

ENDÜKTİF ÖĞRENMEYİ KULLANARAK KONUŞMAYI TANIMA

Ethem Köklükaya, İlhan Coşkun

Özet - Konuşma tanıma, insan makine iletişiminin son derece etkin olarak kullanılmasına imkan tanır. Bir çok bilgisayarda giriş çıkış verileri interaktif olduğu için konuşma yoluyla iletişim en ideal bilgisayar arayüzünü oluşturur. Konuşma tanıma işlemi zeka gerektirir ve bu yüzden konuşma üretmekten çok daha zor bir problemdir. Genel olarak konuşma tanıma işlemi daha baştan cesaret kırıcı bir iş olduğu için halen çok daha basit sistemler yaygın olarak kullanılmaktadır. Bu çalışma bir PC ve bir ses kartıyla gerçekleştirilebilecek basit bir konuşma tanıma sistemini ortaya koymuştur. Sınırlı ve basit bir sistemle bu işlem gerçekleştirilebildiğine göre bu işlemin gerekli özel donanım ve daha hızlı bilgisayarlarla daha gelişmiş olarak ortaya konabileceğini gösterir.

Anahtar Kelimeler - Konuşmayı Tanıma, Bulanık Mantık, Simetrik Dinamik Programlama

Abstract - Recognizing speech enable to use human-machine communication as active and most excellent. The Communication by speech is the most ideal computer interface because of data inputs are interactive on many computers. Recognizing speech necessitates (requires) intelligence so that this process is more hard problem than to produce speech. In general recognizing speech discouraged process, so the simple systems are used. The study, expose the simple recognizing speech system using with a PC and a sound-card . As to realize this process with a simple and limited system. The process puts forward as developed with special hardware and faster computers.

Key words- Recognizing speech, Fuzzy Logic, Symetric Dynamic Programming

I.GİRİŞ

Konuşma insanın en etkin iletişim aracıdır. Etkinliğin ötesinde insanların kullanırken rahat oldukları ve kullanmaya en çok alışkın oldukları, iletişim aracıdır. Diğer modaliteler daha fazla konsantrasyon gerektirir, ve hareketi kısıtlar ve hatta bazıları doğal olmayan pozisyonlar sebebiyle vücut gerginliğine sebep olur.

Çağımızda bilim ve teknolojiye hızlı gelişmeler diğer sistemleri olduğu kadar eğitimsel ve sosyal sistemleri de etkilemektedir. Bilgi, gelişmiş toplumların gelişmelerinin anahtarıdır. Teknoloji ise eğitim sürecinin geliştirilmesinde önemli rol oynamaktadır. Bilgi teknolojisinin hızla gelişmesiyle, toplumların yeni teknolojik gelişmeleri izlemeleri ve kendilerine uyarlamaları zorunlu hale gelmiştir. Bu teknolojik sistemlerden birisi de "en etkili iletişim ve bireysel öğretim aracı" olarak nitelendirilen bilgisayarlardır.

1950 li yıllarda bir çok bilgisayar sisteminde bilgi girişi anahtarların açılıp kapanmasıyla olurdu ve çıkışlar ise ledlerin yanıp sönmesiyle alınırdı. Daha sonra delikli kartlar kullanılmaya başlandı. 1970 lerin sonlarında daha kolay ve etkin giriş-çıkış imkanı tanıyan CRT terminal yaygınlaştı. Klavye girişi anahtarları ayarlamaktan ve delikli kartları ayarlamaktan daha hızlıydı ve yine aynı şekilde karakterleri okumak için CRT monitörler LFD veya delikli kartların ikili sistemdeki sonuçlarını görmekten daha kolay ve daha hızlıydı. Fakat bunun yanında öğrenilmiş ve antrenmanlı bile olsa klavyeden bilgi girişi konuşmaya göre daha yavaştır. Üstelik yazma ve okuma sırasında kullanıcı giriş-çıkış işine odaklanmak zorundadır. Oysa konuşma sırasında kullanıcı çok şekilli giriş-çıkışa imkan sağlayacak diğer işlerini yapabilir.

Normal konuşma sırasında saniyede 2-3.6 kelime arası iletişim sağlanabilir. Tecrübeli yazıcıların ise saniyede yazabilecekleri kelime sayısı 1.5-2.5 kelime arasındadır. Ama bu hızda ancak hazırlanmış bir metni yazarken ulaşabilirler.[8]

El ile yazılan metinlerde ise ortalama hız 0.4 kelime/saniyedir. Aynı şekilde okuma hızını gözönüne alırsak, sessiz okumada bir okuyucu saniyede 2.5-9.8 kelime hızına ulaşabilir.[10] Bundan daha yüksek hızlara

yüksek kalıcılıkta ulaşabilmek için okuyucunun sadece ve sadece okumaya konsantre olması gerekir. bu kıyaslamalar da gösteriyor ki insan makine arayüzü olarak optimum sistem konuşma tanımadır.

Bilgisayarların iş dünyasında, idari sistemlerde ve eğitim öğretimde her geçen gün daha fazla kullanıldığını gözönüne alırsak, daha hızlı ve daha etkin bir arayüz olarak ses tanımanın karşısında büyük bir piyasa olduğu açıktır. bilgisayarlar kullanılırken harcanan zamanın büyük bir bölümü kelime işlemek, data girişi veya bu girişleri beklemek için kullanılır. insanların en doğal yolları ve doğal hızları ile iletişimine imkan vermekle etkinlik, kalite ve hız yükselecektir. Tüm bunların yanında konuşma ile uzun mesafeler üzerinden bile kolay ve hesaplı bir iletişim kullanarak özürülere de ulaşabiliriz ve onların da bilgisayar sistemlerini kolayca ve verimli şekilde kullanılmasını sağlayabiliriz.

Konuşmacı tanıma çok fazla sayıda tanınacak insan olmadığı takdirde, insan sesinin özelliklerinden dolayı kısmen çeşitli kolaylıklar içermektedir. Hatta bu özelliklerden dolayı ses tanıma daha zorlaşmaktadır. Demek istediğim ses tanıma uygulamalarının konuşmacıdan bağımsız olması için ayrıca çaba harcanmaktadır. Ses üretme (speech synthesis) ise kısmen de olsa üzerinde belli başlı bazı algoritmalar geliştirilmiş ve oturmuş bir konudur. Özellikle İngilizce için bu konuda çok iyi uygulamalar vardır.

Ses tanıma teknolojisi ile Türkiye'de ticari manada ciddi olarak ilgilenen pek fazla firma yoktur. Bu konuda daha çok çeşitli üniversitelerde doktora tezleri şeklinde çalışmalar yapılmaktadır. Ancak bu çalışmalardan şanslı olanların dışardan mali destek gördükleri de olmuştur. En son yapılan GZV ses tanıma ve ses analizi projesi; Türkçe ses tanıma tekniğinde önemli bir gelişmedir Yurt dışında ise, özellikle de Amerika'da bu konuda çalışan pek çok firma var. Amerika'da 1994 den beri 1250 civarında kuruluş bu konu ile ilgili çalışmalarda bulunmuştur, bunların 30'a yakını üniversite diğerleri ise ticari ve askeri kuruluşlardır, bu konu ile ilgili çalışmalar yapan kuruluşların içinde US Army ve US Navy de yer almıştır. Bunların haricinde telefon şirketlerinde meşrubat şirketlerine pek çok kuruluş bu çalışmalara katılmıştır.

Ses tanıma ve doğal dil işleme, Microsoft'un hesaplarına göre Dos'tan Windows'a geçişten sonraki en büyük atılım olacak. Onlara göre bu teknoloji normalde cansızmış gibi görünen bir objeyle olan (bilgisayar) ilişkinizi köklü biçimde değiştirecek. Ancak Microsoft'a göre bu teknoloji birden ortaya çıkabilecek bir teknoloji değil. Sonraki on içinde ortak çalışmalar sonucunda yavaş-yavaş gelişip yerine oturacak.

Bu teknolojiyi 4 başlık altında incelemek mümkündür.

1. Telefonda ses (konuşma) tanıma: Komutları anlayan bilgisayarlar ile, bu konuda zaten bayağı yol alındı. Telefonla servis veren veya verebilecek olan şirketler için bu konu büyük önem arz etmektedir.

2. Dikte ettirme: Sizin sürekli konuşmanıza kısıtlı olarak izin veren, mevcut yazılılar var. Örnek olarak "Micro Dictation" ve "Dragon Dictate" verilebilir. programların doğruluk oranları %90-95'ler civarında. Ancak hala çalışmalar devam etmektedir çünkü bu oranı 3000 kelimelik bir makalede pek çok boş kalmasına sebep olmaktadır.

3. Konuyu anlayan tanıyıcılar: Bu alandaki çalışmalar sadece söylenen kelimeyi anlamayı değil ne demek istediğinizi yani söylediğiniz cümlenin anlamı çıkarmayı hedeflemektedir. Bu hedefe ulaşmak öncelikle uzmanlaşmış uygulamalar yapmayı ve öncelikle sınırlı belirli alanlar içinde kalmayı, bunu başardıktan sonra genel kullanıma geçmeyi düşünmektedirler

4. Doğal dil anlama: Bilgisayarlar, sürekli konuşmayı diyalogları anlayabildiğinde, bu, teknolojideki büyük devrim olacaktır. Henüz bu teknoloji gelmesine en az 15 yıl varmış gibi gözükmektedir. Ama bu teknoloji yayıldığında hayatınızda büyük değişikliklere neden olacaktır. Bilgisayarınıza yapmasını istediğiniz şeyi normal bir cümle şeklinde söyleyecek ve isteğinizi anında ulaşacaksınız. Microsoft ve IBM beraber bu hedefe ulaşmak için çalışıyorlar. Özellikle Microsoft işletim sistemine bu teknolojiyi yerleştirmenin yollarını arıyor. Belki inanmak zor olabilir ama gelecekte bir gün bilgisayarınızın etrafında ne söylediğinize dikkat etmeniz gerekebilir.

Ayrıca ses tanıma problemine getirilen farklı çözümler vardır. Bunlar tanınması gereken konuşmanın kesikli mi yoksa sürekli olduğundan etkilenirler. Yani konuşmayı da aynı teknikle tanımak zordur. Kesikli konuşmanın tanınması daha kolaydır ve kelime-kelime yapılması gayet uygundur. Sürekli bir konuşmanın kelime-kelime tanınması daha zordur çünkü kelimeler nerede başlayıp nerede bittiği bilinmemektedir. Dolayısıyla sürekli tanıma genelde fonem bazında yapılmaktadır. Fonem anlam içeren en küçük ses demektir. Yani fonem bir heceden daha kısa bir sestir. Normal bir hecede başlangıç-orta-bitiş olmak üzere genelde üç fonem bulunur. Ancak fonemleri de birbirinden kesin hatlarla ayırmak pek mümkün değildir. Bu nedenle fonemleri tanıyacak ve temsil edecek çeşitli sistemler geliştirilmiştir. Bunların başında Hidden Markov Model (HMM) gelmektedir. Ses tanıma problemi kişiye bağımlı kişiden bağımsız ya da kişiye uyum sağlayan tarzlarla çözülebilir. Ses tanıma probleminde önemli olan diğer bir nokta da tanınacak kelimelerin (kelime haznesi) sayısıdır.

II KONUŞMA DALGA FORMU

II.1 Fiziksel Görünüm

konuşma dalga formu akciğerleri, ses yolunu dili ve ağız kontrol eden ayrı uyarılardan oluşan serilerce meydana getirilir. Sonuçtaki ses sadece o anki uyarıya değil

yardımcı boğumlama [coarticulation] etkisi yapan önceki uyarılara da bağlıdır.

Konuşma dalga formunu oluşturan etkenler farklı farklı iken, dalga formunun kendisi de iki boyutlu nonlinear bir sürekliliktir.[8] Zaman içinde herhangi bir anda dalga formunu ne kadar kesin ölçtüğümüz hiç farketmez; daha ince daha kesin bir ölçüm her zaman mümkündür. Bununla birlikte konuşma dalga formunun sürekliliğinin tanıma ile ilgisi azdır. Bunun delili bizzat, konuşma dalga formunu bir sunumunu beyne taşımadan önce ayrı ayrı örnekleyen ve filtre eden kulağın içindedir. Bu bilgiye dayanarak konuşma dalga formunun tanıma için gerekenden çok daha fazla bilgi içerdiğini anlayabiliriz

II.2 Enformasyon Teorisi

Enformasyon teorisi konuşma dalga boyunun bilgi içeriğinin hesaplanmayı sağlar. Enformasyon teorisi bit ile ölçülen sembol başına ortalama bilgi miktarını entropi olarak adlandırır.[9]

II.3 Konuşma Özellikleri

Konuşma tanıma işlemi data edinme, filtreleme, özellik tanıma ve sınıflandırma başlıkları altında dört işleme bölünebilir.

Aşağıdaki özellikler -tanıma birimi ne olursa olsun-konuşulan bir sesi tanımadaki kullanılabilir.

Yüksek sesli ve alçak sesli. Vokalik ve nonvokalik sesler. Ünsüz ve ünlüler. Gergin ve gevşek.. Burundan ve ağızdan. Strident ve mellow. Grave ve Acute. Sıkı ve yaygın.Flat vs plain.[6]

II.4 Gösterme Ünitesi

Konuşma birkaç bileşene ayrılabilir: Allafonlar, fonemler (sesbirimleri), difonlar (sesler arası geçiş sesleri), heceler ve kelimeler.

Her sunumun kendine göre avantajları ve dezavantajları vardır. [8]Phone (ses) konuşma sesinin en küçük birimini gösterir. Aynı bilgi içeriğini içeren ses kümelerini temsil etmek için allafonlar kullanılır. Fonemler lisan içinde benzer şekilde iş gören allafonlar topluluklarıdır. IPA 44 fonem içerir (Bkz Tablo 1). Her birinin sesi yanındaki kelime içinde altı çizili şekilde gösterilmiştir

Difon, komşu seslerin sabit durum merkezleri arasında geçişli bir sestir. Hece temel olarak 'çekirdekdeki bir sesli harf ile bununla fonksiyonel ilişki içerisindeki komşu sessiz harfler' şeklinde tanımlanabilir.[8] Son olarak kelimeye geldiğimizde, bir dili konuşan herkes içgüdüsel olarak kelimenin ne olduğunu bilse de, fonolojik olarak dili tanımlamak zordur. Bununla beraber kelimeyi "bir tam mesaj içerebilen en küçük bilgi birimi" olarak tanımlayabiliriz.

Tablo 1 IPA Uluslar arası Fonetik Alfabeti (International Phonetic Alphabet)

Phoneme	Example	Phoneme	Example
æ	pat	ŋ	sing
e	pay	o	toe
ə	about	U	book
a:	father	a	pot
er	care	u	boot
b	bet	ɔ	bought
tʃ	church	aU	out
d	debt	ɔɪ	boy
ε	bet	p	pet
i	bee	r	rent
f	fire	s	sat
g	get	ʃ	shut
h	hat	θ	thing
ɪ	bit	t	ten
aɪ	by	ð	that
ɪr	pier	ʌ	but
ɔr	butter	ʒr	term
d	judge	v	vat
k	kit	w	wit
l	let	y	you
m	met	z	zoo
n	net	ʒ	azure

III. KONUŞMA TANIMA MODELLERİ

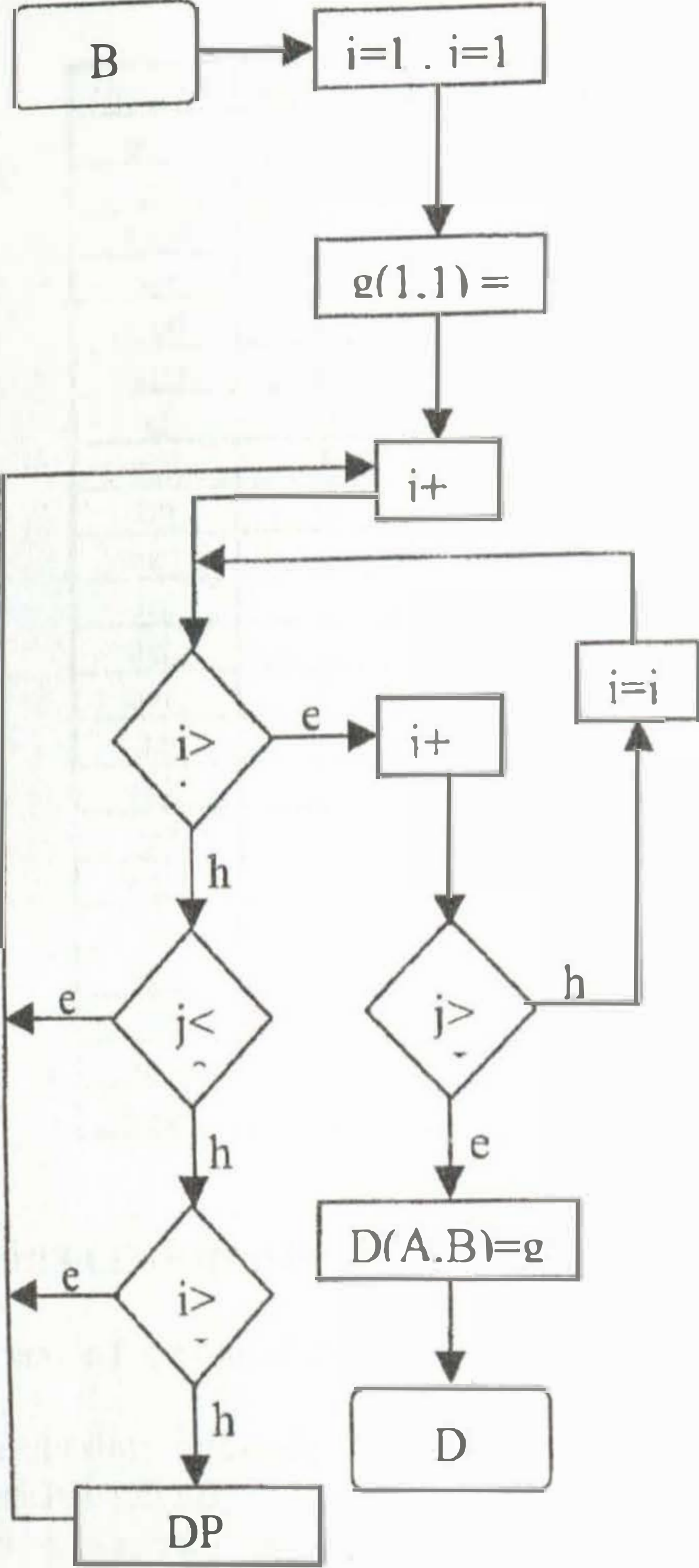
III.1 Bilgisayar Tabanlı Konuşma Tanıma

Konuşma tanıma için bir bilgisayarın kullanılması ilk olarak 1960 yılında gerçekleşmiştir. Bu tarihte Denes ve Matthews zaman normalizasyonu kavramını tanıttı. Zaman normalizasyonu, farklılık gösteren dalga formu uzunluklarının kayıtlı şablonlarla karşılaştırılmak için otomatik olarak sıkıştırılması ve genişletilmesine izin veriyordu. Ayrıca bilgisayar kullanımını araştırmacılar daha geniş sözlüklerle deney yapma ve daha karmaşık algoritmaları tanıma imkanı veriyordu. Konuşma tanımadaki en büyük ilerleme 1971 yılında ARPA'nın "Konuşma Anlama Araştırması" (Speech Understanding Research - SUR) projesini hizmete sokmasıyla sağlandı. Proje, birden fazla konuşmacının sürekli konuşmasını başarıyla (%90'nın üzerinde) tanımasını sağlayan bir sistem geliştirdiyordu. ARPA'nın SUR projesinin getirdiği gelişmeler, çok sayıda yeni tekniğin sayısını arttırıyordu ve konuşma tanıma alanına dikkatleri çekiyordu.

III.2 Konuşma Tanıma Teknikleri

Konuşma tanıma modelleri 4 ana kategoriye ayrılabilir: Akustik sinyaller, konuşma üretimi, duyuşsal netlik [Sensory reception] ve konuşma sezisi.

Şekil 1 Simetrik dinamik programlama akış diyagramı(e:Evet, h:Hayır)



Kısa zaman aralıklarında dalga formu inceleme teknikleri iki kısma ayrılabilir: zaman domeninde inceleme ve frekans domeninde inceleme. Zaman domeni analizi konuşma dalga formunu bir zaman ve genlik fonksiyonu olarak inceler. Frekans domeni analizi konuşma dalga formunu bir zaman ve frekans fonksiyonu olarak inceler. Bu iki temsile dayanarak, konuşma dalga formundaki belirleyici nitelikler ayrılıp kelimeler tanımlanabilir.

Yeni yapılan araştırmalar pitch periyodunu tahmin ederken, dalga formunun ortasını kaldırmanın sinyalin örnekten-örneğe tutarlandırmasını azaltmaya yardımcı olduğunu göstermektedir. Clipping (kesik) adı verilen bu nonlineer teknik alçak genlik bilgisini ortadan kaldırmak için örneklerin değiştirilmesini verimli bir şekilde gerçekleştirebilir.

Lineer tahmin katsayısı, aşağıdaki yinclemeli algoritma kullanılarak bulunabilir. Algoritmanın N^2 işlemi gerektirdiğine dikkat edin.

$$E_0 = R_0$$

$$\alpha_n = \frac{R_n - \sum_{i=1}^{n-1} \alpha_i^{(n-1)} R_{|n-i|}}{E_{n-1}} \quad \text{for } 1 \leq n \leq N$$

$$A_n^{(n)} = \alpha_n$$

$$A_i^{(n)} = A_i^{(n-1)} - \alpha_n A_{n-i}^{(n-1)} \quad \text{for } 1 \leq i \leq n-1$$

$$E_n = (1 - \alpha_n^2) E_{n-1}$$

$$a_n = a_n^N \quad 1 \leq n \leq N$$

Spektrum bir kez elde edildikten sonra, forma frekansları ve konuşma dalga formunun pitch'le belirlenmek üzere analiz edilebilir. Frekans domeninde gerçekleştirilen analizler zaman domeninde gerçekleştirilenlerden daha kesin olmasına rağmen hesapsal bugünün standartlarında bile oldukça yüksektir. Gerçek zamanlı konuşma işleme, pc'lerde uzmanlaşmış aygıtlar olmadan mümkün değildir. Bununla birlikte gelecek yıllarda, görünür bir performans düşüşü olmadan bu tekniklerin avantajlarını almak mümkün olacaktır. [5]

III.3 Zaman Normalizasyonu

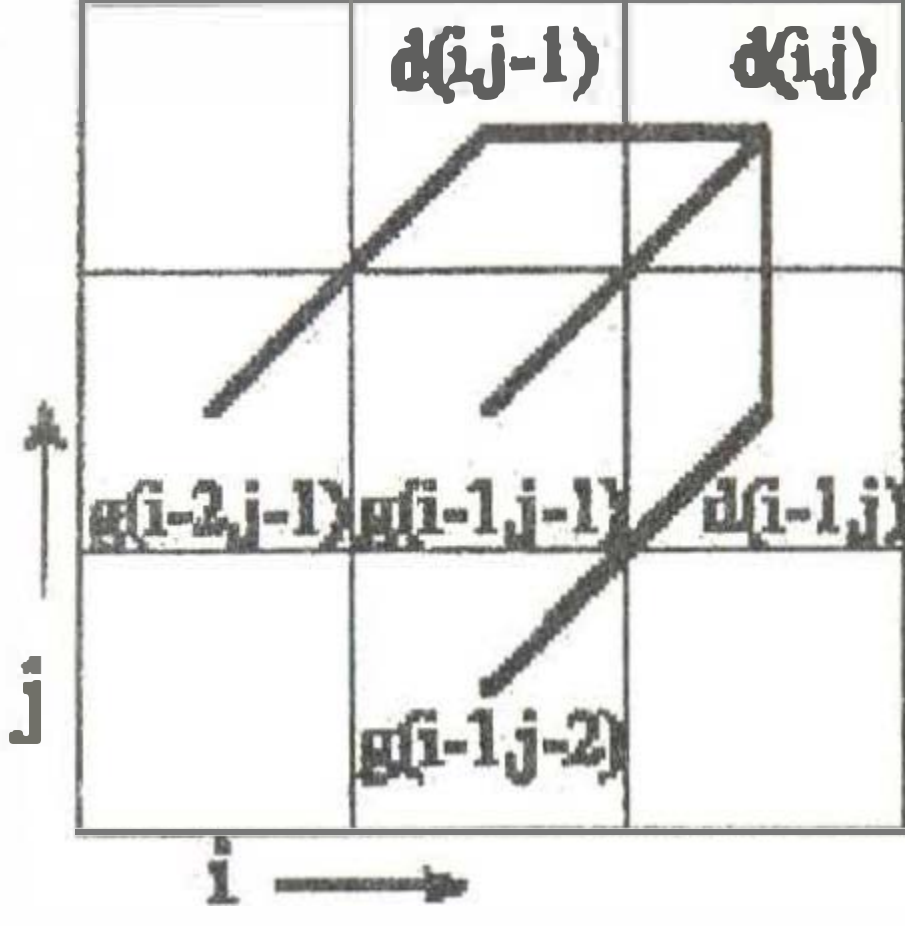
Dinamik Programlama (DP) iki dalgaformu arasındaki optimal biçimde farkı ortadan kaldırmaya çalışan tekniklerden biridir. Algoritma, bir dalgaformunun eksenini diğeriyle üstüne kaydırarak işler. Bununla beraber algoritma, dalgaformunu yalnızca uzatma ve sıkıştırmaktan çok, dalgaformlarını denkleştirmeye çalışır ve böylece benzerlikler sağlanır ve zamanı hizaya sokar.

Simetrik dinamik programlama, farklı zaman eksenlerine sahip iki dalgaformu için ortak bir zaman eksenine meydana getirir.

DP-denklemini $g(i,j)$ en düşük mesafeli yolu ele almayı kapsar ve bu nedenle bir eğim olarak incelenebilir. Şekil 4.2 de gösterildiği gibi eğim sınırlaması 1 olmak üzere, yol incelenmelidir. Eğim sınırlaması için, dikey veya köşegen adımlamadan önce en fazla bir kere yatay veya dikey yönlerde fonksiyon adımı gerektirir.

Eğim sınırlaması bir'e eşitlendiğinde, minimum eğim bulmak için 3 yol incelenmelidir

Tablo 2 Eğim yolları

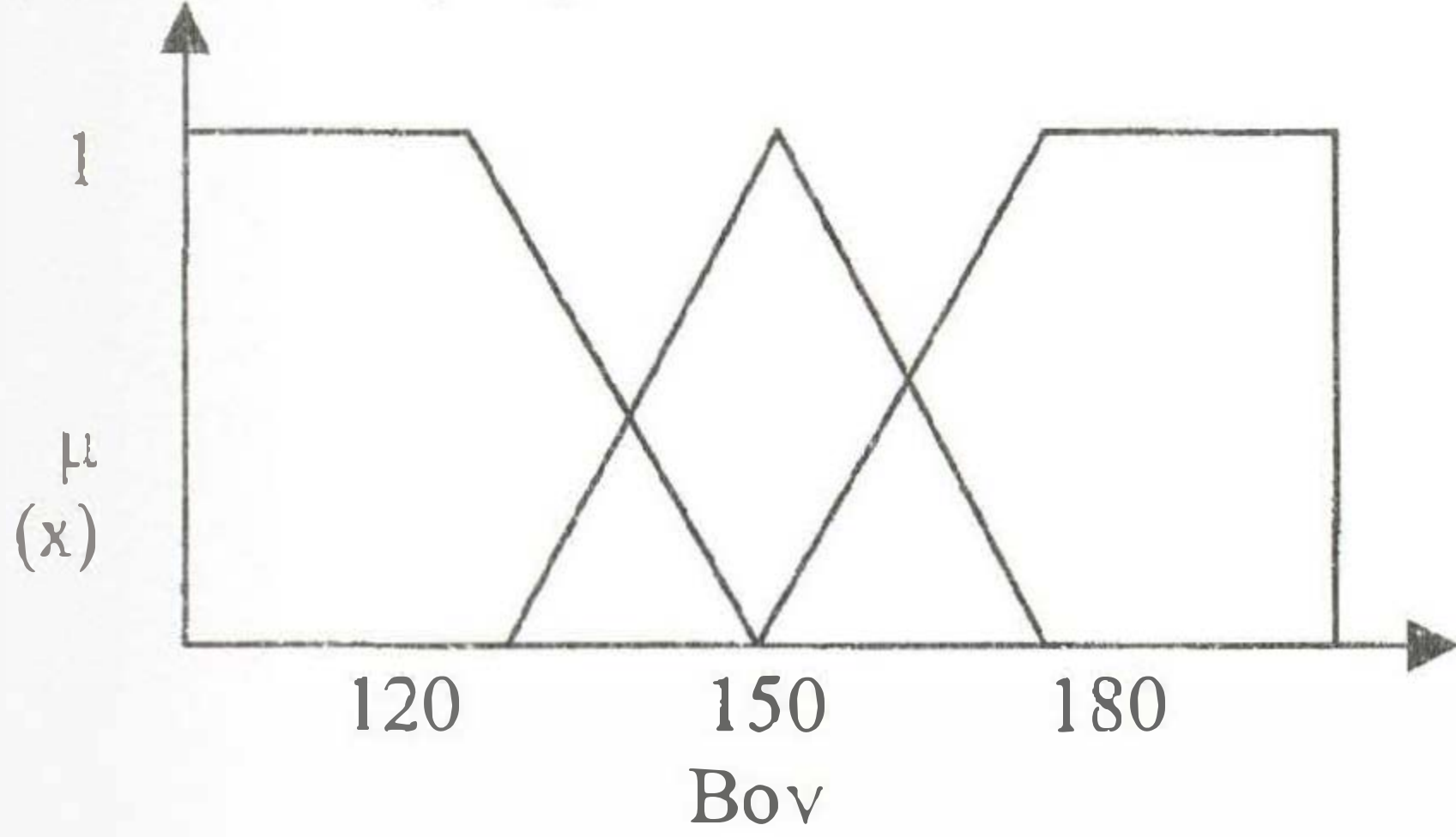


IV. FUZZY LOGIC

IV.1 Bulanık Kümeler (Fuzzy Sets)

Bulanık kümeler klasik kümelerin bir süper kümesidir. Bulanık kümedeki her eleman kümedeki üyelik derecesini gösteren bir reel sayı ile ilişkilidir. Bulanık kümeler genelde üyelikleri veya gerçekleştirme değerleri $[0,1]$ kapalı aralığında değerler alan elemanlar kümesi olarak ifade edilir.

Şekil 2 : Boy uzunluğunu gösteren bir bulanık küme.

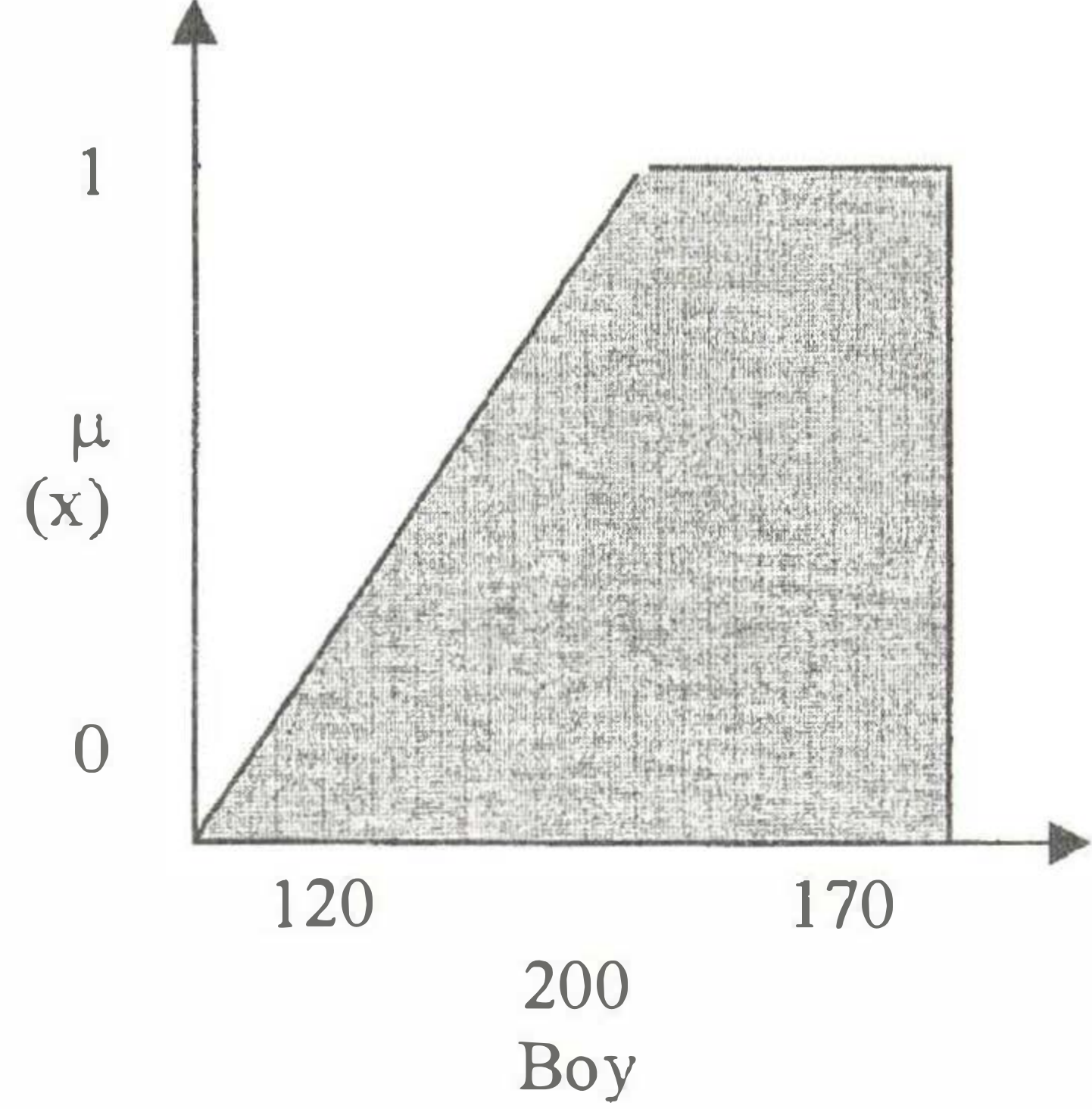


Uzun boyluluk kavramı için basit bir bulanık kümeyi inceleyelim (Bkz.Şekil 2). Bu küme içeriği gökdelenler olsaydı iyi bir gösterim olmayacaktı, fakat 15-20 yaş arasındaki Türk kadınlar için yerinde bir gösterimdir. 180 cm uzunluğundaki kadınlar kümenin tam üyesi iken 120 cm ve altındakiler Uzun boyluluk kümesinde üyeliği olmayacaktır. Özel bir üyeliği tanımlamak için kişinin yüksekliği yatay aksis üzerinde bulunarak, yüzey fonksiyonu izlenerek dikey aksis üzerinden üyelik derecesi $\mu(x)$ okunur. Belirli bir içerik için boy uzunluğu bir bulanık küme olarak gösterilmiştir. 120 cm nin altındaki hiç kimsenin küme üyeliği mevcut değilken, 170 cm üzerindeki kümeni tam üyesidir. Aradaki boylar orantılı olarak dağılmıştır.

Bulanık kümenin bir konsepti ifade etme ve içeriğe sahip olma fikri linguistik (dilsel-sözlü) değişkenler ile daha da genişletilmiştir. Bir sözlü değişken bir bulanık aralığa

atanır, ve bir takım bulanık kümesi tüm içeriği ifade eder. Şekil 5.2 yükseklik konsepti için bir örnektir. Değişken üç bulanık kümeden oluşur: kısa, orta, uzun. Bulanık kümelerle birlikte sözlü değişkenlerin yorumlanmasında kontekst geniş bir rol oynar. Yatay aksis yeni aralık olan temel değişkeni tanımlar. Verilen temel bir değer için, her bulanık kümedeki üyelik derecesi tanımlanabilir. Bu işlem 'bulanıklaştırma' olarak adlandırılır.[1]

Sözlü değişkenler tek bir konsepti ifade eden bulanık kümelerle ilişkilendirilmiştir.



Şekil 3 Yüksekliği ifade eden bir sözlü değişken.

IV.2 Bulanık Mantık

Bulanık mantık klasik mantığın bir süper kümesidir. "Mantık, sistemlerin taslağı ve onun tüm mümkün formlarında nedenselleştirme ilkeleridir. Klasik lojik doğru veya yanlış olması öngörülen önermeler ile uğraşır." Bulanık mantık önermelerin üyeliklerini üyesi olmayan-yanlış ve tam üye-doğru arasında kadencalı üyelikleri içerecek şekilde genişletir. Bir önerme üye değilse veya tam üye ise bu durumda bulanık mantık klasik mantığa indirgenmiş olur. Bulanık mantığın önemli bir faydası belirsiz önermeleri değerlendirme iskeleti oluşturmaktır. Bu değerlendirme yaklaşık değerlendirme olarak adlandırılır ve bulanık kontrol sistemleri oluşturmada kullanılır.

IV.3 Bulanık sistemler

Bir bulanık sistem bir bulanıklaştırma altsistemi, bir bulanık arayüzü, bulanık kural tabanı ve netleştirme (defuzzification) alt sisteminden oluşur. Şekil 5.3 genel bir bulanık kontrol sistemini göstermektedir. İşlemin o anki durumunu gösteren yeni girişlerin kümesi verildiğinde; bulanıklaştırma alt sistemi onları geçerli

bulanık kümelerle çevirir ve o kümelerdeki üyelik derecelerini tanımlar. Bulanık girişler daha sonra arayüz motoru tarafından bulanık çıkışları belirlemek için kullanılır. Arayüz motoru kural tabanı ile etkileşir ve hangi kuralın geçerli olduğunu belirlemek için girişleri kullanır. Kurallar bağımsızdır ve bununla beraber paralel olarak değerlendirilebilirler. Çıkışlar mümkün olan çıkışlar uzayında tanımlı bulanık kümelerin bir kümesidir. Bu bulanık çıkışlar işlemi kontrol etmek için kullanılan çıkışları üretmek üzere netleştirilir. netleştirme için bir çok metod vardır fakat, en popüler olanları: centroid, center of maxima ve mean of maximadır.

Her bulanık kontrolcü en az dört temel alt-sistemden oluşur. Bulanık arayüz motorunun bulanık kural tabanı ile etkileşen tek alt-sistem olduğuna dikkat ediniz. Bu etkileşim paralel olarak tezahür edebilir.

V. İZOLE EDİLMİŞ RAKAM TANIMA

V.1 Rakam Dalgaformları

İzole edilmiş kelime tanıyıcılarının çoğunluğu tanımayı gerçekleştirmek için şablon karşılaştırma denen bir işlem kullanırlar. şablon karşılaştırma kütüphanedeki her kelime için dalgaformu elde etmeyi içerir. her şablon kendi ismiyle ilişkilidir. sisteme bilinmeyen bir dalgaformu sunulduğunda en uygun eşi bulmak üzere her şablonla karşılaştırılır. bir eş bulunduğunda, bilinmeyen dalga formuna şablonun adı verilir.

Bu Çalışma için tasarlanan tanıma sistemi genel kelime tanıma yapabilmesine rağmen, öncelikle izole edilmiş rakam tanıma gerçekleştirmek üzere tasarlanmıştır.

Tablo 3 Rakam temsili rakamlar [0-9] fonem ve hece olarak gösterilmiştir.

Number	Phoneme	Syllable
zero	z i' r o	ze • ro
one	w ʌ n	one
two	t u	two
three	θ r i	three
four	f ɔ r	four
five	f aɪ v	five
six	s i k s	six
seven	s e v' ə n	sev • en
eight	e t	eight
nine	n aɪ n	nine

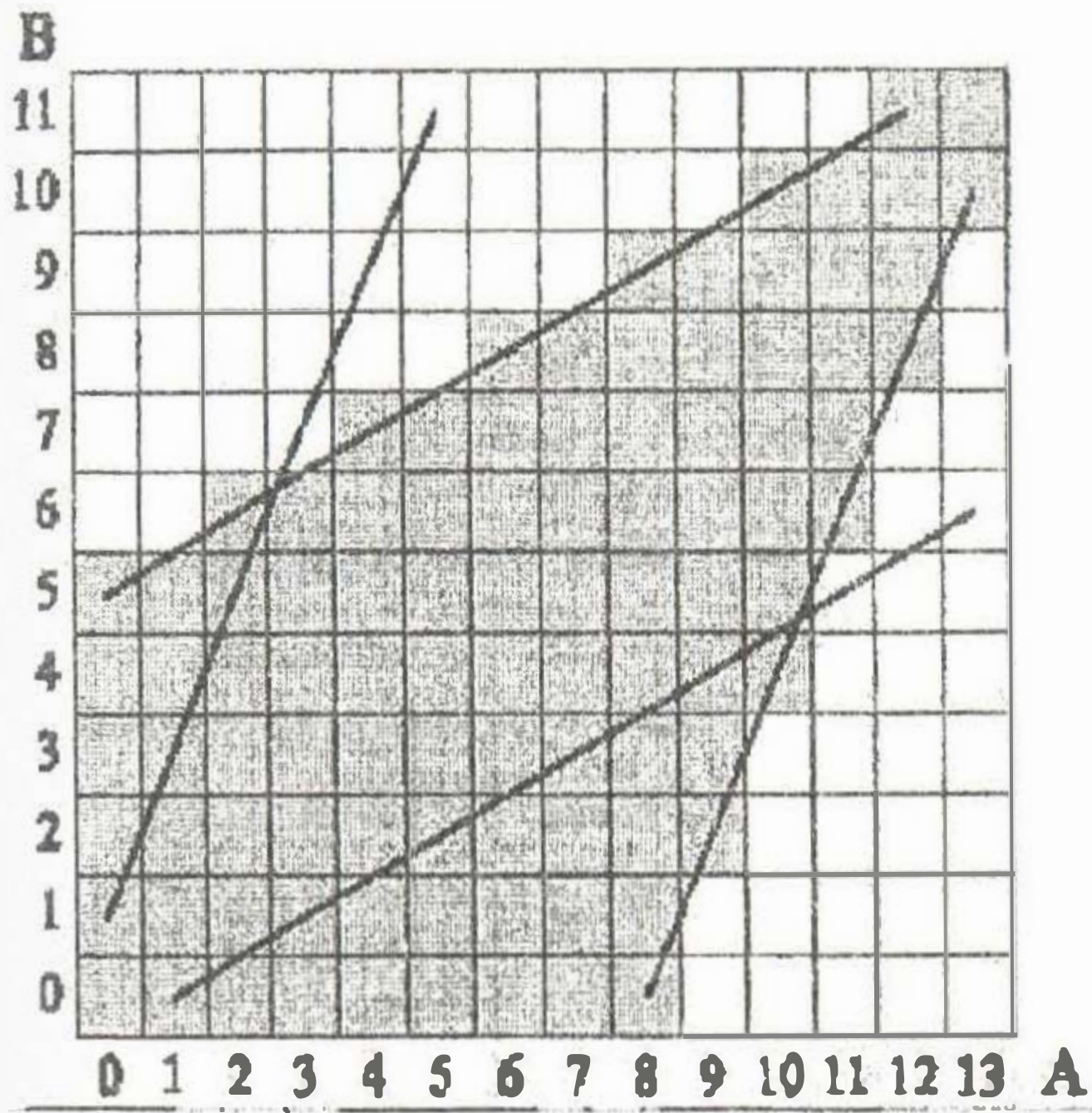
V.2 Bulanık Konuşma Tanıyıcı

Bu çalışmayı geliştirmek için 16-bit ses kartı donatılmış bir PC kullanılmıştır. Dalgaformları dalga kodlu modülasyon (PCM) kullanılarak WAV formatında saklanmaktadır. 8 bit PCM minimum 0h, or 80h ve maksimum FFh ile bir işaretlenmiş formatıdır. Sistem, program açılışında her rakam için şablonları okur. Bir rakam okunduktan sonra, minimumu 0h ve maksimum FFh olması için normalize eder.

Çalışma şu sisteme göre tasarlanmıştır. Tanıma için gerekli olan örnekler, daha önce kaydedilmiş WAV dosyaları veya ses kartı ile kaydedilmiş canlı kayıtlardan alınabilir. İstikrarlı frekansları sağlamak için, örnekler şablonlarla aynı örnekleme frekansında kaydedilir. Kaydetme işlemi tamamlandıktan sonra, sistem örneklerin başını ve sonunu belirlemek üzere bölümlendirme işlemi uygular. Bu noktada, değiştirilmiş dalgaformları hafızada saklanır ve doğru şekilde kayıt edildiğinden emin olmak için hoparlörden çalınır. Bir örnek yüklendikten veya kaydedildikten sonra pencerenin üst kısmında zaman-genlik grafiği görüntülenir. Analiz işlemi sırasında, pencerenin üstünde bilinmeyen dalgaformu görüntülenirken, alt tarafta karşılaştırılacak olduğu şablon görüntülenir. Bir durum çubuğu, her şablon için işlemi gösterir. Analiz tamamlandıktan sonra, sonuç (sınıflandırma) dalgaformunun üzerinde görüntülenir.

V.3 Simetrik Dinamik Programlama

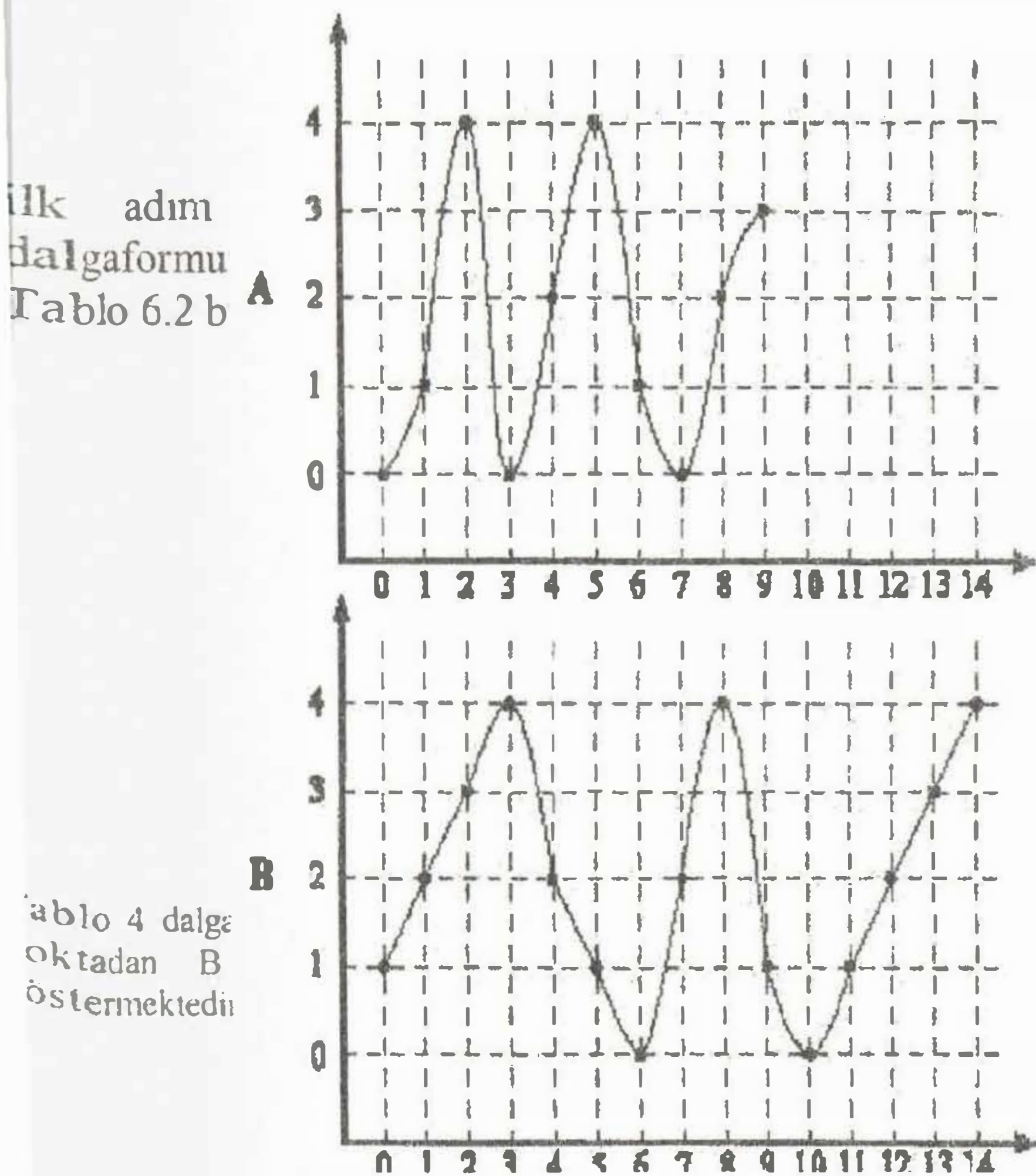
Bu tezde geçmiş başarısından dolayı simetrik dinamik programlama seçilmiştir. Dinamik programlama, bir şablonla bilinmeyen arasındaki en iyi zaman hizalamayı bulurken, özellikleri de izole eder ve karşılaştırır. Bu tez için kullanılan gerçekleştirme zaman-genlik değerlerinde sadece dinamik programlama kullanılmıştır; spectral analiz kullanılmamıştır. Tanıma işleminin toplam kesinliğini sınırlayan spectral bilgiyi göz ardı ederken, uzmanlaşmış donanım olmadan dinamik programlama, Fourier analizi veya doğrusal tahmin kodlaması (Linear Predictive Coding) gibi ek karmaşık analizler için çok zaman alıcıdır. Dinamik programlama, işlem süresi ve sistem belleğinde nedensiz talepte bulunabilecek $O(n^2)$ kompleksliğine sahiptir. Neyse ki karmaşıklığı azaltmak üzere bazı kısıtlamalar kullanılabilir. Bir eğitim kısıtlaması sağlanarak, gözlem alanı sınırlanır. Sakoe ve Chiba'nın çalışmalarına dayanarak, optimal bir eğitim kısıtlaması seçilmiştir.



Şekil 4 Dinamik programlama algoritması tarafından araştırılan alan. 4 çizgi, bütün olası çözümlerin alanını sınırlamaktadır. Gri alan olası çözümlerin aranması gereken alanı göstermektedir.

B \ A	0	1	2	3	4	5	6	7	8	9
0	2	0	6							
1	4	4	4	6						
2	6	5	5	9						
3	8	11	6	10	10					
4	4	10	8	10	10					
5	2	4	7	10	11	15				
6	0	4	8	7	9	18				
7	4	2	4	9	7	9	14			
8	8	5	2	6	9	7	10			
9	2	8	5	4	5	10	9	8		
10	0	4	8	4	7	12	8	9		
11			3	6	6	9	12	8	9	
12					6	8	11	12	8	
13							10	15	9	8
14									16	9

Dinamik Programlama algoritması, belki de en iyi bir örnekle açıklanabilir. Şekil 5 normalize edilecek 2 dalgaformunu göstermektedir; A dalgaformunun uzunluğu 10 ve B dalgaformunun uzunluğu 15'dir.



Şekil 5. Simetrik dinamik programlama örneği. Örnekteki iki dalgaformu A ve B gösterilmektedir.

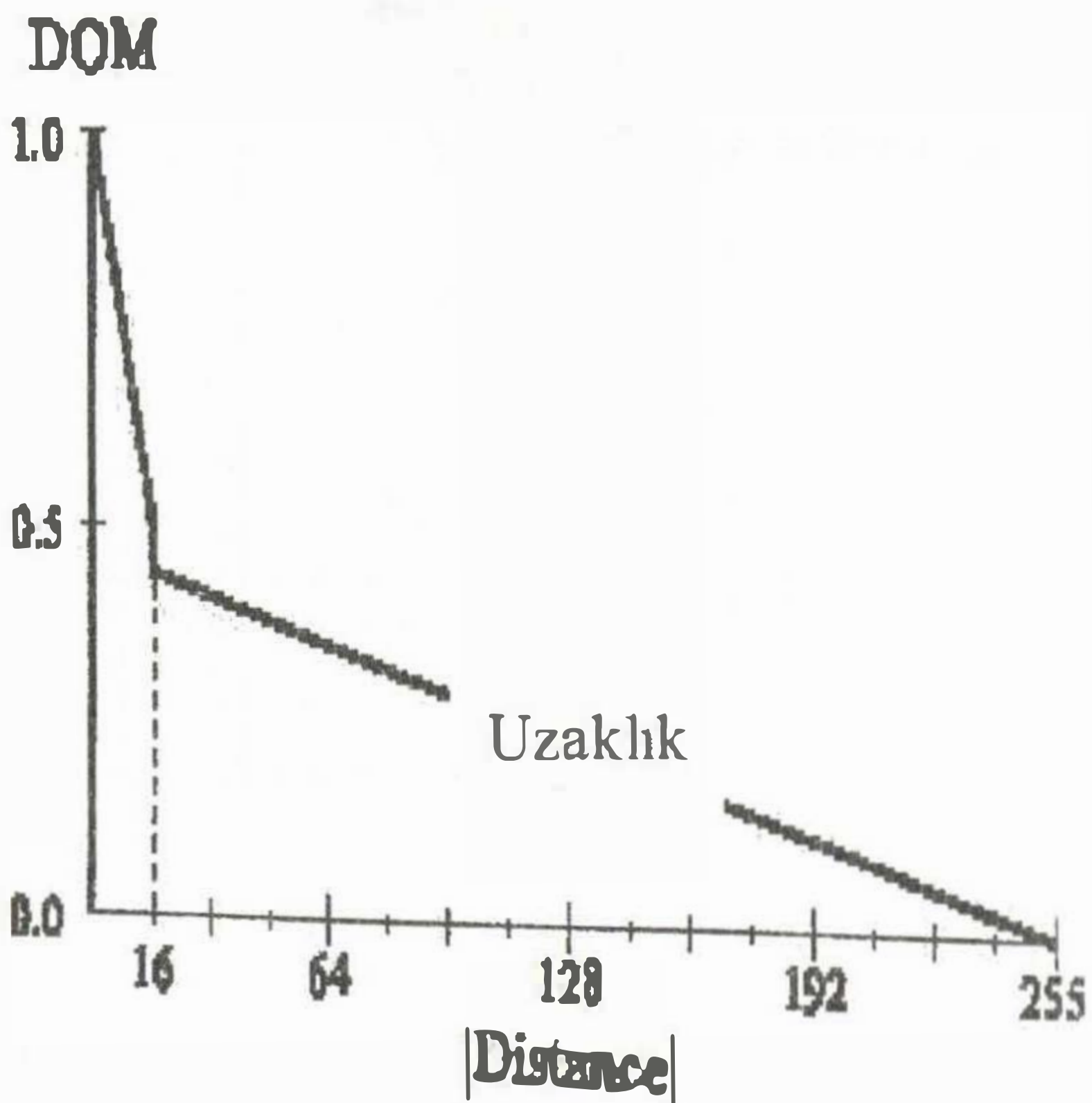
Bir sonraki adım tekrarlamalı eğimleri hesaplamaktır. Tablo 4 eğim hesaplarını göstermektedir. Son adım (A-1,B-1)=(9,14)deki eğimin $A+B=25$ 'e bölümünün normalize edilmesidir. Sonuç 0.36, iki dalgaformu arasındaki en iyi mesafedir. Simetrik dinamik programlama algoritması, bir klasik "crisp" ("crisp"=gevrek,taze,çabuk,soğuk... Seç beğen kullan.) Form ve bir karmaşık form olmak üzere iki formda uygulanmıştır. Crisp form, her adımda iki dalgaformunun genlikleri arasındaki farkın mutlak genliğini alarak mesafeyi hesaplar. Daha sonra iki dalgaformu arasındaki toplam mesafe ölçümü bu mesafelerin toplamının, dalgaformlarının uzunlukları toplamına bölümü ile bulunur. Kısa ve uzun dalgaformlarının ortak mesafe ölçümü olduğundan bu son normalizasyon gereklidir. En iyi eş/karşılaştırma, en kısa toplam uzunluğu olan dalgaformunu seçerek elde edilir. Bundan sonraki kısım bulanık (fuzzy) gerçekleştirilmeyi açıklamaktadır.

Tablo 5 Dalgaformu mesafe matrisi her hücre A dalgaformu üzerindeki bir noktadan B dalgaformu üzerindeki bir noktaya olan uzaklığı içerir

B \ A	0	1	2	3	4	5	6	7	8	9
0	1	0	3							
1	2	1	2	2						
2	3	2	1	3	1					
3	4	3	0	4	2					
4	2	1	2	2	0	2				
5	1	0	3	1	1	3				
6	0	1	4	0	2	4	1			
7	2	1	2	2	0	2	1			
8	4	3	0	4	2	0	3	4		
9	1	0	3	1	1	3	0	1		
10	0	1	4	0	2	4	1	0	2	
11		0	3	1	1	3	0	1	1	
12				2	0	2	1	2	0	1
13						1	2	3	1	0
14								4	2	1

V.4 Bulanık Simetrik Dinamik Programlama

Bulanık gerçekleştirme bütün dalgaformlarının kararsızlık içerdiğini varsayar. Bu kararsızlık konuşmacı değişikliğinden, dalgaformu nicelendirmeden, gürültüden, ve konuşma tanıma işleminin tam olarak tanımlanamayışından ileri gelir. Her genlik bulanık bir sayıyla ifade edilir. Bulanık bir sayı, belirli bir zaman aralığı etrafındaki sayılar kümesi olarak düşünülebilir. Örneğin, x'e yaklaşan tamsayılar, sayının x'e ne kadar yakınsa kümedeki üyelik derecesinin o kadar yüksek olduğu bir bulanık küme ile ifade edilebilir. X'in maksimum üyelikte olması gerektiğinden karmaşık küme normaldir.



toplam mesafenin üyeliği, optimal yol üzerindeki her noktanın derecesinin ortalamasıdır. her şablon bilinmeyen dalgaformuyla karşılaştırılmasından elde edilen sonuçlar, şablonun bilinmeyen dalgaformuna benzerliğini gösteren, şablonlar ve üyelik dereceleri içeren bir karmaşık kümedir. sonuçların karmaşıklığını gidermek ve en benzer kategoriye seçmek için maksimum üyelik derecesi belirlenir. daha sonra maksimum üyelik kümesinin en düşük mesafeli elemanı en uygun eş olarak seçilir.

VI. SONUÇ

Üretilen program ile ses tanıma gerçekleştirildi. Ses kayıtları sade bir ortamında, normal gürültüler (Araba sesleri, uzakta konuşan insanlar) de dahil olacak şekilde yapılmıştır. Kullanılan ses kartı "SoundBlaster" uyumlu (16 Bit) sıradan ses kartıdır. Hatası 1/1000'in altında olan testler başarılı olarak kabul edilmiştir. Ses tanıma testlerinde çıkışlardan en büyük olan 0,7'den büyük olmak ve kendisine en yakın çıkışın kendisinden en az 0,4 kadar ufak olması o çıkışın seçilmiş kabul edilme şartı olarak alınmıştır. Tüm testler başarılı olmuştur. Ancak, ilginçtir, testlerin %95'inde YSA'nın coder devreleri için eğitilmesi, decoder devrelerinden daha hızlı ve kolay olmuştur.

KAYANGLAR

1. Bezdek, James C. Analysis of Fuzzy Information: Volume III Applications, CRC, 1987
2. Chen, C:H: Digital Waveform Processing and Recognition, CRC, 1982
3. Cooper, William R., Speech Perception and Production, Studies in Selective Adaption, Able Publishing, 1979
4. Cox Earl, The Fuzzy Systems Handbook Academic Press, 1994
5. De Mori, Renato, Computer Models of Speech Using Fuzzy Algorithms, Plenum Press, 1983
6. Kandel, Abraham, Fuzzy Mathematical Techniques with Applications, Addison Wesley, 1986
7. Klir, George J., Fuzzy sets and Fuzzy Logic Theory and Applications, Prentice Hall, 1995
8. Lea, Wayne A., Trends in Speech Recognition, Prentice Hall, 1980
9. Newell, Allen, Speech Understanding Systems Final Report of Study Group, North-Holland Publishing, 1973
10. Ross, Timothy J., Fuzzy Logic with Engineering Applications, McGraw-Hill, 1995
11. Waibel, Alex and Kai-Fy Lee, Reading in Speech Recognition, Morgan Kauffmann, 1990