.27	1.35	
MBEFO\si1651	5.53	4.09	3.75	1.53	1.55	1.08		1.3	1.19	
MBEFO\sa111	4.57	5	7	10.33	7	3.75	2.56	2.2	1.58	
Hatalar	7	6	6	9	9	9	9	8	9	

12 TIMIT konuşmacısının bütün cümleleri ile *k*'nın değişik değerleri için tanıma deneyleri yaptık (EK3). Güvenirlik oranı *G* her cümle için, *k*'nın değişik değerlerine karşı TABLO E3-1 'de gösterilmiştir. TABLO E3-1'den alınan bazı sonuç örnekleri TABLO 4.2'de verilmiştir. Tabloda sarıya boyanan hücreler hatalı tanımları gösterir. Tablonun en alt satırında her *k* değeri için kaç hata yapıldığı görülmektedir. Bu deneylerin bazısında güvenirlilik faktörü "∞" olmakta bazılarında ise 1'e yaklaşmaktadır. Hatalı hücrelerdeki 1'den büyük *G* değerleri geçersizdir. Minimum tanıma hatası %5 (120 konuşmada 6 konuşma) olarak gerçekleşmiştir.

TABLO E3-1'deki değerlerden en iyi tanımayı gerçekleştiren k değerini bulmak için basit bir performans fonksiyonu geliştirmek faydalı olacaktır. Bu fonksiyonda kullanılmak üzere, verilen bir S konuşması ve k değeri için bir a ağırlığı tarif edelim:

$$a(k, S) = \begin{cases} -1 & \text{hatalı karar} \\ 2^{-d} & d = 0, 1, 2 \\ 0 & \text{diğer durumlarda} \end{cases} \quad (4.5)$$

(4.5) ifadesinde d tablodaki bir hücre ile $G_{\max}$ hücresi arasındaki uzaklıktır. S konuşması için k 'nın artan değerleriyle tabloda yatay olarak ilerlerken belli bir k değeri için G maksimum olmaktadır. $G_{\max}$ değerine en yakın olan hücrenin uzaklığı $d = 1$, yakınlıkta ondan sonra gelen hücrenin uzaklığı $d = 2$ 'dir. Tablodaki sarı hücrelerde $a(k, S)$ 'ye -1 atanarak performans fonksiyonunda hatalı sınıflamaya "ceza" unsuru getirilmiştir. Daha uzak hücrelerin performansa etkileri olmadıklarını kabulle $a(k, S)$ değerleri 0 alınmıştır. Örnek olarak $sx199$ konuşmasının G değerlerini alalım:

k	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80
$fadg\backslash sx199$	2.51	3.67	6.21	4.74	4.35	2.9	3.11	2.79	2.26
$a(k, sx199)$	0	0	1.00	0.50	0.25	0	0	0	0

Burada $k = 0.03$ olduğunda $G = G_{\max} = 6.21$ olmaktadır. Bu hücrede güvenilirlik $G_{\max}$ 'a eşit olduğu için $d = 0$ ve $a(0.03, sx199) = 1$ olur. Maksimuma en yakın hücreler $k=0.05$ ve $k=0.06$ değerlerinde görülür. Bu hücreler için $G(0.05) = 4.74$, $G(0.06) = 4.35$ olarak okunmaktadır. Dolayısıyla $a(0.05, sx199) = 0.5$ ve $a(0.06, sx199) = 0.25$ olur.

$a(k, S)$ böylece tarif edildikten sonra şimdi k için bir performans fonksiyonu tarif edebiliriz:

$$p(k) = \sum_i a(k, S_i) \quad (4.6)$$

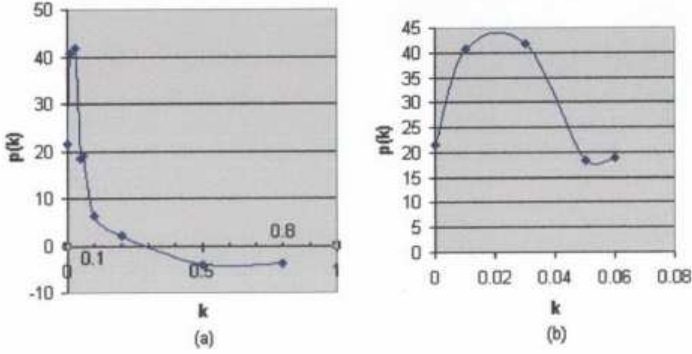
Denklem (4.6)'daki " i " konuşma indeksidir. Toplam tablodaki bütün konuşmaları kapsar. Böylece her k sütunu için bir performans değeri elde edilmiş olur. Aranılan k faktörü (4.6) ifadesini maksimum yapan değerdir:

$$k_{\text{en iyi}} = \arg \max_k [p(k)] \quad (4.7)$$

TABLO 4.3 değişik k değerlerine için elde edilen performans değerlerini vermektedir. Tablonun en alt satırında $p(k)$ değerlerinden en yüksek tanıma performansının $k = 0.03$ civarında gerçekleştiği anlaşılmaktadır. $p(k)$ 'nın grafiği Şekil 4.5'te gösterilmiştir.

TABLO 4.3. TANIMA PERFORMANSININ SKALA FAKTÖRÜ ile DEĞİŞİMİ

	SKALA FAKTÖRÜ, k								
KONUŞMALAR	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80
FADG0\sa1	-1.00			0.50		1.00			0.25
FADG0\sa2		0.50	1.00		0.25				
FADG0\si1279				0.25	1.00	0.50			
FAEM0\si762			1.00	0.50	0.25	-1.00			
MARCO\si1620		0.25	1.00	1.00	1.00	1.00	1.00		0.50
MARCO\si12041							0.50	0.25	1.00
MADCO\sa1	0.25	1.00	0.50		-1.00				
MADCO\sa2	1.00	0.50	0.25						
MADCO\si1367	1.00	0.50	0.25						
MAEBO\sa1			0.25		1.00	0.50			
MAEBO\sa2	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
MAEBO\si1411		1.00	0.50	0.25					
MAEBO\si2250			0.25	1.00				0.50	
MAEBO\si1990		1.00	0.50		0.25		-1.00	-1.00	-1.00
MAEBO\si180		0.50	1.00				0.25		
MAKOO\sa1	-1.00		1.00	-1.00			-1.00	0.25	0.50
MAKOO\sa2		0.50	1.00		0.25				
MAKOO\si1326	0.25	0.50	1.00	-1.00	-1.00	-1.00	-1.00		0.25
MAKOO\si1655	-1.00	-1.00	-1.00	-1.00	-1.00		0.25	1.00	0.50
MAKOO\si1956	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
MAKOO\si156	1.00	0.50	0.25						
MAKOO\si246	1.00	0.50			0.25	-1.00	-1.00	-1.00	-1.00
MAFMD\si1569		1.00	0.50	0.25					-1.00
MAHHO\sa1				0.25		0.50	1.00		
MAHHO\sa2	-1.00	-1.00	-1.00	-1.00	-1.00	0.50	1.00	-1.00	-1.00
MAHHO\si1294			0.25	1.00	0.50				
MAHHO\si664		-1.00	-1.00	-1.00	-1.00	-1.00		-1.00	-1.00
MAHHO\si1924	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
MAHHO\si124		1.00						0.25	0.50
MAHHO\si214		1.00	0.25	0.50					
MBEFO\sa1	0.50	1.00	0.25						
MBEFO\sa2				0.25		1.00	0.50		
MBEFO\si1281	1.00	0.50		0.25		-1.00	-1.00		
MBEFO\si1911	0.50	1.00	0.25	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
MBEFO\si1651	1.00	0.50	0.25						
MBEFO\si111		0.25	0.50	1.00	0.50				
PERFORMANS, $p(k)$	21.50	40.75	41.75	18.50	19.00	6.25	2.25	-4.00	-3.75



Şekil 4.5 Performansın k faktörü ile değişimi. (a) bütün k değerleri, (b) $0 \leq k \leq 0.6$. $p(k)$ 'nin genel davranışının belirlenmesi için k değerlerinin arasında interpolasyon uygulanmıştır.

Şekil 4.5(a), k 'nin küçük değerleri için Şekil 4.5(b)'de açılmıştır. Burada $p(k)$ 'nin maksimal davranışı daha açık olarak görülebilmektedir. Deneyler arasındaki davranış interpolasyonla elde edilmiştir. Sonuç olarak en iyi tanıma için öznitelik vektörümüzün

$$[0.03 \ T_0, c_1, \dots, c_{19}] \quad (4.8)$$

olarak seçilmesi gerekmektedir.

4.5 İstatistik Zenginleştirme

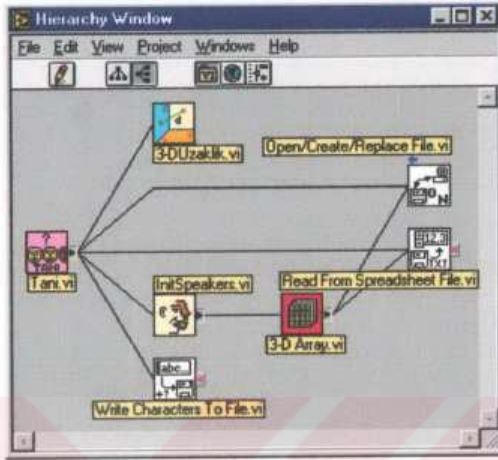
Dikkat edilirse TABLO 4.2'de k faktörü ne olursa olsun *maeb0\sa2*, *maeo0\si1956* ve *mahh0\si1924* konuşmaları hatalı sınıflandırılmaktadır. *mahh0\si664* dokuz deneyden yedisinde hatalı sınıflandırılmakta, ikisinde ise kararsızlık durumu görülmektedir. Buradan *maeb0.cod*, *maeo0.cod* ve *mahh0.cod* kod defterlerinin, yani SOM'ların, eğitim sırasında kendi konuşmacıları hakkında yeterince bilgilendirilemediklerini söyleyebiliriz. Yine *mbef0\si1911* konuşmasında $k = 0.03$ için kararsızlığa çok yakın bir teşhis elde edilmiştir; olayı insanlarla kıyaslarsak kararda bir tereddüd olduğunu söyleyebiliriz.

SOM'ların gerçekten verilen verilerden "öğrendiklerini" göstermek üzere hatalı sınıflandırılan bu konuşmalar kendi sahiplerinin *trn.dat* dosyalarına eklendi; $k = 0.03$ alınarak dört tane SOM yeniden eğitildiler ve tanıma deneyi bütün konuşmacılar için

TABLO 4.4 YENİDEN EĞİTİMDEN SONRA TANIMA PUANLARI

KONUŞANLAR	FADG0	FAEM0	FAKS0	FEAS0	MABCO	MABW0	MADCO	MAFM0	MAES0	MAEO0	MAHH0	MBEFO	TOPLAM	KARAR	G
FADG0\sa1	100	42	41	6	0	0	0	0	0	0	0	0	189	FADG0	2.38
FADG0\sa2	120	17	31	16	0	0	0	0	0	0	0	0	184	FADG0	3.87
FADG0\si1279	72	1	20	0	0	0	0	0	0	0	0	0	93	FADG0	3.6
FAEM0\si1762	12	75	44	23	0	0	0	3	0	2	0	0	159	FAEM0	1.7
MABCO\si1620	0	0	0	0	29	0	0	0	0	0	0	0	29	MABCO	∞
MABCO\si12041	0	0	0	0	44	0	0	0	0	12	4	6	66	MABCO	3.67
MADCO\sa1	0	12	0	3	24	3	62	0	24	33	14	10	185	MADCO	1.88
MADCO\sa2	0	1	0	1	6	4	127	3	14	10	9	2	177	MADCO	9.07
MADCO\sk287	0	0	0	0	9	1	123	1	1	13	4	2	154	MADCO	9.46
MAES0\sa1	0	0	0	0	2	0	1	0	107	6	19	20	155	MAES0	5.35
MAES0\sa2	0	0	0	0	4	0	1	1	69	4	3	43	125	MAES0	1.6
MAES0\si1411	0	0	0	3	11	4	3	1	127	25	30	22	226	MAES0	4.23
MAES0\si12250	0	1	0	0	2	0	2	1	80	8	3	13	110	MAES0	6.15
MAES0\si1990	0	0	0	0	17	4	12	0	56	4	26	23	142	MAES0	2.15
MAES0\sk180	0	1	0	0	1	0	1	0	98	10	4	3	118	MAES0	9.8
MAEO0\sa1	0	3	0	6	17	1	9	11	51	38	12	11	159	MAEO0	1.34
MAEO0\sa2	0	0	0	2	12	2	17	18	8	47	3	8	117	MAEO0	2.61
MAEO0\si1326	0	2	0	4	11	12	55	7	49	72	7	9	228	MAEO0	1.31
MAEO0\si1655	0	0	0	0	1	1	9	0	2	41	6	8	69	MAEO0	4.56
MAEO0\si1956	0	22	0	1	6	0	3	1	19	19	5	12	88	FAEM0	1.16
MAEO0\sk156	0	8	0	4	1	6	18	4	8	58	3	2	112	MAEO0	3.22
MAEO0\sk246	0	0	0	3	19	2	12	7	3	24	4	2	76	MAEO0	1.26
MAFM0\si1569	0	15	1	16	0	11	0	34	6	8	0	0	91	MAFM0	2.12
MAHH0\sa1	0	5	0	6	17	31	4	17	21	8	106	37	252	MAHH0	2.86
MAHH0\sa2	0	0	0	1	15	8	2	2	10	12	75	40	165	MAHH0	1.87
MAHH0\si1294	0	11	0	4	21	8	39	14	28	44	114	23	306	MAHH0	2.59
MAHH0\sk1924	0	16	0	1	3	3	8	5	0	5	59	28	128	MAHH0	2.11
MAHH0\si1664	0	0	0	0	13	3	36	0	21	46	110	16	245	MAHH0	2.39
MAHH0\sk124	0	0	0	0	9	2	22	0	15	1	56	10	115	MAHH0	2.55
MAHH0\sk214	0	1	0	3	13	7	5	12	21	4	95	30	191	MAHH0	3.17
MBEFO\sa1	0	0	0	0	13	4	2	2	18	4	13	82	138	MBEFO	4.56
MBEFO\sa2	0	0	0	0	4	13	2	1	8	10	15	53	106	MBEFO	3.53
MBEFO\si1281	0	0	0	0	13	16	1	13	3	2	9	91	148	MBEFO	5.69
MBEFO\si1911	0	5	0	3	8	0	1	10	3	3	6	66	105	MBEFO	6.6
MBEFO\si1651	0	0	0	10	16	12	1	10	11	7	29	72	168	MBEFO	2.48
MBEFO\si111	0	0	0	0	1	6	0	1	0	8	7	25	48	MBEFO	3.12
MBEFO\sk201	0	0	0	0	7	2	0	7	8	3	2	54	83	MBEFO	6.75
MBEFO\sk21	0	0	0	0	0	1	0	0	4	5	8	32	50	MBEFO	4
MBEFO\sk291	0	0	0	0	0	9	2	3	3	9	7	66	99	MBEFO	7.33
MBEFO\sk381	0	0	0	0	1	0	1	1	7	0	10	59	79	MBEFO	5.9
Toplam 2 hata (sarı ile boyanan konuşmalar)															

tekrarlandı. Bu deneyin sonucu Tani.vi'dan elde edilen çıktı ile EK3 TABLO E3-3'te verilmiştir. TABLO 4.4'te sadece TABLO 4.2 ve TABLO 4.3'deki konuşmaların sonuçları sunulmuştur. Her konuşmacının konuşmalarda aldığı puanlar, Tani.vi'n verdiği karar ve G değerleri verilmiştir. TABLO 4.4'ten yeniden eğitimin fayda verdiği anlaşılmaktadır. Sadece mae0/sa1 ve mae0/si1956'da hata yapılmıştır. Böylece birinci eğitimden sonra hata oranının en iyi % 5 olmasına karşılık, ikinci eğitimden sonra hata oranı % 1.7'ye düştüğü görülmüştür.



Şekil 4.6 Tani.vi'nin subvi hiyerarşisi.

4.6 LabVIEW ile Tanıma

Şekil 4.4'te görülen Tani.vi'nin bağımlı olduğu subvi'lar Şekil 4.6'daki hiyerarşi diyagramında görülmüyor. Bu sanal cihazlardan bizim çalışmamızda gerçekleştirilenler Tani.vi, 3-DUzaklik.vi, InitSpeakers.vi ve 3-DArray.vi'dir. Beyaz ikonlarla gösterilen Open/Create/ReplaceFile.vi ve Read From Spreadsheet File.vi LabView kütüphanesine aittir.

InitSpeakers.vi sınıflama (tanıma) işleminden önce konuşmacıların SOM'larını dosyalardan belleğe okur. SOM'lar 12X8X20 elemanlı 3-boyutlu dizinlerdir. InitSpeakers.vi, bunları okurken 3-DArray.vi'ı kullanır. Tani.vi bilinmeyen konuşmanın bellekteki SOM'lara uzaklığını bulurken 3-DUzaklik.vi'dan faydalanır.

Şekil 4-7 Tani.vi diyagramını göstermektedir. Sanal cihazın iki çerçeveden oluşan bir sequence yapısı vardır (Şekil 4-7 (a) ve (b)). Birinci çerçevede InitSpeakers SOM haritalarını cod dosyalarından belleğe okur. Bellek bir cluster array olarak düzenlenmiş ve "C" diye adlandırılmıştır. Cluster Pascal dilindeki record, C dilindeki struct veri yapılarının LabVIEW'daki karşılığıdır. C şu şekilde tanımlanmıştır:

```
typedef struct {
    float      SOM[12][8][20]; /* SOM (kod defteri) */
    double int SKOR;           /* tanıma puanı */
    char       KONUŞAN[];      /* SOM etiketi */
} kisi;

kisi          C[12];
```

Görüldüğü gibi "C" oniki elemanlı bir cluster dizinidir. Her elemanı, bir konuşmacının kod defterini, tanıma puanını ve ismini (SOM etiketini) içerir. SOM okuma işlemi aşağıda `InitSpeakers.vi`'da kısaca ele alınacaktır. Birinci çerçeve ayrıca tanıma sonuçlarının kaydedileceği bir dosya açar. Bu dosyaya `Tanima.txt` adını verir.

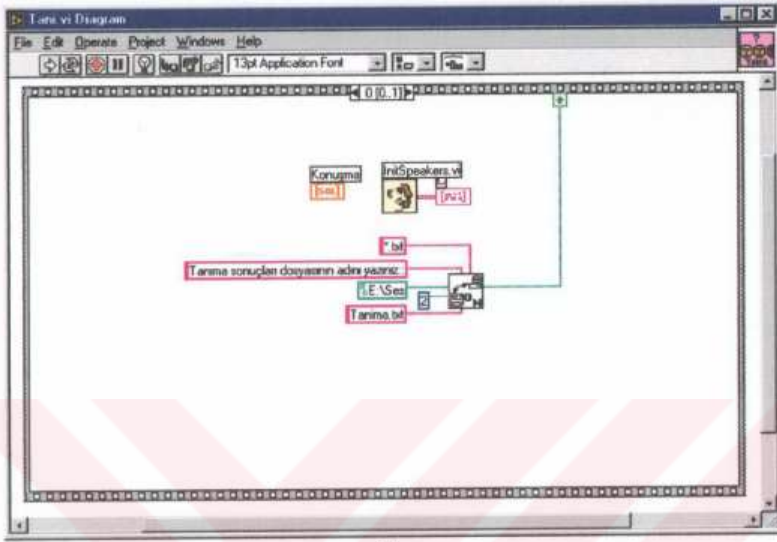
İkinci çerçevedeki sonsuz döngüde dört çerçeveli bir sequence yapısına daha girilir. Birinci çerçevede bilinmeyen konuşma `Konusma` adlı yerel değişkene kaydedilir. Hatırlanacağı üzere, bilinmeyen konuşmalar `2nciGecis` sanal cihazı ile elde edilen `dat` öznitelik dosyalarıdır (Bölüm 3.4.3 İkinci Geçiş). Bu sırada konuşmacıların puanları sıfırlanır.

Şekil 4.7(c)'de 1 nolu çerçeve* konuşma işaretinin her çerçevesi (kod vektörü) için Euklid uzaklığını kullanarak denklem (4.1) ile ifade edilen minimum uzaklığa sahip konuşmacı indisini bulur. Uzaklık bulmak için `3-Duzaklik.vi` kullanılır. Bu çerçeve içinde iki `for` döngüsü vardır. En içteki döngü kod vektörü için 12 elemanlı bir uzaklık dizini oluşturur. Döngünün dışında bu dizinin minimumu bulunarak o kod vektörünün (çerçevenin) hangi konuşmacıya atanacağı belirlenir ve bu konuşmacının puanı 1 arttırılır.

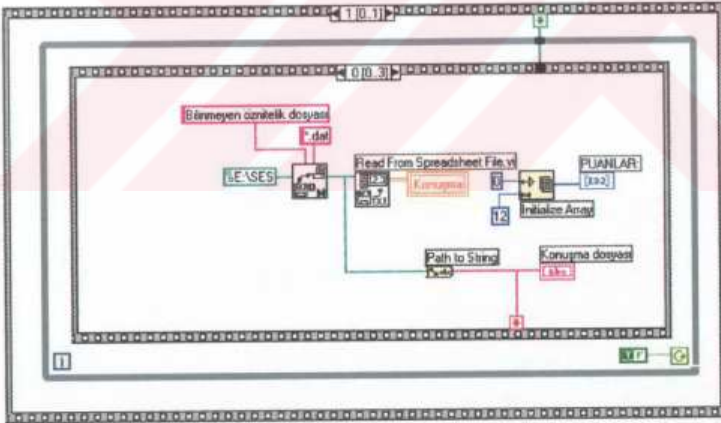
Şekil 4.7(d)'de 2 nolu çerçevede (4.3) ve (4.4) denklemleri gerçekleştirilir. Bunu için en fazla puan alan iki konuşmacı bulunur. Bu iki konuşmacı arasından puanı büyük olanı bilinmeyen konuşmayı yapan olarak panoya yazılır. Kazanan konuşmacının puanının diğer konuşmacının puanına oranı `G` güvenilirlik oranı olarak panoya ve dosyaya kaydedilir. Şekil 4.4, TABLO E3.3 ve TABLO 4.4 `Tani.vi` çıktılarıdır. LabVIEW kütüphanesinden alınan `ConcatenateStrings` fonksiyonu ve `WriteCharactersToFile.vi` sanal cihazı sonuçları normal olarak `Tanima.txt` dosyasına yazar.

Şekil 4.7(e)'de 3 nolu çerçeve kullanıcıya başka bir konuşma ile devam etmek isteyip istemediğini sorar. "HAYIR" denirse sonsuz döngüden çıkılır.

* Bu paragrafta çerçeve iki ayrı anlamda kullanılmaktadır. İlk kullanımda LabVIEW sequence yapısının çerçevesi, ikinci kullanımda ise ses işaretine ait çerçeve kastedilmektedir

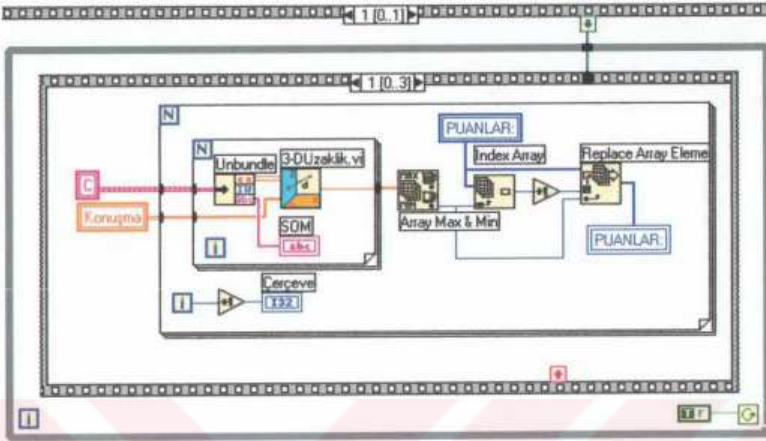


(a)

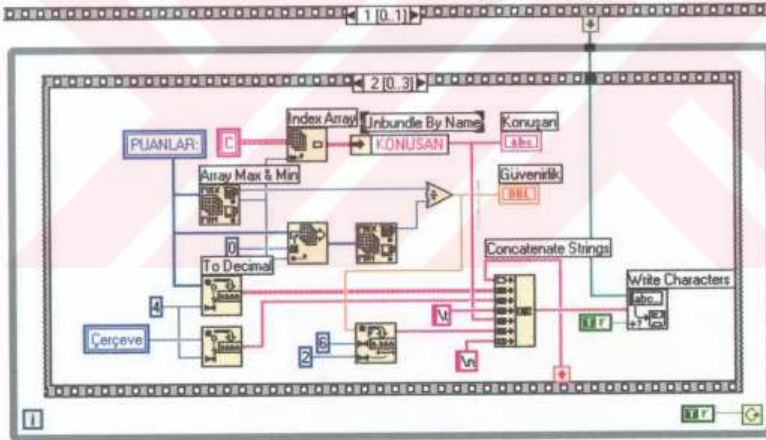


(b)

Şekil 4.7(a, b) Tani.vi diyagramı iki çerçeveli bir "sequence"den meydana gelir. İkinci çerçevedeki sonsuz while döngüsünün içinde dört çerçeveli bir sequence daha vardır. (a) 0 nolu çerçevede kod defterleri (SOM haritaları) belleğe okunur, sonuçların yazılacağı dosya için kullanıcıyla diyalog kurulur. (b) 1 nolu çerçevedeki sonsuz döngü içinde dört çerçeveli bir sequence daha bulunur. 0 nolu çerçevede kullanıcıya tanımak istediği konuşma sorulur ve konuşmacıların puanları sıfırlanır. Konuşma dosyası Konuşma yerel değişkeninde tutulur. Puanlar ve Konuşma dosyası ön panoya yazılır.

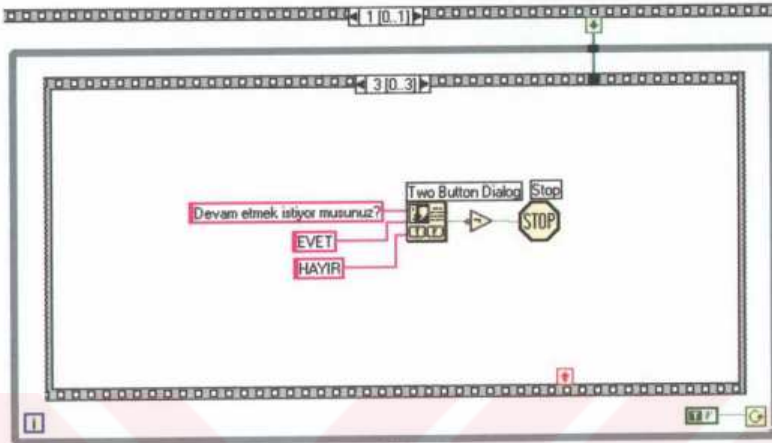


(c)



(d)

Şekil 4.7(c, d) Tani.vi diyagramı. (c) Bilinmeyen konuşmanın çerçeve çerçeve C cluster dizininin bütün elemanlarına olan uzaklığı hesaplanır. Bir çerçevede minimum uzaklığı tutturun SOM'un konuşmacısına bir puan verilir. Bu sırada panoda puanlar, SOM isimleri ve çerçeve numaraları sürekli yenilenir. (d) En fazla puan alan konuşmacı panoda "Konuşan" olarak ilan edilir. Bu arada maksimuma en yakın puan bulunur ve güvenirlilik oranı hesaplanarak panoya yazılır.



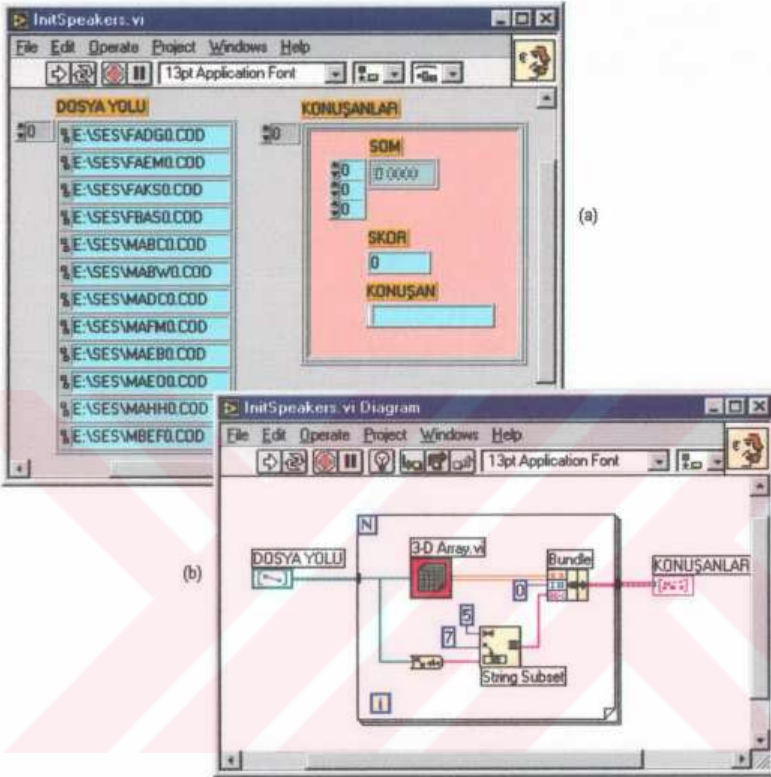
(e)

Şekil 4.7(e) Tani.vi diyagramı. Bilinmeyen konuşmanın kime ait olduğu tahmin edildikten sonra kullanıcıya başka bir konuşma ile devam etmek isteyip istemediği sorulur. Konuşmacı "HAYIR" diye cevap verirse sonsuz döngüden çıkar. "EVET" diye cevap verilirse sonsuz döngünün 0 nolu çerçevesine dönülür ve başka bir konuşma için aynı işlemler tekrarlanır.

Şekil 4.8'de InitSpeakers sanal cihazı görülmektedir. DOSYA YOLU panoda görüldüğü gibi konuşmacı SOM'larının yollarını içeren bir dizindir. For döngüsü içindeki 3-DArray sanal cihazı her konuşmacı SOM'unu 3-boyutlu bir dizin olarak okur, ve Bundle fonksiyonunu kullanarak sayfa 86'da tanımlanan "cluster" yapısını oluşturur. Bu yapının diğer iki elemanı, KONUŞAN ve SKOR, Bundle fonksiyonunun diğer iki girişi olarak şekilde görülmektedirler. For döngüsü çıkışında oluşan KONUŞANLAR dizini Tani sanal cihazındaki C dizinine bağlanır.

Şekil 4.9'da 3-DUzaklik 12X8X20 elemanlı 3-boyutlu bir dizini, her koordinatında 20 elemanlı bir vektör bulunan 96 elemanlı 2-boyutlu bir harita olarak alır. Konuşma işaretinin çerçevesi ile haritanın 96 vektörü arasındaki Euklid uzaklıklarının karelerinin minimumu bulunur. Bu uzaklık konuşma çerçevesinin haritaya uzaklığı olarak kabul edilir.

Şekil 4.10'da 3-DArray sanal cihazının diyagramı görülmektedir. SOM_PAK yazılımında cod (SOM) dosyaları metin dosyaları (ASCII text file) halinde aşağıdaki formata sahiptir:



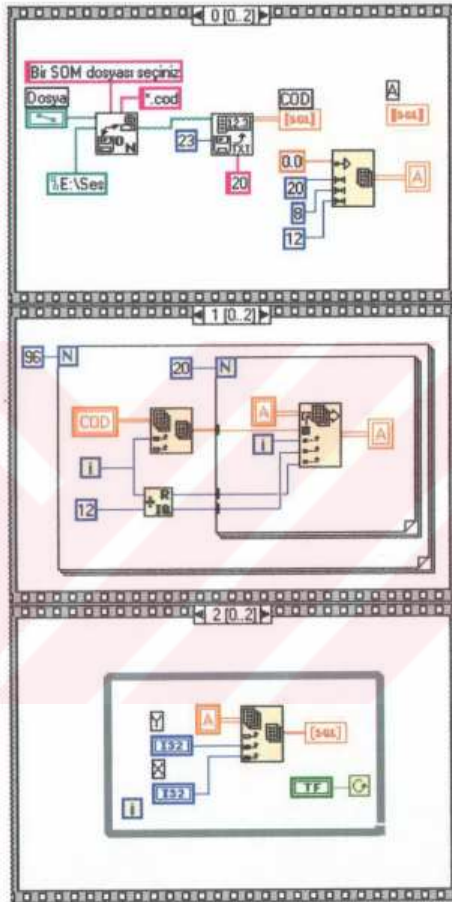
Şekil 4.8 InitSpeakers sanal cihazı. Konuşmacıların SOM'ları (kod defterleri) 3-DArray sanal cihazıyla okunup Konuşanlar cluster dizinine kaydedilir.

```

20 hexa 12 8 gaussian
1.016 -0.0255769 0.214393 0.343647...-0.00448917 (x=0, y=0)
1.01215 0.039127 0.204791 0.327775...-0.0108389 (x=1, y=0)
.....
1.0129 0.679992 -0.0563998 -0.295892...-0.0187654 (x=11, y=7)
(kT0 C1 C2 C3 ... C19)

```

Şekil 4.10'da birinci çerçevedeki ve Read From Spreadsheet File.vi, cod dosyasının başında bulunan "20 hexa 12 8 gaussian" 23 karakterlik kısmı atar. Vi'a



Şekil 4.10 3-D Array.vi diyagramı. Birinci çerçevede SOM_PAK formatında kaydedilmiş bir SOM dosyası (metin dosyası olarak) açılır ve 3-boyutlu bir dizi hazırlanır. İkinci çerçevede SOM haritasının kod vektörleri bu 3-boyutlu dizine kopyalanır. Üçüncü çerçeve sanal cihazın testi için hazırlanmıştır ve normal olarak kapalı tutulur (TF = False).

5 SONUÇLARIN TARTIŞILMASI ve ÖNERİLER

5.1 YSA Mimarisi

Biyolojik olarak öğrenmenin, kortekste sinir hücrelerinin oluşturduğu öznitelik haritalarında sinaptik ağırlıkların bir düzen ve tertibe girmeleri şeklinde olduğu gösterilmiştir (Willshaw ve von der Malsburg, 1976, Kohonen, 1982). SOM yapay sinir ağı süpervizörsüz (denetleyicisiz) bir ağıdır ve kortekste öğrenmeyi modeller. Tek katmanlı bu YSA şu özelliklere sahiptir:

1. Eğitimi kolaydır.
2. Eğitildiği verileri 2-boyutlu öznitelik haritalarına çevirerek özniteliklere bir görsellik kazandırır.

Metinden bağımsız konuşmacı tanıma, konuşan kişinin anatomisi ve konuşmacının zaman içinde bu anatomik özellikleri kullanma tarzından yararlanır. Konuşmacının öznitelikleri, bilgice "zengin" bir eğitim setinin olasılık yoğunluk fonksiyonlarına yansır ve öznitelik haritaları yaklaşık olarak veri setinin olasılık yoğunluk fonksiyonlarını temsil eder.

Bu çalışmamızda konuşmacı tanıma için SOM yapay sinir ağını kullandık. Diğer yapay sinir ağlarına göre oldukça basit olan bu mimari EK-2 ve 4. Bölümde gösterildiği gibi gerçekten başarılı tanıma sonuçları vermiştir. TIMIT her konuşmacı için kısıtlı bir konuşma tabanı sunmaktadır. Bu kısıtlamaya rağmen, metinden bağımsız tanıma gibi zor bir amacın gerçekleştirilebilmiş olması SOM yapay sinir ağının gücünü gösterir.

Yöntem olarak literatürde anlatılan uygulamanın dışına çıktık. SOM genellikle farklı sınıfların özniteliklerinin öne çıkarıldığı tek bir haritadır (Kangas, 1990 ve Haykin, 1994). Klasik uygulamada bütün sınıflar bir tek haritayı eğitirler. Aynı yaklaşımı kullansaydık oniki TIMIT konuşmacısı, özniteliklerindeki farklılıklara göre büyük bir SOM'un farklı bölgelerinde oniki öbek oluşturacaktı. Yukarıda değinilen kaynaklar bu haliyle SOM'un bilinmeyen veriyi sınıflandırmak için kullanılmasını uygun görmemektedirler. SOM'la

beraber süpervizörlü doğrusal sınıflayıcı veya LVQ önerilmektedir. Bu öneriler bütün sınıflar için bir SOM kullanılması durumundadır. Hatta Helsinki Teknoloji Üniversitesi SOM_PAK yazılım grubunun bu iş için hazırlamış olduğu bir LVQ_PAK yazılım paketi bile vardır.

Bir tek SOM kullanıldığında sınıflar arasında kesin sınırlar olmayacağını ve konuşmacı kimliğini belirleyen çok sayıda özneliğin sınıflar arasındaki belirsiz bölgede kaybolacağını düşündük. Bu yüzden oniki TIMIT konuşmacısı için bir tek SOM kullanmak yerine, her konuşmacı için ayrı bir SOM kullanmayı uygun gördük. Eğer bir tek SOM kullanılmış olsaydı bütün konuşmacıların konuşmaları aynı dosya içinde olacak ve hep beraber bir tek SOM'u eğiteceklerdi. Olasılık yoğunluk fonksiyonları birbirine benzeyen konuşmacılar birbirlerinin öbekleri üzerinde çekecekler ve öğrenmeyi olumsuz yönde etkileyebileceklerdi. Bu durumda ise belirleyici özneliklerdeki kontrastın flulaşması kaçınılmaz olacaktı. Her konuşmacıya bir SOM ayrıldığında ise, her konuşmacı için ayrı bir eğitim dosyası kullanılır. Bir dosyanın verileri ile bir konuşmacının SOM'u eğitilir. Bu şekilde eğitilen SOM'da ortaya çıkan öznelik öbekleri de o konuşmacının zaman içinde farklılık gösteren öznelikleri olur. Konuşmacının eğitim konuşmalarından öğrenilmiş böyle bir öznelik haritasının, olasılık yoğunluk fonksiyonuna daha yakın olacağı ve içinde statik ve dinamik özellikleri barındıracağı açıktır. Statik özneliklerin yanı sıra temporal özneliklerin korunması metinden bağımsız tanıma için son derece önemlidir.

Haritaların büyüklüğü 12X8 olarak seçilmiştir. Bu büyüklükteki bir SOM her konuşmacıyı 96 kod vektörü ile temsil eder. Bu seçim üzerinde herhangi bir irdelememiz olmadı. Gerek SOM_PAK gerekse NENET gibi SOM yazılımları örneklerinde bunu seçtikleri için biz de aynen bu büyüklüğü benimsedik. 96 kod vektörünün iyi bir tanıma için yeterli temsil gücü var mıdır? Biz bu soruyu cevaplandırmayı amaç edinmedik. SOM büyüklüğünün tanıma üzerindeki etkisi ilerideki çalışmalara konu olabilir. Aynı şekilde SOM eğitim parametreleri de ya SOM_PAK grubunun önerdiği parametreler veya bizim "hissettiğimiz" parametreler olmuştur.

5.2 SOM Performansı

TABLO E3-3'te $k = 0.03$ alınarak oniki TIMIT konuşmacısının 120 konuşmasıyla yapılan tanıma deneyleri görülmektedir. Oniki sütunun herbirinde konuşmacıların elde ettikleri

puanlar, konuşmaların toplam çerçeve sayıları, varılan teşhis ve G güvenilirlik oranı verilmiştir. Bu deneylerin sonucunda iki hata göze çarpmaktadır. Dolayısıyla doğru karar verebilme oranı %98.3'tür. $1 < G < 2$ aralığında 11 karar, $2 \leq G < 10$ aralığında 89 ve $10 \leq G < \infty$ aralığında 18 karar vardır. Tani.vi, hata yaptığı *maeo0\sa1* ve *maeo0\si1956* konuşmaları için güvenilirlik oranlarını 1.34 ve 1.16 olarak vermektedir. 1.34 den küçük güvenilirlikle teşhis edilen iki konuşma bulunmaktadır (*maeo0\si1326* için $G = 1.31$, *maeo0\sx246* için $G = 1.26$). Dolayısıyla KARAR = *maeo* ve $G \leq 1.34$ olduğu zaman kullanıcının verilen karar hakkında ihtiyatlı olması ihtar edilebilir.

İstatistik zenginleştirme öncesinde üç konuşmacıda (*maeb0*, *maeo0* ve *mahh0*) hata yapılmakta iken, zenginleştirmeden sonra hatalar sadece *maeo0*'a indirgenmiştir. Bu noktadan sonra hataları sıfıra indirmek veya G oranını iyileştirmek için yeniden istatistik zenginleştirmeye başvurmamak. Çünkü amacımız ortaya hatasız bir ticari sistem koymak değil, SOM'ların verilen ek bilgileri öğrenebildiğini göstermekti. Daha zengin bir konuşma eğitimi ve belki de 12X8'den daha büyük bir haritayla hatasız tanımaya ulaşmak mümkün olabilir.

Deneylerde dikkat çeken başka bir husus da, ilk anda sanılabileceği gibi, eğitimde kullanılan konuşma uzunluğuyla öğrenme arasında doğrudan bir ilişki bulunmamasıdır. TABLO 4-1 konuşmacıların trn.dat eğitim dosyalarının çerçeve olarak uzunluklarını göstermektedir. Bu tabloda en fazla çerçeve *mahh0*'da (1273 çerçeve) en az çerçeve de *mabc0* konuşmacısındadır (336 çerçeve). Fakat EK3'teki tablolara bakıldığında zaman *mahh0* konuşmalarındaki tanıma performansı *mabc0* konuşmalarındaki tanıma performansının çok gerisindedir. Bir kere istatistik zenginleştirme işleminden önce $k = 0.03$ için *mahh0*'a ait konuşmalarda 3 yanlış tanıma bulunmakta ve maksimum güvenilirlik oranı $G = 8.5$ olmaktadır. Buna karşılık *mabc0*'a ait konuşmalarda hiç hata bulunmamakta, maksimum güvenilirlik oranı $G = \infty$ olmaktadır. Bu konuşmacıya ait minimum güvenilirlik oranı ise 2.56'dır. İkinci olarak istatistik zenginleştirme işleminden sonra bile *mahh0*, *mabc0*'ın performansına yetişememektedir. En iyi güvenilirlik oranı sadece bir konuşma için 11 olmakta diğer dokuz konuşmada 10'un altında kalmaktadır. Bunu *mahh0*'ın eğitim setinin istatistik olarak yeterli bilgiyi içermemesiyle açıklayabiliriz. Gerçekten *mahh0*'ın konuşmaları dinlendiğinde *mahh0*'ın konuşma hızının diğer konuşmacıdan çok daha yavaş

olduğu görülmüştür. Dolayısıyla 1273 çerçevelik eğitim konuşması bu uzunlukla orantılı fazla bilgiyi içermediği anlaşılmaktadır.

Deney sonuçlarında hemen göze çarpmayan bir yan ürün de, geliştirdiğimiz bu tanıma sisteminin konuşmacıları *erkek* ve *kadın* olarak da sınıflayabilmesidir. Bunun tek istisnası *maeo0\si1956* konuşmasının *faem0*'a sınıflanmasıdır (*maeo0* erkek, *faem0* ise kadın konuşmacıdır). Yukarıda anlatılan ek tedbirlerle bunda da iyileşme sağlanabilir.

5.3 Glotal Uyarım Miktarı

Bizce bu çalışmanın en önemli bulgusu, en iyi tanıma için ses tellerinin öznitelik vektörüne katkı miktarının belirlenmesidir. Öznitelik vektöründe kepstrum katsayıları sabit tutulup temel peryod bir skala faktörü ile çarpılmaktadır. kT_0 ile ifade ettiğimiz bu öznitelik glotal uyarım miktarını belirler.

Glotal uyarımı (k faktörünü) değiştirerek yaptığımız deneylerden elde edilen performans değerleri TABLO E3-2 ve Şekil 4.5'te verilmiştir. Bunlardan $k = 0$ değeri glotal uyarımın bulunmadığı durumdur. Üçüncü bölümde ortaya konan bilgiler ışığında bu durumun fisiltı ile konuşma olduğunu görürüz. Performans değerlerine yakından bakıldığı takdirde konuşmacıyı tanımak açısından $k = 0$ hali, k 'nın uygun seçilmediği pekçok durumdan daha iyi performans sağlamaktadır. Diğer bir deyişle temel peryod T_0 'a büyük bir ağırlık vermektense fisiltıdan konuşmacıyı tanımak daha başarılı olmaktadır. Şekil 4.5'e göre 20 bileşenli öznitelik vektörü ile k 'nın en iyi değeri 0.028 ~ 0.03 değerleri arasındadır. Biz bu iki değer arasında tahmin edilen maksimum değeri aramadık. Şekil 4.5'ten ona çok yakın olduğu anlaşılan $k = 0.03$ değerini benimsedik.

Bizimkinden farklı öznitelik vektörü seçimlerinde de glotal uyarım miktarına özen gösterilmesi gerektiği inancını taşımaktayız.

5.4 Yarım Perde Seslendirmesi

32 ms 'lik çerçevelerimizde sadece kadın konuşmacılarda YPS gözlenebilmektedir. Erkeklerde temel peryod 10 ~ 12.5 ms olabilmektedir. 256 örnekli 32ms uzunlukta bir çerçevenin kepstrum domenindeki Nyquist kefransı 16 ms'dir. YPS'nin görülebilmesi için temel peryodun 8ms'den daha büyük olmaması veya temel frekansın 125 Hz'den daha

küçük olmaması gerekir. Erkeklerde de YPS olduğuna inanıyoruz. Ancak bu saydığımız sebeplerden dolayı erkeklerde YPS gözleyemedik. Durağanlık kaygılarını bir tarafa bırakarak çerçeve uzunluğumuzu 64 ms yapsaydık büyük bir olasılıkla erkeklerde de YPS görülecekti.

5.5 LabVIEW

En son olarak kısaca LabVIEW yazılımına değinmek gerektiğini düşünüyoruz. 1990'lı yıllarda kişisel bilgisayar işletim sistemlerinde DOS ortamından GUI (Graphic User Interface) tabanlı WINDOWS işletim sistemine büyük bir kayma oldu. Teknik arka planı bulunmayan vasat bir kullanıcıya sunduğu kullanma kolaylığı, DOS'taki 640KB bellek sınırına karşılık CPU (Central Processing Unit) biriminin görebildiği maksimum belleği kullanabilme olanağı ve multitasking özellikleri sebebiyle çok kısa bir zamanda bu işletim sistemi bilgisayarlara hakim oldu. Bilimsel uygulamaların vazgeçilmez işletim sistemi UNIX'in kullanıldığı makinaların fiyatının kişisel bütçelerin kaldıramayacağı kadar yüksek olması sebebiyle, WINDOWS hızla yayıldı.

WINDOWS'un sunduğu avantajların bedeli ise profesyonel olmayan programcılara yüklediği ağır külfettir. DOS sistem kaynakları üzerinde cimri değildi. DOS ve BIOS fonksiyonlarını kullanarak, 80x86 Assembler dili ile veya BASIC, PASCAL, C gibi dillerle mühendisler olarak hemen hemen her istediğimizi bilgisayara yaptırabiliyorduk. WINDOWS bütün bunları değiştirdi. Artık multitasking ve grafik tabanlı bir işletim sisteminin bütün ince ayrıntılarını göz önüne alarak program yazmak mümkün olabilmektedir. WINDOWS'un binin üzerinde API (Application Program Interface) fonksiyonları bulunmaktadır. Alt seviyede program yazarken API fonksiyonları ile program yazmak kaçınılmaz olmuştur. Bu ince ayrıntılarla uğraşmak bizim için çoğu kez hedeften sapmakla eş anlamlı hale gelmiştir. İşte LabVIEW bu sınıf uygulamacılarla WINDOWS işletim sistemi arasına giren yüksek seviyeli bir uygulama geliştirme platformudur. LabVIEW mühendislik ve bilim çevrelerine hitap eden veri toplama ve işlemeye yoğunlaşmış grafik olarak programlanan bir platformdur. Programcı uygulamasını yazmayarak fakat çizerek oluşturur. LabVIEW'nun kullandığı bu dil G dili (Grafik dili) olarak da anılmaktadır.

LabVIEW olmasaydı bu çalışma belki de olmayacaktı. Bu çalışmanın bir orijinalliği de G diliyle yapılmış olmasıdır. LabVIEW bazı az sayıda kısıtlamalarla, C, PASCAL gibi dillerle yapılabilecek hemen her türlü bilimsel uygulamaya yatkındır. Fakat tersine LabVIEW ile beş on dakikada gerçekleştirilen uygulamaları bu klasik dillerle gerçekleştirmek belki de günler alabilir. Eğer uygulamanın bir de gerçek zamanda çalışması isteniyorsa mutlaka profesyonel WINDOWS programcılarının yardımı gerekli olacaktır.

Çalışmamızın başlarında SOM_PAK yazılımını DLL (Dynamic Link Library) haline dönüştürme veya kendimize özgü bir SOM DLL yazma veya alternatif olarak SOM_PAK'ın fonksiyonlarını üstlenecek bir alt sanal cihaz hazırlama düşüncemiz vardı. LabVIEW, "Call Library Function" fonksiyonu ile kullanıcıya herhangi bir dil ile yazdığı DLL fonksiyonunu çağırma imkanı sağlar. SOM_PAK yazılımının çok sayıda modülden oluşması ve yüzlerce satırdan oluşan *source code*'u bu yazılımın WINDOWS platformuna tercüme edilmesini güçleştiren bir etkidir. Onun için bu düşüncemizi gerçekleştiremedik ve SOM_PAK'ı olduğu gibi kullandık.

5.6 İleriye Dönük Öneriler

Gerçek zamanda gerçekleştirmedığımız konuşmacı tanıma oniki TIMIT konuşmacısı için her çerçeveyi 0.4 saniyede sınıflamaktadır. LabVIEW "data-flow execution" esasına göre çalışır, fakat WINDOWS işletim sisteminin kısıtlamaları ile karşı karşıyadır. WINDOWS 95 ve NT her uygulamaya sırayla 20 ms kadar bir çalışma zamanı verir. Bu yüzden PC ortamında tanıma uzun zaman almaktadır. Yapay sinir ağlarının temel fikri olan paralel işlem kullanılırsa sınıflama zamanları kabul edilebilir değerlere inecektir. İleriki çalışmalarda SOM ile sınıflama için çoklu DSP (Digital Signal Processor) ortamı kullanmayı öneririz. Bugünkü teknoloji bize 30 MFLOP (Million Floating-Point Operations Per second) ve yukarısında çalışabilen Floating-Point DSP çipleri sunmaktadır. İki DSP çipi kullanan bir mimari kullanarak bir DSP çipi ile gelen konuşma işaretininin kepstum analizini yapabilir ve özneliliklerini çıkarabiliriz. Diğer DSP çipi ile de SOM haritalarına bakıp çerçeveyi sınıflayabiliriz.

Konuşmacı sayısını yükseltmek, gürültülü ortamda konuşulan sesi sınıflamak, SOM haritasının büyüklüğünü optimize etmek de ileride yapılabilecek çalışmalardır.

Son olarak, YPS olgusunu arařtırmak isteriz. Bebek ađlamaları üzerine yapılan yeni alıřmalar larink sisteminin nonlinearer dinamikle sıkı bađlantıları olduđunu ortaya koydu (Mende vd., 1990). Kpek larinkslerinde titreřen ses telleri ile yapılan alıřmalarda larinksin asimetrik bir yapı ve nonlinearite gsterdiđi bulundu. Buna bađlı olarak altharmonik ieren, modulasyonlu ve kaosumsu ses retimine rastlandı (Berry vd., 1996). Fackrell (1996) da lineer ses retim modeline ve kaynak-filtre bađımsızlıđına itiraz etti. Akustik borular modelinde ve bunun elektriksel eřdeđerinde, larinkse bađlanan ilk borudan geriye yansıyan akustik basın dalgasının ses tellerinin osilasyon modu zerinde etkili olduđunu dřnyoruz. Zaman serileri analizi ile bu yansımayı bulup YPS ile paralellik gsterip gstermeyeceđini arařtırmak istiyoruz.



6 KAYNAKLAR

- Atal, B.S. ve Hanauer, S.L., (1971), "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", *Journal of the Acoustic Society of America* 50: 637-655.
- Atal, B.S., (1976), "Automatic Recognition of Speakers from Their Voices", *Proceedings of the IEEE* 64(4): 460-475.
- Baum, L. ve Petrie, T., (1966), "Statistical Inference for Probabilistic Functions of Finite State Markov Chains", *Annals of Mathematical Statistics*, 37:1554-1563.
- Berry, W., Herzel, H., Titze, I.R., ve Story, B. (1996), "Bifurcations in Excised Larynx Experiments", *Journal of Voice* 10: 129-138.
- Bogert, B.P., Healey, M.J.R. ve Tukey, J.W., (1963), "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-aoutocovariance, Cross-cepstrum and SapheCracking", *Proc. Symposium Time Series Analysis*, M. Rosenblatt, Ed., John Wiley and Sons, New York: 209-243.
- Clifford, J.W., 1991, "Opportunities for the Advanced Speech Processing in Military Computer-Based Systems", *Proceedings of the IEEE* 79(11): 1627-1639.
- Colombi, J.M., (1996), Generalized Hidden Filter Markov Models Applied to Speaker Recognition, Doktora tezi, Faculty of School of Engineering, Air Force Institute of Technology, Air University, <http://www.au.af.mil/> Defense Technical Information Center (DTC), Scientific and Technical Information Network (STINET).
- Durbin, R., Miall, C., ve Mitchison, G., editörler (1989), *The Computing Neuron*, Addison-Wesley, Reading, Massachusetts.
- Elomaa, J., Halme, J., Hassinen, P., Hodju, P. ve Rönkkö, J., (1998), NENET 1.0b, Nenet-Team@hut.fi
- Fackrell, J.W.A., (1996), Bispectral Analysis of Speech Signals, Doktora tezi, Edinburgh Universitesi.
- Faggin, F., (1991), "VLSI Implementation of Neural Networks," tutorial notes. International Joint Conference on Neural Networks, Seattle, Washington.
- Fakotakis, N., Tsopanoglou, A. ve Kookinakis, G., (1993), "A Text-Independent Speaker Recognition System Based on Vowel Spotting", *Speech Communications* 12: 57-68.
- Fischler, M.A ve Firschein, O., (1987), *Intelligence: The Eye, the Brain and the Computer*, Addison-Wesley, Reading, Massachussts.
- Flanagan, J.L., (1972), *Speech Analysis, Synthesis and Perception*, Dowden, Ssprinter-Verlag, New York.
- Flanagan, J.L. ve Rabiner, L. R. editörler, (1973), *Speech Synthesis*, Dowden, Hutchinson and Ross, Inc., Stroudsburg, Pennsylvania.
- Fukushima, K., (1975), "Cognitron: A Self-Organizing Multilayered Neural Network", *Biological Cybernetics* 20: 121-136.

- Furui, S., (1994), "An Overview of Speaker Recognition Technology", Proceedings of the 1994 ECSCA Workshop on Speaker Recognition, Identification and Verification.
- Furui, S., (1996), "Speaker Recognition", Survey of the State of the Art in Human Language Technology, (Editorial Board: Ronald A. Cole, Joseph Mariani, Hans Uszkoreit, Annie Zaenen, Victor Zue, Giovanni Battista Varile, Antonio Zampolli), National Science Foundation European Commission, <http://www.cse.ogi.edu/CSLU/HLTSurvey/ch1node1.html>.
- Gregory, S., 1986, "Voice Spectrography Evidence: Approaches to Admissibility", University of Richmond Law Review 20: 357-376.
- Haykin, S., (1994), Neural Networks, A Comprehensive Foundation, Prentice-Hall, Upper Saddle River, New Jersey.
- Hebb, D.O., (1949), The Organization of Behavior: A Neuropsychological Theory, Wiley, New York.
- Hertz, J., Krogh, A. ve Palmer, R.G., (1991), Introduction to the Theory of Neural Computation, Addison-Wesley, Reading, Massachusetts.
- Hubel, D.H. ve Wiesel, T.N., (1962), "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex", Journal of Physiology (London) 160: 106-154.
- Hubel, D.H. ve Wiesel, T.N., (1977), "Functional Architecture of Macaque Visual Cortex", Proceedings of the Royal Society of London, Series B 198: 1-59.
- Kaas, J.H., Merzenich, M.M. ve Killackey, H.P., (1983), "The Reorganization of the Somatosensory Cortex Following Peripheral Nerve Damage in Adult and Developing Mammals", Annual Review of Neurosciences 6: 325-356.
- Kangas, J., Kohonen, T., ve Laaksonen, J., (1990), "Variants of Self-Organizing Maps", IEEE Transactions on Neural Networks 1: 93-99.
- Kohonen, T., (1982), "Self-Organized Formation of Topologically Correct Feature Maps", Biological Cybernetics 43: 59-69.
- Kohonen, T., (1988a), "An Introduction to Neural Computing", Neural Networks 1: 3-16.
- Kohonen, T., (1988b), "The 'Neural' Phonetic Typewriter", Computer 21: 11-22.
- Kohonen, T., (1988c), Self-Organization and Associative Memory, 3.ncü baskı, Springer-Verlag, New York.
- Kohonen, T., (1990), "The Self-Organizing Map", Proceedings of the IEEE 78: 1464-1480.
- Kohonen, T., (1993), "Physiological Interpretation of the Self-Organizing Map Algorithm", Neural Networks 6: 895-905.
- Kohonen, T., Hynninen, J., Kangas, J. ve Laaksonen, J., (1995), SOM_PAK, The Self-Organizing Map Program Package version 3.1, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland (<ftp://cochlea.hut.fi>).
- Leinonen, L., Kangas, J., Torkkola, K. ve Juvas, A., (1991), Artificial Neural Networks, (ed. T. Kohonen, T. Mäkisara, O. Simula, J. Kangas), North-Holland, Amsterdam, Hollanda.

- Levinson, S.E., Rabiner, L.R. ve Sondhi, M.M., (1983), "An Introduction to the Applications of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition", *The Bell System Technical Journal* 62(4): 1035-1074.
- Liu, L., He, J. ve Palm, G., (1997), "A Comparison of Human and Machine in Speaker Recognition", *Eurospeech 97 Vol.5: 2327-2330*, 5th European Conference on Speech Communication and Technology, 22-25 Eylül 1997, Rodos.
- Luttrell, S.P., (1989), "Self Organization: A Derivation from First Principle of a Class of Learning Algorithms", *IEEE Conference on Neural Networks: 495-498*, Washington DC.
- Luttrell, S.P., (1991), "Code Vector Density in Topographic Mappings: Scalar Case", *IEEE Transactions on Neural Networks* 2: 427-436.
- Markel, J.D. ve Gray, A.H., Jr., (1976), *Linear Prediction of Speech*, Springer-Verlag, New York.
- Matsui, T. ve Furui, S., (1992), "Comparison of Text-Independent Speaker Recognition Methods Using VQ Distortion and Discrete/Continuous HMMs", *Proceedings of the 1992 ICASSP*, vol.2: 157-160.
- McCulloch, W.S. ve Pitts, W., (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics* 5: 115-133.
- Mende, W., Herzel, H. ve Wermke, K., (1990), "Bifurcations and Chaos in Newborn Infant Cries", *Physical Letters A* 145: 418-424.
- Minsky, M.L. ve Papert, S.A., (1969), *Perceptrons*, Cambridge, Massachusetts: MIT Press.
- Nolan, F., (1983), *The Phonetic Bases of Speaker Recognition*, Cambridge University Press, New York.
- Oglesby, J. ve Mason, J.S., (1991), "Radial Basis Function Networks for Speaker Recognition", *Proceedings of the 1991 ICASSP: 393-396*.
- Poritz, A. (1982), "Linear Prediction of Hidden Markov Models", *Proceedings of the 1982 ICASSP: 1291-1294*.
- Proakis, J.G., Deller, J.R. ve Hansen, J.H.L., (1993), *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co.: New York.
- Rabiner, L.R. ve Schafer, R.W., (1978), *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Rabiner, L.R., (1989), "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE* 77(2).
- Rabiner, L.R. ve Juang, B.H., (1993), *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Ramón y Cajál, S., (1911), *Histologie du Système Nerveux de l'Homme et des Vertébrés*, Paris: Maloin; Edition Français Revue: Tome I , 1952; Tome II, 1955; Madrid: Consejo Superior de Investigaciones Científicas.
- Reynolds, D.A., (1992), *Gaussian Mixture Modeling Approach to Speaker Identification*, Doktora tezi, Georgia Institute of Technology.

Reynolds, D.A. ve Rose, R.C., (1995), "Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models", *IEEE Transactions on Speech and Audio Processing* 3(1): 72-83.

Rivarol, V. ve O'Shaugnessy, D., (1997), "A Double Gaussian Mixture Modeling Approach to Speaker Recognition", *Eurospeech 97 Vol.5: 2287-2290*, 5th European Conference on Speech Communication and Technology, 22-25 Eylül 1997, Rodos.

Rosenberg, A.E., (1997), "Speaker Identification with User-Selected Password Phrases", *Eurospeech 97 Vol.3: 1371-1374*, 5th European Conference on Speech Communication and Technology, 22-25 Eylül 1997, Rodos.

Rosenblatt, F., (1958), "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", *Psychological Review* 65: 386-408.

Rosenblatt, F., (1960), "On the Convergence of Reinforcement Procedures in Simple Perceptrons", Report, VG-1196-G-4. Cornell Aeronautical Laboratory, Buffalo, New York.

Sharma, M., ve Mammone, R.J., (1996), "Subword-based Text-dependent Speaker Verification System with User-selectable passwords", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP): 93-96*.

Sheperd, G.M. ve Koch, C., (1990), "Introduction to Synaptic Circuits.", *Synaptic Organization of the Brain*, (G.M. Sheperd, ed): 3-31. New York: Oxford University Press.

Soong, F.K., Rosenberg, A.E., Rabiner, L.R. ve Juang, B.H., (1985), "A Vector Quantization Approach to Speaker Recognition", *Proceedings of the 1982 ICASSP*, vol. 1: 387-390.

Sönmez, M.K., Heck, L., Weintraub, M. ve Shriberg, E., (1997), "A Log-normal Tied Mixture Model of Pitch for Prosody Based Speaker Recognition", *Eurospeech 97 Vol.3: 1391-1394*, 5th European Conference on Speech Communication and Technology, 22-25 Eylül 1997, Rodos.

Sprint FONCARD®, Texas Instruments, <http://www.ti.com/research/docs/speechv>.

Suga, N., (1985), "The Extent to which Binaural Information is Represented in the Bat Auditory Cortex", *Dynamic Aspects of Neocortical Function: 653-695*, (Ed: Edelman, G.M., Gall, W.E. ve Cowan, W.M.) Wiley (Interscience), New York.

Texas Instruments, (1998), Logic Selection Guide and Data Book CD-ROM.

VoicePass Calling Card, (1998), VoiceID Systems, Inc., <http://www.voicepass.com/techinfo.htm>, San Diego, California.

von der Malsburg, C., (1973), "Self-Organization of Orientation Sensitive Cells in the Striate Cortex", *Kybernetik* 14: 85-100.

Willshaw, D.J. ve von der Malsburg, C., (1976), "How Patterned Neural Connections Can Be Set Up by Self-Organization", *Proceedings of the Royal Society of London, Series B* 194: 4331-445.

Wolf, J., (1972), "Efficient Acoustic Parameters for Speaker Recognition", *Journal of the Acoustic Society of America* 51(6): 2044-2056.

Yu, G. ve Gish, H., (1993), "Identification of Speakers Engaged in Dialog", *Proceedings of the 1993 ICASSP*, vol.2: 383-386.

Zeek, E.J., (1996), **Speaker Recognition by Hidden Markov Models and Neural Networks**, Master tezi, Faculty of School of Engineering, Air Force Institute of Technology, Air University, <http://www.au.af.mil/> Defense Technical Information Center (DTC), Scientific and Technical Information Network (STINET).

Zhu, X., (1994), "Text-Independent Speker Recognition Using VQ, MixtureGaussian VQ and Ergodic HMMs", Proceedings of the 1994 ECSA Workshop on Speaker Recognition, Identification and Verification.



EK1 KONUŞMACILAR

Çalışmamızda konuşmaları kullanılan TIMIT konuşmacıları ile ilgili bilgiler aşağıda TABLO E1-1'de verilmiştir. Konuşmacıların isimlerinin baş harflerinden önce kadın konuşmacılar için "F", erkek konuşmacılar için "M" harfi eklenmiştir. İsimlerin baş harflerinden sonra 0-9 arasında bir rakam kullanarak aynı isme sahip konuşmacılar birbirinden ayırt edilmiştir. Konuşmacıların hepsi de ABD'de konuşulan sekiz şive (dialect) grubundan birine mensuptur. Bu sekiz şive grubu şunlardır: 1 - New England, 2 - Kuzey, 3 - Orta Kuzey, 4 - Orta Güney, 5 - Güney, 6 - New York City, 7 - Batı ve 8 - Gezgin.

Genel olarak TIMIT konuşmacılarının ait oldukları ırklar A - Asyalı, B - Beyaz, K - Kızılderili, S - İspanyol asıllı Amerikalı, Z - Zenci olup bizim kullandığımız grupta 11 beyaz ve 1 zenci bulunmaktadır. Konuşmacıların eğitim durumları BS (lisans), MS (yüksek lisans) ve PHD (doktora) ile gösterilmiştir.

TABLO E1-1KONUŞMACILAR

Etiket	Cinsiyet	Bölge	Doğum tarihi	İrk	Eğitim	Notlar
FADG0	K	4	11/04/1959	B	MS	
FAEM0	K	2	13/05/1960	B	BS	
FAKS0	K	1	01/04/1986	B	PHD	
FBAS0	K	4	04/09/1933	B	BS	
MABCO	E	6	17/06/1960	B	BS	
MABW0	E	2	26/05/1959	B	BS	
MADC0	E	3	27/04/1954	B	HS	
MAEB0	E	4	20/09/1937	B	HS	
MAEO0	E	7	19/07/1955	B	BS	
MAFM0	E	7	22/12/1952	B	BS	
MAHH0	E	5	22/02/1953	Z	BS	
MBEF0	E	3	12/02/1986	B	PHD	Hafif işitme kaybı var

TIMIT veri tabanında her konuşmacı on cümle okumaktadır. Bu cümleler *sa*, *sx* ve *si* cümleleri olarak adlandırılır. Her konuşmacı *sa1* ve *sa2* konuşmasını okur. *si* cümleleri sadece bir defa okunmuştur. Bir *sx* cümlesi bir kaç konuşmacı tarafından okunabilirse de bizim 12 kişilik grubumuzda tekrarlanan *sx* cümlesi olmamıştır.

Her konuşmacının okuduğu cümleler aşağıda gösterilmiştir. Cümle isimleri ve cümleler *italik* harflerle yazılmıştır.

Standard cümleler:

sa1 she had your dark suit in greasy wash water all year

sa2 don't ask me to carry an oily rag like that

fadg0:

- sx199 young children should avoid exposure to contagious diseases*
- sx379 pam gives driving lessons on thursdays*
- sx109 she slipped and sprained her ankle on the steep slope*
- sx289 weatherproof galoshes are very useful in seattle*
- sx19 aluminum silverware are often flimsy*
- si649 it suffers from a lack of unity of purpose and respect for heroic leadership*
- si1279 bricks are an alternative*
- si1909 fat showed in loose rolls beneath the shirt*

faem0:

- sx312 those answers will be straightforward if you think them through carefully first*
- sx42 biblical scholars argue history*
- sx222 you always come up with pathological examples*
- sx402 we'll serve rhubarb pie after rachel's talk*
- sx132 publicity and notoriety often go hand in hand*
- si762 fill small hole in bowl with clay*
- si1392 assume for example a situation where a farm has a packing shed and fields*
- si2022 what outfit does she drive for*

faks0:

- sx43 elderly people are often excluded*
- sx223 put the butcher block table in the garage*
- sx403 her wardrobe consists of only skirts and blouses*
- sx133 pizzerias are convenient for a quick lunch*
- sx313 drop five forms in the box before you go out*
- si943 production may fall far below expectations*
- si1573 his captain was thin and haggard and his beautiful boots were worn and shabby*
- si2203 the reasons for this dive seemed foolish now*

fbas0:

- sx307 the meeting is now adjourned*
- sx37 critical equipment needs proper maintenance*
- sx217 how permanent are their records*
- sx397 tim takes sheila to see movies twice a week*
- sx127 the emperor had a mean temper*
- si1472 she greeted her husband's colleagues with smiling politeness offering nothing*
- si1387 several factors contributed to this change*
- si2066 he saw a pint sized man with a graying spade beard and an unusually large head*

mabc0:

- sx331 the big dog loved to chew on the old rag doll*
- sx61 chocolate and roses never fail as a romantic gift*
- sx241 the hallway opens into a huge chamber*
- sx421 those who teach values first abolish cheating*
- sx151 the previous speaker presented ambiguous results*
- si781 again these blocks were set in resin saturated glass cloth and nailed*
- si1620 his head flopped back*

si2041 the shot reverberated in diminishing whiplashes of sound

mabw0:

sx134 december and january are nice months to spend in miami
sx314 f people were more generous there would be no need for welfare
sx44 basketball can be an entertaining sport
sx224 how good is your endurance
sx404 barb burned paper and leaves in a big bonfire
si1230 princes and factions clashed in the open street and died on the open scaffold
si1664 there were other farmhouses nearby
si2294 just long enough to make you feel important

madc0:

sx287 in the long run it pays to buy quality clothing
sx17 carl lives in a lively home
sx197 draw each graph on a new axis
sx377 she always jokes about too much garlic in his food
sx107 birthday parties have cupcakes and ice cream
si737 have a test run on the family first to be sure timing and seasoning are right
si1367 the battery median grade equivalent was used in data analysis in this study
si1997 like as if it were built of books

maeb0:

sx90 the thinker is a famous sculpture
sx270 are holiday aprons available to us
sx450 roy ignored the spurious data points in drawing the graph
sx180 almost all colleges are now coeducational
sx360 i assume moisture will damage this ship's hull
si990 and these hardy travelers are not unappreciated today
si1411 resolved that the anti slavery sentiment is becoming ripe for resolute action
si2250 all your wishful thinking won't change that

maeo0:

sx246 how ancient is this subway escalator
sx426 a moth zig zagged along the path through otto's garden
sx156 our experiment's positive outcome was unexpected
sx336 rob made hungarian goulash for dinner and gooseberry pie for dessert
sx66 we got drenched from the uninterrupted rain
si1655 so he understood her panic
si1326 the larvae kept warm by the queen are full grown in about ten days
si1956 a single kick made it spring open shuddering

mafm0:

sx39 barb's gold bracelet was a graduation present
sx219 nonprofit organizations have frequent fund raisers
sx399 they assume no burglar will ever enter here
sx129 angora cats are furrier than siamese
sx309 the proof that you are seeking is not available in books
si939 impressions often appear in a symbolic form and cannot be taken at face value
si1569 this is our singular goal
si2199 will you tell me why

mahh0:

sx214 *the morning dew on the spider web glistened in the sun*
 sx394 *calcium makes bones and teeth strong*
 sx124 *objects made of pewter are beautiful*
 sx304 *cheap stockings run the first time they're worn*
 sx34 *don't do charlie's dirty dishes*
 si664 *look at the things we do to escape our fears and to forget our worries*
 si1294 *f you are a life tenant you deduct allowable depreciation and depletion*
 si1924 *getting the boss rattlesnake to help you*

mbef0:

sx201 *approach your interview with statuesque composure*
 sx381 *we welcome many new students each year*
 sx111 *his sudden departure shocked the cast*
 sx291 *the avalanche triggered a minor earthquake*
 sx21 *alfalfa is healthy for you*
 si651 *far more frequently overeating is a result of a psychological compulsion*
 si1281 *the same shelter could be built into an embankment or below ground level*
 si1911 *ten years ago they blew up some of our ditches*



EK2 SOM_PAK YAZILIMININ KULLANIMI

Self-Organizing Map Program Package, SOM_PAK Versiyon 3.1, Teuvo Kohonen ve SOM programlama ekibi tarafından Helsinki Teknoloji Üniversitesi Bilgisayar ve Enformasyon Bilimleri Bölümü'nden Internet aracılığıyla temin edilebilmektedir. Paket UNIX ve MS-DOS platformlarında çalışabilmektedir. Biz MS-DOS versiyonu kullandık Bu versiyon som_pak3r1.exe adında kendi kendine açılan sıkıştırılmış bir ZIP arşiv dosyasıdır.

som_pak3r1.exe çalıştırıldığı zaman mevcut direktori altına Som_pak.3r1 adında bir direktori açılır ve içindeki dosyalar bu direktoriye kaydedilir. Bu dosyalar şunlardır:

Dosya adı	Büyükölük	Tarih	Saat	Açıklama
Readme	4,763	07/04/1995	03:01	
Version	22	07/04/1995	03:01	
Ex.dat	197	07/04/1995	03:01	Örnek data dosyası
Ex_fdy.dat	4,894	07/04/1995	03:01	Örnek data dosyası
Ex_fts.dat	13,780	07/04/1995	03:01	Örnek data dosyası
Ex_ndy.dat	5,028	07/04/1995	03:01	Örnek data dosyası
Makefile.dos	2,839	07/04/1995	03:01	Make dosyası
Header.ps	2,914	07/04/1995	03:01	PostScript Dosyası
Makefile.uni	3,064	07/04/1995	03:01	UNI Dosyası
Datafile.c	29,617	07/04/1995	03:01	
Fileio.c	10,514	07/04/1995	03:01	
Header.c	4,064	07/04/1995	03:01	
Labels.c	10,899	07/04/1995	03:01	
Lvq_pak.c	15,720	07/04/1995	03:01	
Map.c	34,692	07/04/1995	03:01	
Mapinit.c	6,344	07/04/1995	03:01	
Median.c	1,985	07/04/1995	03:01	
Planes.c	14,426	07/04/1995	03:01	
Qerror.c	4,361	07/04/1995	03:01	
Sammon.c	13,516	07/04/1995	03:01	
Som_rout.c	22,542	07/04/1995	03:01	
Umat.c	16,782	07/04/1995	03:01	
Vcal.c	7,280	07/04/1995	03:01	
Version.c	219	07/04/1995	03:01	
Vfind.c	9,876	07/04/1995	03:01	
Visual.c	7,103	07/04/1995	03:01	
Vsom.c	6,590	07/04/1995	03:01	
Config.h	3,903	07/04/1995	03:01	
Datafile.h	4,399	07/04/1995	03:01	
Errors.h	2,673	07/04/1995	03:01	
Fileio.h	2,972	07/04/1995	03:01	
Labels.h	2,726	07/04/1995	03:01	
Lvq_pak.h	9,555	07/04/1995	03:01	
Som_rout.h	2,494	07/04/1995	03:01	
Umat.h	3,855	07/04/1995	03:01	
Version.h	71	07/04/1995	03:01	

Bu dosyaların büyük çoğunluğu C dili ile yazılmış "source" dosyalarıdır. Ayrıca örnek veri dosyaları, açıklama, proje dosyaları vb bulunur. Uygun bir C derleyicisi kullanarak Kohonen ağı (SOM haritasını) eğitecek bir program elde edilebilir. Biz bu maksatla Borland C++ versiyon 3.0'ı kullandık.

SOM_PAK kullanıcıya bir çok program sunmaktadır. Fakat bize Kohonen ağını eğitmek için iki program gereklidir: a) SOM haritasını oluşturacak ve düğümlere rastgele sayılar atayacak olan randinit.exe, b) oluşturulan SOM haritasını eğitecek vsom.exe. Diğer programlar bize gerekli olmayıp, bunlarla ilgili fazla bilgi için SOM_Pak1.txt dosyasına başvurulabilir.

randinit.exe ve vsom.exe programlarını oluşturmak için Borland C++ içinde iki ayrı proje oluşturulabilir. Ancak daha önce "header" dosyalarının (uzantısı h olan dosyaların) ya Borland C++ "include" dosyalarının bulunduğu direktoriye kaydedilmesi veya bunların bulunduğu direktörinin Borland C++ Integrated Development Environment içinde tanıtılması gereklidir. OPTIONS/DIRECTORIES seçeneğine girerek bu dosyaların yerleri belirtilmelidir. Örnek olarak Borland C++, C:\BORLANDC direktorisinde bulunuyorsa

C:\BORLANDC\INCLUDE

denebilir. Daha sonra OPTIONS menüsünde bazı seçenekleri işaretlemek gerekmektedir:

```
OPTIONS /APPLICATIONS: "Dos Standard"
OPTIONS /COMPILER/CODE GENERATION: "Compact"
OPTIONS /COMPILER/ADVANCED CODE GENERATION: Floating point
emulation: 80287
OPTIONS /COMPILER/ADVANCED CODE GENERATION: Instruction Set: 80286
OPTIONS /COMPILER/OPTIMIZATIONS: Optimize for speed
```

Bundan sonra iki proje dosyası hazırlanmalıdır. Örnek olarak birincisine rand.prj ikincisine ise vsom.prj adı verilebilir. Bu projelere dahil edilen "source" dosyaları aşağıdaki gibi olmalıdır:

<u>rand.prj:</u>	<u>vsom.prj</u>
mapinit.c	vsom.c
datafile.c	datafile.c
lvq_pak.c	lvq_pak.c
som_rout.c	som_rout.c
file_io.c	ile_io.c
labels.c	labels.c
version.c	version.c

Bundan sonra menüden COMPILE/Build all opsiyonu seçildiğinde bu programlar derlenir, "link" edilerek rand.prj'den mapinit.exe, vsom.prj'den vsom.exe elde edilir. Mapinit.exe dosyasının adı randinit.exe olarak değiştirilmelidir.

Tipik bir SOM eğitimi, randinit.exe programının rastgele bir SOM oluşturmasıyla başlar. Oluşturulan SOM (xdim x ydim) boyutlarındadır. Her koordinat (Kohonen ağı düğümü) öznitelik vektörü boyutundadır. İkinci ve üçüncü aşamalar bu ağı eğitilmesidir. vsom.exe eğitim verilerini iki defa kullanarak SOM haritasını düzenler. Birinci geçişte ağı düğümleri kabaca sıraya girer. Kazanan nöron etrafındaki etki bölgesinin yarıçapı neredeyse tüm haritayı kaplayacak kadar geniş ve öğrenme katsayısı da büyük tutulur.

İterasyon sayısı birkaç bin mertebelerindedir. İkinci geçişte daha fazla iterasyon kullanılır, etki bölgesinin yarıçapı bir kaç nöronu kapsayacak kadar daraltılır ve öğrenme katsayısı için birinci geçişten daha küçük değerler kullanılır. Böylece harita düzenine "ince ayar" yapılmış olur.

Örnek:

```
randinit -xdim 12 -ydim 8 -din e:\ses\fadg0\trn.dat -cout
fadg0.cod -topol hexa -neigh gaussian -rand 123

vsom -buffer 250 -din e:\ses\fadg0\trn.dat -cin fadg0.cod -cout
fadg0.cod -rlen 2000 -alpha 0.05 -radius 10

vsom -buffer 250 -din e:\ses\fadg0\trn.dat -cin fadg0.cod -cout
fadg0.cod -rlen 15000 -alpha 0.03 -radius 3
```

Birinci DOS kumandası 12 X 8 büyüklüğünde fadg0.cod adıyla altıgen, gauss komşuluklu bir ağ (SOM haritası) oluşturur. Ağın düğümleri 123 tohum değeri kullanılarak rastgele belirlenir.

İkinci kumanda e:\ses\fadg0\trn.dat eğitim dosyasını kullanarak 2000 adımda bu ağ kabaca düzenler. Öğrenme katsayısı 0.05, etki bölgesinin yarıçapı 10'dur. Konvansiyonel belleğin tüketilmesini önlemek amacıyla eğitim dosyasından her seferinde 250 satır okunur.

Üçüncü kumanda yine aynı eğitim dosyasını kullanarak fadg0.cod ağını eğitmeye devam eder. Fakat bu sefer etki yarıçapı 3 nöron, öğrenme katsayısı 0.03'tür.

Eğitim dosyasının ilk satırı eğitim (öznitelik) vektörlerinin boyutunu bildirmelidir. Aynı örneğe dönersek e:\ses\fadg0\trn.dat dosyasının ilk iki satırı şöyledir:

```
20
0.840 -0.156 0.198 0.315 -0.038 -0.184 -0.080 -0.099 -
0.060 -0.203 -0.090 -0.127 -0.163 -0.041 -0.065 -0.092
0.012 -0.054 -0.033 0.026
```

Eğitimde kullandığımız SOM_PAK Program parametreleri:

SOM_PAK yazılımı beraberinde pekçok yardımcı programı içerir. Bu programlar MS-DOS kumanda satırından verilen kumandalarla, kumandalar da program parametreleriyle yönlendirilir. SOM_PAK bu amaçla pekçok parametre tarif etmiştir. Ancak bizim kullandığımız randinit ve vsom programları ile ilgili parametreler aşağıya alınmıştır. Yukarıdaki örneklerde bu parametrelerin kullanılış biçimleri verilmiştir.

din	Giriş dosyasının adı.
dout	Çıkış dosyasının adı.
cin	Referans vektörlerinin okunduğu dosyanın adı.
cout	Referans vektörlerinin kaydedildiği dosyanın adı.
alpha	Başlangıçtaki öğrenme hızı. Eğitim boyunca doğrusal olarak sıfıra iner.
radius	SOM algoritmasının başlangıçtaki eğitim bölgesi yarıçapı. Eğitim boyunca doğrusal olarak bire iner.
xdim	x-yönündeki nöron sayısı.

ydim	y-yönündeki nöron sayısı.
topol	Haritada kullanılan topoloji tipi. Altıgen haritalar için hexa, dikdörtgen haritalar için rect olabilir.
neigh	Kullanılan komşuluk fonksiyonu tipi. Kullanılabilecek değerler basamak fonksiyonu (bubble) ve Gaussian (gauss)'dur.
rand	Rastgele sayı jeneratörü için yeni bir tohum değeri tanımlanıp tanımlanmadığını belirler.
buffer	Veri giriş dosyasından bir seferde kaç satır okunacağını belirler.



EK3 TANIMA SONUÇLARI

Bu ekte üç tablo sunulmuştur. Skala faktörü ile güvenilirlik oranının değişimi TABLO E3-1'de, bu tablodan hesaplanan performans değerlerinin skala faktörü ile değişimi TABLO E3-2'de, en iyi skala faktörü için 120 konuşmadan konuşmacıların aldıkları puanlar ve hesaplanan performans değerleri TABLO E3-3'te gösterilmiştir.

Tablolarda sarı renkli hücreler hatalı teşhisleri gösterir. Bu hücrelerdeki güvenilirlik oranları geçersizdir. Sarı hücrelerde $G < 1$ 'dir. Konuşmanın gerçek sahibi puan sıralamasında ikinci sırada ise hatalı G değerinden doğru olan değer $1/G$ ifadesiyle bulunabilir. Ancak bazı kereler konuşmanın gerçek sahibinin puan sıralamasında üçüncü sıraya düştüğü olmuştur. Bu durumda $1/G$ ifadesiyle gerçek G değerini bulmak imkansızdır. Bunu yapmak için Tani.vi sanal cihazında değişiklik yapma zorunluluğu vardır.



TABLO E3-1 SKALA FAKTÖRÜ İLE GÜVENİRLİK ORANININ DEĞİŞİMİ

KONUŞMALAR	SKALA FAKTÖRÜ, <i>k</i>									
	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80	
FADG0\Sa1	1.02	1.32	2.02	2.19	1.94	2.23	1.62	2.06	2.12	
FADG0\Sa2	1.41	4.28	4.52	2.8	3.32	2.39	1.95	2.34	1.7	
FADG0\Si1279	2.26	3.56	2.19	8.44	11.29	8.89	3.94	3.3	3.14	
FADG0\Si1909	1.16	2.03	1.73	3.37	3.7	3.56	2	1.47	1.29	
FADG0\Si649	1.47	2.15	2.33	2.15	2.13	1.89		1.68	1.58	
FADG0\Sx109	2.79	5	3.06	5.33	4.38	3.08	1.88	1.74	1.84	
FADG0\Sx19	3.03	2.52	2.67	3.25	5.4	3.57	2.24	2.08	1.62	
FADG0\Sx199	2.51	3.67	6.21	4.74	4.35	2.9	3.11	2.79	2.26	
FADG0\Sx289	3.81	7.11	5.32	8.18	6.33	5.04	2.8	2.79	3.15	
FADG0\Sx379	2.43	9.23	7.22	9.69	6.25	3.97	3.96	2.67	2.83	
FAEM0\Sa1	3	3.61	6.53	1.91	3.36	1.73	2.1	1.44	1.51	
FAEM0\Sa2	3.95	4.38	3.33	3.68	5.11	3.79	3.33	2.36	2.52	
FAEM0\Si1392	2.61	6.09	5.81	2.31	4.56	2.02	1.83	1.36	1.48	
FAEM0\Si2022	3.37	2.42	2.12	2.55	3.33	2.36	2.07	1	1.37	
FAEM0\Si762	1.18	1.28	2.22	2.17	2.06		1.5	1.63	1.58	
FAEM0\Sx132	4.9	10.08	11.91	5.9	10	7.76	5.13	3	3.48	
FAEM0\Sx222	8.1	6.54	11.25	11.25	12.12	9.27	8.73	6.29	4	
FAEM0\Sx312	4.72	7.46	9.36	10.78	11.89	7.53	5.56	5	4.27	
FAEM0\Sx402	4.19	6.43	11	3.91	7.67	4.65	8.75	4.05	4.19	
FAEM0\Sx42	5.5	13.6	25.33	7.11	14.2	9.86	7.33	5.08	4.75	
FAKS0\Sa1	5.04	9.35	5.03	5.64	4.44	2.34	2.16	1.76	1.86	
FAKS0\Sa2	2.14	2.66	2.62	2.35	3.28	1.25	1.48	1.18	1.5	
FAKS0\Si1573	2.27	2.9	2.05	3.83	4.34	3.74	3.81	4.4	4.27	
FAKS0\Si2203	5.42	8.13	4.54	3.48	3.39	2.47	2.91	2.05	2.32	
FAKS0\Si943	3.52	3.52	3.06	2.26	2.39	1.77	1.98	2.5	2.49	
FAKS0\Sx133	9.5	9	8.33	4.65	5.7	3.06	3.44	3.52	2.92	
FAKS0\Sx223	3.9	5.94	7.33	7.2	5.2	7.19	8.29	6.76	3.1	
FAKS0\Sx313	7.07	10.08	8.71	8.5	7.06	3.87	4.41	3.33	3.11	
FAKS0\Sx403	9.58	18	11.83	11.75	11.83	7.1	6.52	5.2	3.94	
FAKS0\Sx43	6.54	12.62	17.33	8	7.46	7.29	8.58	5.28	4.04	
FBAS0\Sa1	2.45	5	5.32	3	4.3	2.54	2.76	3.6	3	
FBAS0\Sa2	9	9.73	12	3.8	5.14	2.79	3.56	3.85	4	
FBAS0\Si1387	6.2	5.25	8.56	2.87	5.91	4.92	2.88	2.38	3.06	
FBAS0\Si1472	1.43	1.66	2.02	2.36	2.81	1.76	1.21	1.54	1.54	
FBAS0\Si2066	3.47	4.08	11.55	4.83	10.42	3.45	4.3	3.87	3.43	
FBAS0\Sx127	25	40.5	85	26.33	42	19.75	12.33	7.22	6.3	
FBAS0\Sx217	8.86	11.17	26.67	6.11	15.2	9.43	9.86	5.73	3.62	
FBAS0\Sx307	15	22.67	83	34	41	12.5	12	8.37	6.3	
FBAS0\Sx37	3.81	4.42	5.15	2.71	3.5	2.39	2.43	2.32	2.55	
FBAS0\Sx397	2.7	10.2	18.5	4.94	8.58	4.05	4.5	4.56	2.57	
MABC0\Sa1	1.53	3	5.17	6.2	3.33	5.4	6	4	2.44	
MABC0\Sa2	1.9	2.5	2.36	2.89	3.5	6.71	7.67	3.92	2.33	
MABC0\Si1620	3.17	15	∞	∞	∞	∞	∞	13	28	
MABC0\Si2041	2	2	2.56	3.25	2.5	3.83	4.8	4.36	5.56	
MABC0\Si781	1.35	2.56	3.47	4.82	4.42	4.65	4.87	4.28	5.33	
MABC0\Sx151	16	48	10.5	6.5	6.33	10.17	7.12	6.5	5.1	
MABC0\Sx241	3.6	8.2	15	23.5	11.25	7.43	8.5	8.17	5.75	
MABC0\Sx331	7.33	6.6	18	11.67	18	27.5	27.5	17.67	6.57	
MABC0\Sx421	4.44	19	18.5	20	13.33	11	9.33	6.71	8.17	
MABC0\Sx61	9.4	11	17.5	18	11.67	18.33	13.75	7.57	8.83	
MABW0\Sa1	6.29	4.11	1.5	1.71	1.28	2	2.14	2.44	2.81	
MABW0\Sa2	3.35	2.09	1.13	1.56	1.5	1.37	1.35	1.76	1.86	
MABW0\Si1230	5.31	4.4	1.96	2.41	2.08	2.81	1.72	1.24	1.38	
MABW0\Si1664	1.77	3.05	1.91	1.72	1.55	2.04	1.85	1.57	1.4	
MABW0\Si2294	2.94	2.12	2.06	1.33	1	2.32	2.4	1.87	1.65	
MABW0\Sx134	18.12	12.44	5.22	4.5	4.83	3.59	2.77	1.92	1.89	
MABW0\Sx224	7.14	2.92	5.29	4.33	5.12	5.56	4.6	4.3	3.36	
MABW0\Sx314	7.27	8.17	7.83	10	13.71	6.79	7.77	3.71	3.33	
MABW0\Sx404	6.13	7.73	10.37	5.07	5.54	7.33	6.43	4.53	4.07	
MABW0\Sx44	9.12	9.67	4.58	3.19	3.33	3.11	1.64	1.23	1.44	

TABLO E3-1 SKALA FAKTÖRÜ İLE GÜVENİRLİK ORANININ DEĞİŞİMİ (devam)

KONUŞMALAR	SKALA FAKTÖRÜ, <i>k</i>								
	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80
MADC0\Sa1	1.87	2.46	2.29	1.05	1.02	1.1	1.22	1.1	1.17
MADC0\Sa2	7.67	7.64	6.81	2.83	2.7	3.28	3.61	3.28	2.29
MADC0\Si1367	7.22	5.26	4.96	3.54	3.97	4.03	4.65	3.79	2.97
MADC0\Si1997	3.53	4.58	2.83	3.86	3.93	2.17	2.74	3.06	2.14
MADC0\Si737	4.03	7.37	5.96	2.07	1.92	3.6	3.23	2.74	2.74
MADC0\Sx107	8.37	8.37	6	4.67	3.71	3.93	2.88	2.55	1.91
MADC0\Sx17	11.2	10.7	10.6	15.43	11.33	8.83	6.93	5.06	4.47
MADC0\Sx197	6	6.33	4.93	4.79	4.33	4.41	4.24	3.05	2.83
MADC0\Sx287	15	17.43	9.67	10.09	9.5	8.64	7.93	6.62	6.12
MADC0\Sx377	8.58	7.92	4	8.45	9.2	9.09	5.06	4.28	3.7
MAEO0\Sa1	1.13	1.27	1.88	1	1.08	1.07	1.14	1.56	1.59
MAEO0\Sa2	1.49	2	3.4	1.18	1.75	1.63	1.55	1.05	1.5
MAEO0\Si1326	1.33	1.34	1.54	1.16	1.07	1.3	1.33	1.15	1.33
MAEO0\Si1655	2.11	1.23	1.13	1.33	1.31	1.25	1.27	1.92	1.75
MAEO0\Si1956	1.04	2	1.24					1.14	
MAEO0\Sx156	5.54	5.46	5.17	4.91	4.25	3.27	3.29	2.92	2.5
MAEO0\Sx246	4.22	2	1.53	1.56	1.75	1	1.35	1.62	2
MAEO0\Sx336	2.28	1.98	2.62	3.15	2.85	2.81	2.86	2.57	2.7
MAEO0\Sx426	3.22	2.88	3.15	2.26	2.53	1.77	1.74	2.08	2.03
MAEO0\Sx66	6.89	3.79	6	3.06	2.44	2.05	2.45	2.04	2.33
MAFMO\Sa1	2.88	3.23	4.4	8.14	5.4	3.06	2.88	3.14	2.9
MAFMO\Sa2	5.93	5.81	16.33	5.67	8.27	5.79	6.92	5.13	3.35
MAFMO\Si1569	1.25	3.18	1.95	1.67	1.4	1.14	1.11	1.25	1.03
MAFMO\Si2199	1.71	2.45	3.7	6.83	7.17	2.47	2.17	3.67	3.33
MAFMO\Si939	4.61	4.24	5.65	4.46	4.2	2.67	4.81	4.03	4.28
MAFMO\Sx129	10.3	13.11	25	9.91	10.9	8.08	5.71	5.93	3.82
MAFMO\Sx219	6.57	7.17	7.73	7.91	7.42	9.8	6.21	4.71	5.53
MAFMO\Sx309	10.6	13	9.55	12.37	6.36	6.79	6.71	5.11	4.47
MAFMO\Sx39	6.18	4.8	4	2.95	3.21	2.71	2.21	2.18	2.15
MAFMO\Sx399	6.15	8.8	7.08	8.8	6.38	4.11	5	4.17	4.47
MAHMO\Sa1	1.67	2	3.32	4.54	4.33	5.09	5.32	3.06	2.26
MAHMO\Sa2	1.31	1.28	1.53	1.69	1.5	1.1	1.15	1.14	1.14
MAHMO\Si1294	1.14	1.25	2.31	2.34	2.32	2.08	1.87	1.38	1.44
MAHMO\Si664	1	2	2.67		1.03	1.21	1	1.17	1.28
MAHMO\Si1924	1.82	1.22	1.09				1.15	1	1.13
MAHMO\Sx124	2.95	3.88	2.85	2.75	2.5	1.55	2.43	3.53	3.81
MAHMO\Sx214	2.41	5.35	4.81	4.95	4.26	3.41	2.89	2.23	2.69
MAHMO\Sx304	5.42	7.71	7.31	17.86	11.18	13.1	10.08	4.78	5.45
MAHMO\Sx34	8.1	9.33	8.5	6.31	4.76	5.43	6.82	7.09	5.69
MAHMO\Sx394	5.62	5.37	7.75	10.33	6.54	4.56	6.54	3.95	2.41
MBEFO\Sa1	9.09	9.78	3.25	1.67	1.83	1.8	1.26	1.16	1.05
MBEFO\Sa2	2.17	2.5	3.13	3.92	2.94	6.8	5	3.71	2.7
MBEFO\Si1281	3.36	2.58	2.24	2.25	1.87	1.24	1.28	1.06	1.03
MBEFO\Si1911	2.32	3.06	1.04	1.17	1.03	1	1.1	1.27	1.35
MBEFO\Si651	5.53	4.09	3.75	1.53	1.55	1.08		1.3	1.19
MBEFO\Sx111	4.57	5	7	10.33	7	3.75	2.56	2.2	1.58
MBEFO\Sx201	9.86	15	5.11	7.17	5.57	5.5	3.08	6.71	5.25
MBEFO\Sx21	10.5	5.67	5.83	3.12	3.25	4.71	3.67	4	1.92
MBEFO\Sx291	4.21	10	8.57	8.29	10.33	10.43	7.33	4.12	3.56
MBEFO\Sx381	5.6	6	3.53	5.2	6.87	7.12	5.3	5.3	4.5
Hatalar	7	6	6	9	9	9	9	8	9

TABLO E3-2 SKALA FAKTÖRÜ İLE PERFORMANSIN DEĞİŞİMİ

KONUŞMALAR	SKALA FAKTÖRÜ, <i>k</i>								
	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80
FADG0\Sa1	-1.00			0.50		1.00			0.25
FADG0\Sa2		0.50	1.00		0.25				
FADG0\Si1279				0.25	1.00	0.50			
FADG0\Si1909				0.25	1.00	0.50			
FADG0\Si649		0.50	1.00	0.50	0.25				
FADG0\Sx109		0.50		1.00	0.25				
FADG0\Sx19				0.25	1.00	0.50			
FADG0\Sx199			1.00	0.50	0.25				
FADG0\Sx289		0.50		1.00	0.25				
FADG0\Sx379		0.50	0.25	1.00					
FAEM0\Sa1		0.50	1.00		0.25				
FAEM0\Sa2	0.25	0.50			1.00				
FAEM0\Si1392		1.00	0.50		0.25				
FAEM0\Si2022	1.00			0.25	0.50				
FAEM0\Si762			1.00	0.50	0.25	-1.00			
FAEM0\Sx132		0.50	1.00		0.25				
FAEM0\Sx222			0.50	0.50	1.00	0.25			
FAEM0\Sx312			0.25	0.50	1.00				
FAEM0\Sx402			1.00		0.25		0.50		
FAEM0\Sx42		0.25	1.00		0.50				
FAKS0\Sa1	0.25	1.00		0.50					
FAKS0\Sa2		0.50	0.25		1.00				
FAKS0\Si1573					0.50			1.00	0.25
FAKS0\Si2203	0.50	1.00	0.25						
FAKS0\Si943	1.00	1.00	0.50					0.25	
FAKS0\Sx133	1.00	0.50	0.25						
FAKS0\Sx223			0.50	0.25			1.00		
FAKS0\Sx313		1.00	0.50	0.25					
FAKS0\Sx403		1.00	0.50	0.25	0.50				
FAKS0\Sx43		0.50	1.00				0.25		
FBAS0\Sa1		0.50	1.00		0.25				
FBAS0\Sa2	0.25	0.50	1.00						
FBAS0\Si1387	0.50		1.00		0.25				
FBAS0\Si1472			0.25	0.50	1.00				
FBAS0\Si2066			1.00	0.25	0.50				
FBAS0\Sx127		0.25	1.00		0.50				
FBAS0\Sx217		0.25	1.00		0.50				
FBAS0\Sx307			1.00	0.25	0.50				
FBAS0\Sx37	0.25	0.50	1.00						
FBAS0\Sx397		0.50	1.00		0.25				
MABC0\Sa1				1.00		0.25	0.50		
MABC0\Sa2						0.50	1.00	0.25	
MABC0\Si1620		0.25	1.00	1.00	1.00	1.00	1.00		0.50
MABC0\Si2041							0.50	0.25	1.00
MABC0\Si781				0.25			0.50		1.00
MABC0\Sx151	0.50	1.00	0.25						
MABC0\Sx241			0.50	1.00	0.25				
MABC0\Sx331			0.50		0.50	1.00	1.00	0.25	
MABC0\Sx421		0.50	0.25	1.00					
MABC0\Sx61			0.25	0.50		1.00			
MABW0\Sa1	1.00	0.50							0.25
MABW0\Sa2	1.00	0.50							0.25
MABW0\Si1230	1.00	0.50				0.25			
MABW0\Si1664		1.00	0.25			0.50			
MABW0\Si2294	1.00					0.25	0.50		
MABW0\Sx134	1.00	0.50							
MABW0\Sx224	1.00		0.25			0.50			
MABW0\Sx314		0.25		0.50	1.00				
MABW0\Sx404		0.50	1.00			0.25			
MABW0\Sx44	0.50	1.00	0.25						

TABLO E3-2 SKALA FAKTÖRÜ İLE PERFORMANSIN DEĞİŞİMİ (devam)

KONUŞMALAR	SKALA FAKTÖRÜ, k								
	0.00	0.01	0.03	0.05	0.06	0.10	0.20	0.50	0.80
MADC0\Sa1	0.25	1.00	0.50		-1.00				
MADC0\Sa2	1.00	0.50	0.25						
MADC0\Si1367	1.00	0.50	0.25						
MADC0\Si1997		1.00		0.25	0.50				
MADC0\Si737	0.25	1.00	0.50						
MADC0\Sx107	1.00	1.00	0.50	0.25					
MADC0\Sx17	0.25			1.00	0.50				
MADC0\Sx197	0.50	1.00	0.25						
MADC0\Sx287	0.50	1.00		0.25					
MADC0\Sx377	0.25				1.00	0.50			
MAEO0\Sa1	-1.00		1.00	-1.00			-1.00	0.25	0.50
MAEO0\Sa2		0.50	1.00		0.25				
MAEO0\Si1326	0.25	0.50	1.00	-1.00	-1.00	-1.00	-1.00		0.25
MAEO0\Si1655	-1.00	-1.00	-1.00	-1.00	-1.00		0.25	1.00	0.50
MAEO0\Si1956	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
MAEO0\Sx156	1.00	0.50	0.25						
MAEO0\Sx246	1.00	0.50			0.25	-1.00	-1.00	-1.00	-1.00
MAEO0\Sx336				1.00	0.25		0.50		
MAEO0\Sx426	1.00	0.25	0.50						
MAEO0\Sx66	1.00	0.25	0.50						
MAFM0\Sa1			0.25	1.00	0.50				
MAFM0\Sa2			1.00		0.50		0.25		
MAFM0\Si1569		1.00	0.50	0.25					-1.00
MAFM0\Si2199			0.25	0.50	1.00				
MAFM0\Si939	0.25		1.00				0.50		
MAFM0\Sx129		0.50	1.00		0.25				
MAFM0\Sx219			0.25	0.50		1.00			
MAFM0\Sx309	0.25	1.00		0.50					
MAFM0\Sx39	1.00	0.50	0.25						
MAFM0\Sx399		1.00	0.50	1.00	0.25				
MAHH0\Sa1				0.25		0.50	1.00		
MAHH0\Sa2	-1.00	-1.00	-1.00	-1.00	-1.00	0.50	1.00	-1.00	-1.00
MAHH0\Si1294			0.25	1.00	0.50				
MAHH0\Si664		-1.00	-1.00	-1.00	-1.00	-1.00		-1.00	-1.00
MAHH0\Si1924	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
MAHH0\Sx124		1.00						0.25	0.50
MAHH0\Sx214		1.00	0.25	0.50					
MAHH0\Sx304				1.00	0.25	0.50			
MAHH0\Sx34	0.25	1.00	0.50						
MAHH0\Sx394			0.50	1.00	0.25		0.25		
MBEF0\Sa1	0.50	1.00	0.25						
MBEF0\Sa2				0.25		1.00	0.50		
MBEF0\Si1281	1.00	0.50		0.25		-1.00	-1.00		
MBEF0\Si1911	0.50	1.00	0.25	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
MBEF0\Si651	1.00	0.50	0.25						
MBEF0\Sx111		0.25	0.50	1.00	0.50				
MBEF0\Sx201	0.50	1.00		0.25					
MBEF0\Sx21	1.00	0.25	0.50						
MBEF0\Sx291		0.25			0.50	1.00			
MBEF0\Sx381		0.25			0.50	1.00			
PERFORMANS, $p(k)$	21.50	40.75	41.75	18.50	19.00	6.25	2.25	-4.00	-3.75

TABLO E3-3 YENİDEN EĞİTİMDEN SONRA TANIMA PUANLARI

KONUŞMALAR	FADG0	FAEM0	FAKS0	FBAS0	MABC0	MABW0	MADC0	MAFM0	MAEB0	MAEO0	MAHO0	MBEFO	TOPLAM	KARAR	G
FADG0\Sa1	100	42	41	6	0	0	0	0	0	0	0	0	189	FADG0	2.38
FADG0\Sa2	120	17	31	16	0	0	0	0	0	0	0	0	184	FADG0	3.87
FADG0\Si1279	72	1	20	0	0	0	0	0	0	0	0	0	93	FADG0	3.6
FADG0\Si1909	69	22	38	2	0	0	0	0	0	0	0	0	131	FADG0	1.82
FADG0\Si1649	138	26	61	1	0	0	0	0	0	0	0	0	226	FADG0	2.26
FADG0\Sx109	111	6	36	0	0	0	0	0	0	0	0	0	153	FADG0	3.08
FADG0\Sx19	124	2	42	1	0	0	0	0	0	0	0	0	169	FADG0	2.95
FADG0\Sx199	131	18	31	2	0	0	0	0	0	0	0	0	182	FADG0	4.23
FADG0\Sx289	145	13	24	1	0	0	0	0	0	0	0	0	183	FADG0	6.04
FADG0\Sx379	139	2	17	0	0	0	0	0	0	0	0	0	158	FADG0	8.18
FAEM0\Sa1	7	110	6	27	0	4	0	9	0	0	0	1	164	FAEM0	4.07
FAEM0\Sa2	33	80	8	13	0	2	0	3	0	8	0	3	150	FAEM0	2.42
FAEM0\Si1392	26	112	22	47	0	4	6	13	0	3	0	0	233	FAEM0	2.38
FAEM0\Si2022	14	33	3	12	0	0	0	2	0	0	0	0	64	FAEM0	2.36
FAEM0\Si762	12	75	44	23	0	0	0	3	0	2	0	0	159	FAEM0	1.7
FAEM0\Sx132	12	130	8	21	0	1	0	4	0	3	0	3	182	FAEM0	6.19
FAEM0\Sx222	2	96	10	14	0	0	0	1	1	2	0	1	127	FAEM0	6.86
FAEM0\Sx312	15	112	11	11	0	4	1	4	0	1	0	1	160	FAEM0	7.47
FAEM0\Sx402	4	98	11	20	0	0	0	2	0	2	0	0	137	FAEM0	4.9
FAEM0\Sx42	5	74	2	7	0	0	0	0	0	0	0	0	88	FAEM0	10.57
FAKS0\Sa1	33	23	157	1	0	0	0	2	0	0	0	0	216	FAKS0	4.76
FAKS0\Sa2	51	20	104	2	0	0	0	0	0	0	0	0	177	FAKS0	2.04
FAKS0\Si1573	72	31	172	0	0	0	0	0	0	0	0	0	275	FAKS0	2.39
FAKS0\Si2203	24	16	115	1	0	0	0	6	0	0	0	0	162	FAKS0	4.79
FAKS0\Si943	37	3	114	0	0	0	0	0	0	0	0	0	154	FAKS0	3.08
FAKS0\Sx133	21	3	127	0	0	0	0	0	0	0	0	0	151	FAKS0	6.05
FAKS0\Sx223	19	8	117	0	0	0	0	0	0	0	0	0	144	FAKS0	6.16
FAKS0\Sx313	16	10	129	0	0	0	0	0	0	0	0	0	155	FAKS0	8.06
FAKS0\Sx403	9	8	151	1	0	0	0	0	0	0	0	1	170	FAKS0	16.78
FAKS0\Sx43	7	9	113	1	0	0	0	0	0	0	0	0	130	FAKS0	12.56
FBAS0\Sa1	8	30	1	96	0	0	1	21	0	0	0	0	157	FBAS0	3.2
FBAS0\Sa2	10	12	4	107	0	5	1	17	0	1	0	0	157	FBAS0	6.29
FBAS0\Si1387	9	12	1	69	0	8	0	0	0	0	0	0	99	FBAS0	5.75
FBAS0\Si1472	21	52	17	82	0	6	0	12	0	0	0	0	190	FBAS0	1.58
FBAS0\Si2066	12	43	10	209	0	23	0	18	0	0	0	3	318	FBAS0	4.86
FBAS0\Sx127	1	5	0	75	0	0	0	7	0	0	0	0	88	FBAS0	10.71
FBAS0\Sx217	0	5	2	73	0	4	0	2	0	0	0	1	87	FBAS0	14.6
FBAS0\Sx307	0	2	0	81	0	0	0	1	0	0	0	0	84	FBAS0	40.5
FBAS0\Sx37	15	16	2	58	0	0	1	2	0	2	0	0	96	FBAS0	3.62
FBAS0\Sx397	11	11	0	100	0	3	0	4	0	2	0	0	131	FBAS0	9.09
MABC0\Sa1	0	0	0	0	52	0	4	0	1	10	4	8	79	MABC0	5.2
MABC0\Sa2	0	1	0	0	47	5	2	0	0	6	4	1	66	MABC0	7.83
MABC0\Si1620	0	0	0	0	29	0	0	0	0	0	0	0	29	MABC0	∞
MABC0\Si2041	0	0	0	0	44	0	0	0	0	12	4	6	66	MABC0	3.67
MABC0\Si781	0	0	0	0	80	0	14	0	4	7	4	1	110	MABC0	5.71
MABC0\Sx151	0	0	0	0	62	0	0	0	2	0	2	9	75	MABC0	6.89
MABC0\Sx241	0	0	0	0	60	0	0	1	3	0	0	5	69	MABC0	12
MABC0\Sx331	0	0	0	0	52	0	0	0	1	1	5	1	60	MABC0	10.4
MABC0\Sx421	0	0	0	0	55	0	4	0	0	0	5	0	64	MABC0	11
MABC0\Sx61	0	0	0	0	56	0	0	0	2	2	2	0	62	MABC0	28
MABW0\Sa1	3	5	1	1	2	55	0	39	5	3	10	15	139	MABW0	1.41
MABW0\Sa2	0	1	0	5	0	61	5	45	1	3	6	10	137	MABW0	1.36
MABW0\Si1230	0	2	0	11	3	72	13	26	1	20	6	9	163	MABW0	2.77
MABW0\Si1664	0	2	0	0	0	65	3	15	3	0	24	8	120	MABW0	2.71
MABW0\Si2294	0	2	0	5	3	40	0	25	1	2	3	10	91	MABW0	1.6
MABW0\Sx134	5	5	0	8	0	104	5	31	0	4	3	3	168	MABW0	3.35
MABW0\Sx224	0	0	1	3	1	49	0	11	0	0	2	2	69	MABW0	4.45
MABW0\Sx314	0	8	0	4	1	107	0	9	7	4	5	1	146	MABW0	11.89
MABW0\Sx404	0	4	0	2	1	105	2	7	0	0	6	6	133	MABW0	15
MABW0\Sx44	0	1	0	5	0	52	1	25	1	2	3	8	98	MABW0	2.08

TABLO E3-3 YENİDEN EĞİTİMDEN SONRA TANIMA PUANLARI (Devam)

KONUŞMALAR	FADGO	FAEMO	FAKSO	FBASO	MABCO	MABWO	MADCO	MAFMO	MAEBO	MAEOO	MAHHO	MBEFO	TOPLAM	KARAR	G
MADCO\Sa1	0	12	0	3	24	3	62	0	24	33	14	10	185	MADCO	1.88
MADCO\Sa2	0	1	0	1	6	4	127	3	14	10	9	2	177	MADCO	9.07
MADCO\Si1367	0	7	0	0	12	26	149	6	5	15	26	11	257	MADCO	5.73
MADCO\Si1997	0	4	0	0	13	1	61	0	0	12	1	0	92	MADCO	4.69
MADCO\Si737	0	1	0	1	27	7	151	0	8	12	11	7	225	MADCO	5.59
MADCO\Sx107	0	1	0	1	4	6	71	3	10	4	6	3	109	MADCO	7.1
MADCO\Sx17	0	2	0	1	3	0	119	0	3	4	1	2	135	MADCO	29.75
MADCO\Sx197	0	0	0	0	3	0	82	0	1	13	7	4	110	MADCO	6.31
MADCO\Sx287	0	0	0	0	9	1	123	1	1	13	4	2	154	MADCO	9.46
MADCO\Sx377	0	1	0	0	7	0	99	0	6	14	7	3	137	MADCO	7.07
MAEOO\Sa1	0	3	0	6	17	1	9	11	51	38	12	11	159	MAEBO	1.34
MAEOO\Sa2	0	0	0	2	12	2	17	18	8	47	3	8	117	MAEOO	2.61
MAEOO\Si1326	0	2	0	4	11	12	55	7	49	72	7	9	228	MAEOO	1.31
MAEOO\Si1655	0	0	0	0	1	1	9	0	2	41	6	8	68	MAEOO	4.56
MAEOO\Si1956	0	22	0	1	6	0	3	1	19	19	5	12	88	FAEMO	1.16
MAEOO\Sx156	0	8	0	4	1	6	18	4	8	58	3	2	112	MAEOO	3.22
MAEOO\Sx246	0	0	0	3	19	2	12	7	3	24	4	2	76	MAEOO	1.26
MAEOO\Sx336	0	0	0	0	8	0	25	0	32	97	7	2	171	MAEOO	3.03
MAEOO\Sx426	0	0	0	0	3	5	27	0	11	83	9	6	144	MAEOO	3.07
MAEOO\Sx66	0	0	0	0	10	3	20	0	15	47	5	11	111	MAEOO	2.35
MAFMO\Sa1	0	9	0	14	0	27	0	112	1	1	3	4	171	MAFMO	4.15
MAFMO\Sa2	0	3	0	5	0	7	0	103	2	0	4	1	125	MAFMO	14.71
MAFMO\Si1569	0	15	1	16	0	11	0	34	6	8	0	0	91	MAFMO	2.12
MAFMO\Si2199	0	3	0	3	0	15	1	39	0	1	0	0	62	MAFMO	2.6
MAFMO\Si939	2	25	0	12	0	17	4	132	1	1	3	9	206	MAFMO	5.28
MAFMO\Sx129	0	6	0	2	1	2	3	121	0	1	5	6	147	MAFMO	20.17
MAFMO\Sx219	0	11	0	9	0	8	0	98	0	1	6	2	135	MAFMO	8.91
MAFMO\Sx309	0	6	0	7	0	6	0	113	9	2	3	6	152	MAFMO	12.56
MAFMO\Sx39	0	1	0	17	6	20	0	69	0	7	7	10	137	MAFMO	3.45
MAFMO\Sx399	0	2	0	9	0	18	0	88	6	8	1	6	138	MAFMO	4.89
MAHHO\Sa1	0	5	0	6	17	31	4	17	21	8	106	37	252	MAHHO	2.86
MAHHO\Sa2	0	0	0	1	15	8	2	2	10	12	75	40	165	MAHHO	1.87
MAHHO\Si1294	0	11	0	4	21	8	39	14	28	44	114	23	306	MAHHO	2.59
MAHHO\Si1924	0	16	0	1	3	3	8	5	0	5	59	28	128	MAHHO	2.11
MAHHO\Si1664	0	0	0	0	13	3	36	0	21	46	110	16	245	MAHHO	2.39
MAHHO\Sx124	0	0	0	0	9	2	22	0	15	1	56	10	115	MAHHO	2.55
MAHHO\Sx214	0	1	0	3	13	7	5	12	21	4	95	30	191	MAHHO	3.17
MAHHO\Sx304	0	0	0	1	2	10	0	8	4	2	132	12	171	MAHHO	11
MAHHO\Sx34	0	0	0	0	12	0	2	3	21	2	66	11	117	MAHHO	3.14
MAHHO\Sx394	0	0	0	0	1	18	1	3	0	4	91	20	138	MAHHO	4.55
MBEFO\Sa1	0	0	0	0	13	4	2	2	18	4	13	82	138	MBEFO	4.56
MBEFO\Sa2	0	0	0	0	4	13	2	1	8	10	15	53	106	MBEFO	3.53
MBEFO\Si1281	0	0	0	0	13	16	1	13	3	2	9	91	148	MBEFO	5.69
MBEFO\Si1911	0	5	0	3	8	0	1	10	3	3	6	66	105	MBEFO	6.6
MBEFO\Si651	0	0	0	10	16	12	1	10	11	7	29	72	168	MBEFO	2.48
MBEFO\Sx111	0	0	0	0	1	6	0	1	0	8	7	25	48	MBEFO	3.12
MBEFO\Sx201	0	0	0	0	7	2	0	7	8	3	2	54	83	MBEFO	6.75
MBEFO\Sx21	0	0	0	0	0	1	0	0	4	5	8	32	50	MBEFO	4
MBEFO\Sx291	0	0	0	0	0	9	2	3	3	9	7	66	99	MBEFO	7.33
MBEFO\Sx381	0	0	0	0	1	0	1	1	7	0	10	59	79	MBEFO	5.9

Toplam 2 hata (sarı ile boyanan konuşmalar)

ÖZGEÇMİŞ

Doğum tarihi	24.05.1951	
Doğum yeri	Kırıkkale	
Lise	1966-1969	Denizli Lisesi
Lisans	1969-1973	Orta Doğu Teknik Üniversitesi, Mühendislik Fakültesi, Elektrik Mühendisliği Bölümü
Yüksek Lisans	1973-1977	Orta Doğu Teknik Üniversitesi, Mühendislik Fakültesi, Elektrik Mühendisliği Bölümü
Doktora	1994-1999	Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektrik Mühendisliği Ana Bilim Dalı

Çalıştığı kurumlar:

1973-1977	Orta Doğu Teknik Üniversitesi, asistan
1978-1980	Vakıf Gureba Hastanesi, Biyomedikal mühendis, özel kuruluşlarda elektronik mühendisi
1980-1989	Suudi Arabistan İçişleri Bakanlığı, Elektronik/Telekom mühendisi
1990-1993	Serbest mühendis
1993-Devam ediyor	Türkiye Gazetesi Hastanesi Göz Kliniği, excimer laser servis mühendisi

**T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ**