

Otomobil İçindeki Cihazların Sesle Kontrolüne Yönelik Konuşma Tanıma Sisteminin Gerçek Zamanlı Laboratuar Uygulaması

Ersoy KELEBEKLER, Melih İNAL

ÖZET

Bu çalışmada konuşmacıdan bağımsız konuşma tanıma sisteminin gerçek zamanlı laboratuar uygulaması amaçlanmıştır. Konuşma tanıma sistemlerinin verimini etkileyen en önemli etken, sistemin konuşmacıya bağımlı veya konuşmacıdan bağımsız olmasıdır. Konuşmacıdan bağımsız ve metne bağımlı konuşma tanıma sistemlerinin gerçekleştirilmesinde sistemin oluşturulmasında kullanılan konuşmacı sayısı ile sistemin verimi doğru orantılıdır. Oluşturduğumuz konuşmacıdan bağımsız konuşma tanıma sisteminin veritabanı 10 erkek ve 10 bayan konuşmacının telaffuzları kullanılarak oluşturulmuş ve sistemin gerçek zamanlı laboratuar uygulaması MATLAB 6.5 sürümünde Data Acquisition Toolbox kullanılarak gerçekleştirilmiştir.

Anahtar Kelimeler: Konuşma Tanıma; Dinamik Zaman Eğreltimi; Doğrusal Öngörülü Kodlama (DÖK), Mel Frekans Ölçekli Kepstral Katsayılar

Real Time Laboratory Application of Speech Recognition System for Controlling Equipments in a Car by Voice

ABSTRACT

This paper has focused on real time laboratory application of speaker independent speech recognition system. The most important effect of altering speech recognition system's performance is being dependent speaker or independent speaker of system. For performing of speaker independent and text depended speech recognition systems, number of speaker used for training the system is directly proportional to the system's performance. Our performed speaker independent speech recognition system's database has been constituted by using 10 male's and 10 female's utterance and system's real time laboratory application has been performed by using Data Acquisition Toolbox on MATLAB version 6.5.

Keywords: Speech Recognition; Dynamic Time Warping (DTW), Linear Predictive Coding (LPC), Mel Frequency Cepstral Coefficient (MFCC)

1. GİRİŞ

Konuşma tanıma alanındaki çalışmalardan bazıları; Örün (web) sayfalarında dolaşımı sağlayabilecek, klavyeye dokunmadan e-posta veya metin yazmaya imkan sağlayacak, programları açma kapamayı sağlayabilecek, haber veya bilgi içerikli internet sayfalarının görme özürü insanlar tarafından takip edilebilmesine imkan sağlayabilecek uygulamaları içeren insan-bilgisayar iletişimi (1-3), konuşma tanıma sistemlerinin eğitim alanında kullanılması (4), robotların sesle kontrolü (5), çocukların dil gelişiminde otomatik dil tanıma sistemlerinin kullanılması (6) gibi uygulamalar ve bu çalışmamızda amaçlanan araç içerisindeki cihazların sesle kontrolü olarak gösterilebilir.

.....
Makale 23.07.2007 tarihinde gelmiş, 22.10.2007 tarihinde yayınlanmak üzere kabul edilmiştir.

E. KELEBEKLER, M.İNAL, Kocaeli Üniversitesi, Teknik Eğitim Fakültesi, Elektronik ve Bilgisayar Eğitimi Bölümü, Umuttepe Yerleşkesi, 41380, İzmit, Kocaeli
ersoy@kou.edu.tr, minat@kou.edu.tr
Digital Object Identifier 10.2339/2008.11.2.109-114

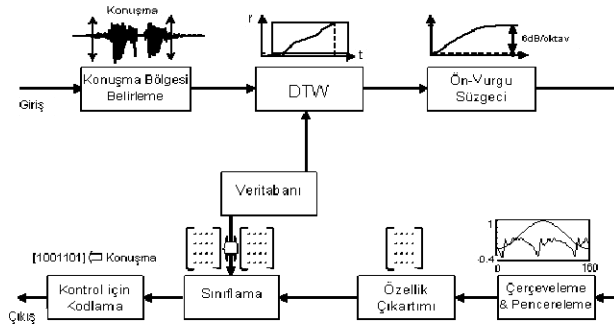
1980'lerde yüksek hızlı işaret işlemcilerinin üretilmesiyle birlikte konuşma tanıma sistemlerinin araç içerisinde kullanılmasına yönelik çalışmalar başlamıştır. Bu alanda ticari ve akademik olarak birçok çalışma yapılmaktadır. Mitsubishi bünyesinde, SpokenQuery (SQ) ses ile nesne seçme ve menü tabanlı arayüz (Menu-Based Interface) yaklaşımları araç içerisindeki müzik sisteminin kontrolü amacıyla karşılaştırılmıştır (7). Cavedon ve arkadaşları, araç içerisindeki ses sistemini kontrol etmek amacıyla Saklı Markov Model (HMM) kullanan istatistiksel doğal dil anlama modülünü geliştirmişlerdir (8-9). Diğer bir çalışmada, mel-frekans kepsral katsayıları (MFCC) kullanan ve doğrusal diskriminant analizi yapan bir sistem araç içerisindeki cihazların kontrolü amacıyla gerçekleştirilmiştir (10). Yapanel ve arkadaşları MFCC ve HMM kullanarak araç içi cihazları ses ile kontrol etmeyi amaçlamışlardır (11). Ayrıca Toyota ve Mercedes-Benz otomobil şirketleri bünyesinde çeşitli çalışmalar yapılmaktadır (12-13).

Günümüzde birçok araba üreticisi kendi ürettikleri araçlarda konuşma tanıma, sesle yönlendirme, komut ve kontrol sistemlerinin kullanılacağını ifade etmişlerdir. Fakat bu konu ile ilgili güvenilirlik ve maliyet üzerindeki teknolojik çalışmalar sonuçlandırılmamıştır ve hala devam etmektedir (13).

Bu alanda yapılan çalışmalar göz önünde bulundurularak, çalışmamızda araç içi cihazların gerçek zamanlı kontrolüne yönelik, gürültüye karşı güçlü, Türkçe dili için konuşmacıdan bağımsız bir konuşma tanıma sistemi oluşturulması amaçlanmıştır. Tanımayı zorlaştıran ve sistemin performansını düşüren gürültü faktörünün ortadan kaldırılması amacıyla farklı bir çalışmada, ses işareti üzerine bindirilmiş farklı İşaret Gürültü Oranlarında (Signal Noise Ratio-SNR) renkli ve beyaz gürültüler için çeşitli algoritmaların gürültü bastırma performansları incelenmiştir (14). Çalışmada; konuşma tanıma sistemi Bölüm 2’de verilmiştir. Bölüm 3’te ise sistemin laboratuvar ortamında gerçek zamanlı uygulaması anlatılarak elde edilen sonuçlar Bölüm 4’te yorumlanmıştır.

2. KONUŞMA TANIMA SİSTEMİ

Konuşma tanıma çalışmalarının temel amacını oluşturan insan-makine iletişiminin bir kolu olarak insan-arac iletişimi veya araç içerisindeki cihazların (pencere, klima ve müzik sistemi) sesle kontrolü gösterilebilir. Çalışmamızda araç içindeki cihazların ve bileşenlerin sesle kontrolünü sağlamak amaçlanmaktadır. Konuşma tanıma yöntemi olarak ayırık kelime tanıma yöntemi kullanılmıştır. Bu yöntemin başarılı bir şekilde çalışmasının temel koşulu konuşma bölgesinin çok iyi şekilde belirlenmesi ve dolayısıyla Dinamik Zaman Eğreltisi (Dynamic Time Warping-DTW) işleminin en iyi biçimde yapılmasıdır (15-17).



Şekil 1. Konuşma tanıma işleminin blok şeması

Hazırlanan sistemin blok diyagramı Şekil 1’de verilmiştir. Bir mikrofon aracılığıyla alınan konuşma işareti, Analog-Sayısal Çevirici (Analog Digital Converter-ADC) ile ayırık zamanlı işaret haline getirilmiş ve 8 KHz örnekleme frekansı ile örneklenerek Şekil 1’de gösterilen konuşma tanıma sisteminin girişine uygulanmıştır.

Şekil 1’deki her blok sırasıyla aşağıda açıklanan işlemleri gerçekleştirirler:

- **Konuşma Bölgesini Belirleme:** Girişten gelen konuşma işaretinin sınırlarını, işaretin enerji seviyesine ve sıfır geçiş sayısına bakarak belirler. Bu sınırlar dışında kalan sessiz bölge atılarak bloğun çıkışında konuşma işareti yalın olarak elde edilmiş olur.
- **Dinamik Zaman Eğreltisi:** Sessiz bölgeden ayrıştırılmış konuşma işareti veritabanındaki kelimeler ile iki işaretin zaman düzleminde boyutlarının eşitlenmesi işlemi olan dinamik zaman eğreltisi işlemi uygulanır.
- **Ön-Vurgu Süzgeci:** Konuşma üretme sisteminde sürtünme, ısı kayıpları ve dudaklardaki saçılma kaybı gibi nedenlerden bir miktar enerji tüketilir. Yüksek frekanslarda meydana gelen bu kayıpları karşılamak, işareti spektral anlamda pürüzsüzleştirmek ve daha sonra işaret işleme içindeki etkilere karşı hassasiyetini azaltmak için işaretin birinci dereceden bir FIR süzgeçten geçirilmesi işlemidir. Ön-Vurgu süzgeci olarak genellikle birinci dereceden bir süzgeç sistemi aşağıdaki denklemde görüldüğü gibi kullanılır (18)

$$H(z) = 1 - az^{-1}, \quad 0.9 \leq a \leq 1 \quad (1)$$

Burada a süzgeç katsayısı ve $H(z)$ ise süzgeç yantıdır. Bu durumda Ön-Vurgu süzgecinin çıkışı, giriş işaretine bağlıdır ve Eşitlik 2 kullanılarak elde edilir.

$$\hat{s}(n) = s(n) - as(n-1), \quad 0.9 \leq a \leq 1 \quad (2)$$

a değeri genellikle çalışmamızda da kullanıldığı gibi 0.95 civarında seçilir. $s(n)$, ayırık ses işaretleri dizisi olup, $\hat{s}(n)$ ise $s(n)$ dizisinin birinci dereceden bir süzgeçten geçirilmiş şeklidir.

• **Çerçeveleme ve Pencereleme:** Ön-Vurgu süzgeçinden geçirilen ve 20 ms ‘lik parçalara bölünerek çerçevesiz konuşma işareti Hanning pencere katsayıları ile çarpılarak pencerelenir. 20 ms ‘lik parçalara bölünerek pencerelenen konuşma işareti, pencere geçişlerinde meydana gelebilecek bilgi kayıplarını önlemek amacıyla her pencere bir diğerinin içine geçecek şekilde 10 ms ‘lik aralıklarla başlatılır.

• **Özellik Çıkartımı:** Sistemde kullanılacak özellik çıkartım yöntemini belirleme aşamasında kepsstral katsayıların elde edilmesi için üç farklı yöntem kullanılmıştır. Bu yöntemlerden ikisi Doğrusal Öngörülü Kodlama (DÖK) analizi yapan; Özilinti Analizi (Autocorrelation Analysis) ve Ortak Değişinti Metodu (Covariance Method), üçüncüsü ise Mel Frekans Ölçekli Kepsstral Katsayılar (MFCC) yöntemidir. Bu yöntemler içerisinde bu çalışmada amaçlanan gerçek zamanlı uygulama için en uygun yöntem belirlenmeye çalışılmıştır. DÖK analizi konuşma kodlama, konuşma işaretinden konuşmayı veya konuşmacıyı belirleyici parametrelerin çıkartılması gibi birçok alanda kullanılır (18). Ancak çok iyi bir

tanıma verimi sağlanamaz. DÖK analizi sonucu elde edilen değerler kepstral katsayılara dönüştürülerek tanıma verimleri artırılır ve konuşma-konuşmacı tanıma sistemlerinde kullanılabilir (19). MFCC yönteminin insan kulağının algılama özelliğini temel alarak çalışması, MFCC'nin konuşma tanıma sistemi için çok etkili bir yöntem olmasını sağlar. Yapılan deneysel çalışmalar konuşma ve konuşmacı tanıma sistemlerinde MFCC'nin çok başarılı olduğunu göstermiştir (20). Gerçek zamanlı uygulamalarda kullanılan sistemin verimi kadar işlemlerin gerçekleşme süreleri de önemlidir. İşlem süreleri açısından yöntemler, oluşturulan veritabanından 3 bayan ve 3 erkek, toplam 6, konuşmacının "pencere" telaffuzları alınarak her yöntemin aynı telaffuz için özellik matrislerini üretme süreleri incelenerek karşılaştırılmıştır. Özilinti (Autocorrelation) Analizi-kepstral katsayı dönüşümü, Ortak Değişinti (Covariance) Metodu-kepstral katsayı dönüşümü ve MFCC Yönteminin aynı telaffuzlar için işlem süreleri Tablo 1'de verilmiştir. Tablo 1 aynı telaffuz için yöntemlerin işlem süreleri incelendiğinde MFCC yönteminin en hızlı yöntem olduğu görülmektedir. Tüm bu karşılaştırmalar sonucunda, bu çalışmada amaçlanan gerçek zamanlı konuşma tanıma sisteminde özellik çıkartım yöntemi olarak MFCC yönteminin kullanılması kararlaştırılmıştır.

Tablo 1. Yöntemlerin işlem süreleri (Saniye)

Konuşmacı Etiketleri	Konuşma Verilerinin Uzunlukları	Özilinti-Kepstral Katsayılar	Ortak Değişinti-Kepstral Katsayılar	MFCC
F1	0,61	1,062	0,060	0,040
F2	0,65	1,101	0,070	0,060
F3	0,69	1,231	0,070	0,040
M1	0,59	1,112	0,070	0,060
M2	0,58	0,981	0,061	0,060
M3	0,56	1,011	0,060	0,060
Ortalama	0,61	1,083	0,065	0,053

- **Sınıflandırma:** Konuşma işareti özellik matrisi ile veritabanında tutulan konuşma işaretlerinin özellik matrislerinin benzerliği, Eşitlik 3'de denklemleri verilen Euclidean uzaklığı kullanılarak hesaplanmıştır:

$$d(r, t) = \sqrt{\sum_{k=1}^N |r_k - t_k|^2} = \sqrt{[r - t]^T [r - t]} \quad (3)$$

Burada r veritabanındaki referans özellik matrisini, t ise o anki karşılaştırılacak konuşma verisinin özellik matrisini ifade etmektedir.

- **Kontrol için Kodlama:** Önceden belirlenen eşik değere göre giriş konuşma işaretine veritabanındaki en yakın konuşma işareti belirlendikten sonra araç içi cihazların veya bileşenlerinin kontrolü için kullanılacak Şekil 6'da gösterilen kodlar üretilir.

DÖK modeli, n . ayrık zamanda verilmiş $s(n)$ konuşma örneği, önceki p konuşma örneğinin yaklaşık olarak doğrusal bir birleşiminden, a_p DÖK katsayıları ile aşağıdaki eşitlikte gösterildiği gibi elde edilebilir:

$$s(n) \cong a_1 s(n-1) + a_2 s(n-2) + a_3 s(n-3) + \dots + a_p s(n-p) \quad (4)$$

$a_1, a_2, a_3, \dots, a_p$ katsayıları ilgili konuşma penceresi için sabit kabul edilir ve Doğrusal Öngörü katsayıları olarak adlandırılırlar.

Konuşma işareti $s(n)$ 'i üretmek için $u(n)$ uyarım kaynağı, G kazancı ile çarpılır ve tüm-kutup sisteme ($H(z)=1/A(z)$), giriş olarak verilir. Eşitlik 4'e uyarım terimi, $Gu(n)$, eklenerek eşitlik aşağıdaki şekle dönüştürülebilir:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (5)$$

Eşitliği z boyutunda yazarsak aşağıdaki eşitlik elde edilir:

$$S(z) = \sum_{k=1}^p a_k z^{-k} S(z) + GU(z) \quad (6)$$

Doğrusal öngörü analizinde temel problem doğrudan konuşma işaretinden öngörü katsayı parametrelerini, $\{a_1, a_2, \dots, a_p\}$, hesaplamaktır.

3. LABORATUAR ORTAMINDA GERÇEK ZAMANLI KONUSMA TANIMA UYGULAMASI

Şekil 1'de gösterilen sistemin gerçek zamanlı uygulaması işlemine, sistemde kullanılacak kelimelerin belirlenmesiyle başlamıştır. Konuşma tanıma işleminde araç içi cihazların ve bileşenlerinin kontrolüne uygun kelimeler seçilmiştir. Çalışmamızda araç bileşenlerinden müzik sisteminin, klimanın ve pencerenin sesle kontrolü amaçlanmaktadır. Komut listesi Şekil 2'de verilen ağaç yapısı şeklindedir. Sistem, ses komutlarına göre içeriği değiştirilecek bir kontrol kod kümesine sahiptir. Bu kod kümesinin içeriği başlangıç değeri olarak altı adet -1'den oluşmaktadır. Ağaç yapısı üç bloktan oluşmaktadır ve her blok kontrol kod kümesinin iki değerini değiştirmektedir. Sistem başlangıç aşamasında kullanıcının ağaç yapısının ilk bloğunu oluşturan "yukarı", "aşağı", "değiştir" kelimelerinden birini beklemektedir. Eğer bu aşamada söylenen telaffuzun özellik matrisi bu kelimelerden herhangi birinin özellik matrisi için belirlenen eşik değerinden daha yakınsa, kontrol kod kümesinin içeriği Şekil 2'de ilgili kelimelerin altında verilen kodlar olarak değiştirilecektir. İlk telaffuz beklenen kelimelerden biri ise ağaç yapısı ilgili kelimenin dallarında gösterilen kelimelerden birini bekleyecektir. Örneğin "yukarı" kelimesinden sonra kontrol kod kümesinin içeriği '00-1-1-1-1' şeklinde değiştirilecek ve ikinci telaffuz olarak "müzik", "pencere", "klima" kelimelerinden biri beklenecektir. Eğer ikinci telaffuz bu kelimelerden herhangi biri ile eşleştirilirse kontrol kod içeriğinin üç ve dördüncü değerleri kelimeye karşılık gelen kodlar olarak değiştirilecektir. İkinci telaffuz bu

kelimelerden hiç biri ile eşleştirilemez ise kontrol kod kümesinin içeriği başlangıç değeri olan ‘-1-1-1-1-1’ şekline döner. Şekil 2’de görüldüğü gibi üçüncü dal bloğuna sahip iki komut vardır. Bu komutlar birden fazla birleşene sahip “pencere” ve “müzik sistemi” komutlarıdır. Üçüncü blok için de ilk iki bloktaki çalışma yapısı geçerlidir. Sistem ağaç yapısının her aşamasında söylenen telaffuz ile “durdur” kelimesinin telaffuzunu kontrol etmektedir. Eğer telaffuz edilen kelime “durdur”

Bu amaçla yapılan verim testlerinin sonuçları Tablo 3’te verilmiştir. Tablo 3 incelendiğinde, doğal olarak en iyi sonuçlar, bayan konuşmacılar için bayan konuşmaların telaffuzlarından elde edilen özellik matrisleri kullanıldığında ve erkek konuşmacılar için ise erkek konuşmaların telaffuzlarından elde edilen özellik matrisleri kullanıldığında elde edilmiştir. Aksine, en kötü sonuçlar bayan konuşmacılar için erkek konuşmaların telaffuzlarından elde edilen özellik matrisleri kul-

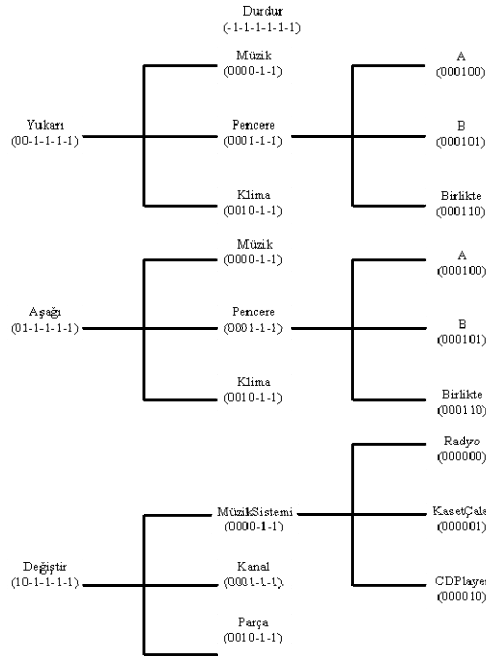
Tablo 2. Sistemde kullanılan komutların listesi ve komutların işlevleri

Komut Listesi	Kontrol Kod Kümesi	Komuta Karşılık Gelen Görev
Durdur	-1-1-1-1-1-1	Sistemi başlangıç durumuna getirir
Yukarı → müzik	0000-1-1	Müzik sesini artırır
Yukarı → pencere → A	000100	Sol pencereyi kapatır
Yukarı → pencere → B	000101	Sağ pencereyi kapatır
Yukarı → pencere → birlikte	000110	İki pencereyi birlikte kapatır
Yukarı → klima	0010-1-1	Klimanın derecesini artırır
Aşağı → müzik	0100-1-1	Müzik sesini azaltır
Aşağı → pencere → A	010100	Sol pencereyi açar
Aşağı → pencere → B	010101	Sağ pencereyi açar
Aşağı → pencere → birlikte	010110	İki pencereyi birlikte açar
Aşağı → klima	0110-1-1	Klimanın derecesini azaltır
Değiştir → müzik sistemi → radyo	100000	Müzik sisteminde radyoyu çalar
Değiştir → müzik sistemi → kasetçalar	100001	Müzik sisteminde kaseti çalar
Değiştir → müzik sistemi → cdplayer	100010	Müzik sisteminde CD’yi çalar
Değiştir → kanal	1001-1-1	Radyo kanalı değiştirilir
Değiştir → parça	1010-1-1	Çalan parça değiştirilir (kaset veya CD)

kelimesi ile eşleştirilirse kontrol kod kümesi içeriği ‘-1-1-1-1-1-1’ olarak değiştirilir ve sistem başlangıç durumuna getirilmiş olur. Böylelikle herhangi bir aşamada sistemin yapabileceği yanlış önlenemez veya telaffuz edilecek yanlış bir komuttan geri dönüş sağlanabilir. Komutlar, yapacağı işler ve bu işlere karşılık gelen kontrol kod kümesi içerikleri Tablo 2’de verilmiştir. Pencere kontrolünü sağlayan komutlarda “sol” ve “sağ” kelimelerinin telaffuzları dolayısıyla özellik matrisleri birbirine çok yakın olduğu için “sol” kelimesinin yerine “A”, “sağ” kelimesinin yerine de “B” seslerinin telaffuz edilmesi uygun görülmüştür. Sisteme Tablo 2’nin son satırındaki komutlar; sırasıyla “değiştir” ve “parça” telaffuz edildiğinde, sistemde o anda kasetçalar veya CD oynatıcıdan hangisi aktif ise çalmakta olan parçayı değiştirerek sonraki parçaya geçilir. Tablo 2’deki komutlar, 10 bayan ve 10 bay konuşmacıya beşer kez tekrar ettirilerek; konuşma beyirlerinin üç tanesi referans özellik matrislerini içeren veritabanının oluşturulmasında, diğer ikisi ise sistemin verimini test etmek amacıyla kullanılmıştır.

lanıldığında ve erkek konuşmacılar için ise bayan konuşmaların telaffuzlarından elde edilen özellik matrisleri kullanıldığında elde edilmiştir.

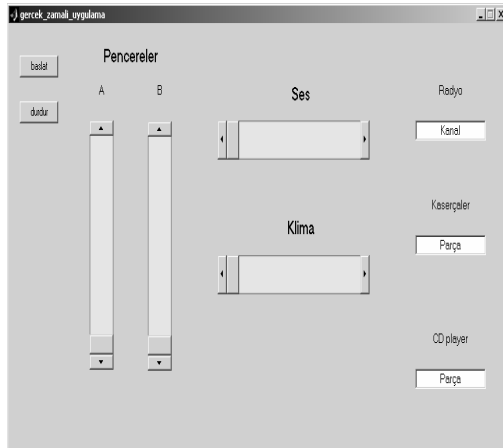
Sistemin gerçek zamanlı laboratuvar uygulaması Matlab 6.5 sürümünde aşağıda özellikleri verilen Data Acquisition Toolbox (21) kullanılarak gerçekleştirilmiştir: Analog giriş: Winsound, Kanal sayısı: 1, Örneklem oranı: 8000 Hz, Tetikleme tekrar sayısı: 0, Tetikleme tipi: Manual, Zamanlayıcı fonksiyon: Veri_al, Tetiklemede alınan örnek sayısı: 4000, Zamanlayıcı fonksiyon periyodu: 0.5 saniye, Tampon konfigürasyonu: max. Çalışmamızda kullanılan Data Acquisition Toolbox, ses kartından veri okuma amacıyla kullanılmıştır. Analog giriş “winsound” seçildiğinde Matlab bir saniyede örneklem oranı kadar veriyi ses kartından okur. Oluşturduğumuz zamanlayıcı fonksiyon, “zamanlayıcı fonksiyon periyodu” kadar sıklıklarla çalıştırılır. Oluşturduğumuz veri edinim yapısıyla ses kartından her 0.5 s.’de bir 4000 veri, veri_al fonksiyonu aracılığı ile okunmaktadır. Gerçek zamanlı uygulamanın çalışmasını görsel olarak izleyebilmek amacıyla Şekil 3’te verilmiş olan GUI penceresi hazırlanmıştır.



Şekil 2. Komut listesinin ağaç yapısı

Tablo 3. Sistem verimi sonuçları (%)

Konuşmacı Etiketleri	Erkek Konuşmacılar Ortalama Özellik Vektörleri	Bayan Konuşmacılar Ortalama Özellik Vektörleri	Tüm Konuşmacılar için Ortak Ortalama Özellik Vektörleri
M1	100,00	16,67	83,33
M2	96,67	53,33	100,00
M3	100,00	33,33	53,33
M4	100,00	73,30	100,00
M5	100,00	33,33	70,00
M6	100,00	90,00	100,00
M7	73,33	46,67	60,00
M8	100,00	60,00	100,00
M9	93,33	100,00	86,67
M10	93,33	66,67	100,00
F1	86,67	100,00	86,67
F2	53,33	93,33	86,67
F3	46,67	100,00	100,00
F4	50,00	100,00	73,33
F5	46,67	100,00	93,33
F6	83,33	100,00	100,00
F7	56,67	100,00	100,00
F8	53,33	100,00	100,00
F9	30,00	93,33	73,33
F10	46,67	100,00	100,00
Erkek Konuşmacıların Ortalamaları	95,67	57,33	85,33
Bayan Konuşmacıların Ortalamaları	55,33	98,67	91,33



Şekil 3. Gerçek zamanlı GUI penceresi

4. SONUÇ VE ÖNERİLER

Çalışmamızda amaçladığımız araç içindeki cihazların sesle kontrolüne yönelik konuşma tanıma sisteminin laboratuvar ortamında gerçek zamanlı uygulaması işlemi için bir sistem oluşturulmuştur. Bu sistemde kullanılan yöntemlerin seçiminde konuşma tanıma işlemini gerçekleştirme verimleri kadar işlem hızları da dikkate alınmıştır. Bu amaçla yapılan çalışmaların sonucunda gerçek zamanlı uygulamalarda; gürültü bastırma işleminde RLS algoritmasının (14), özellik çıkartım yöntemi olarak da MFCC'nin kullanılması kararlaştırılmıştır. Türkçe dili için böyle bir çalışmanın yapılması gelecekte yapılacak çalışmalara örnek olacağı düşünülmektedir. Gerçekleştirilen konuşmacıdan bağımsız konuşma tanıma sisteminden elde edilen sonuçlar göz önünde bulundurulduğunda; gelecekteki çalışmalarda, erkek ya da bayan sesleri ile oluşturulmuş iki ayrı veritabanına sahip bir sistemin, sürücünün bayan veya erkek olmasına göre kullanacağı veritabanını otomatik olarak belirleyen ve böylelikle daha iyi konuşma tanıma verimi sağlayan sistemler gerçekleştirilebilir. Ayrıca çalışmamızda özellik vektörlerinin birbirinden uzaklığı Eşitlik 3'te gösterilen Euclidean denklemiyle hesaplanmıştır. Bunun seçilme nedeni sistemin gerçek zamanlı uygulanabilirliğini arttırmaktır. Bunun yerine örüntü sınıflandırmada kullanılan Bayesian, Yapay sinir Ağları (YSA), karar ağaç yapıları ve K-means algoritmaları gibi yöntemlerin kullanılması önerilmektedir. Yaygın olmakla birlikte bedensel engelli insanlar için tasarlanan araçlarda konuşma işareti ile araç kontrolü gerçekleştirilmektedir. Bu tür uygulamalarda sürücünün özelliğine göre çeşitli değişiklikler yapılmakta, dolayısıyla konuşma tanıma sistemi konuşmacıya bağımlı bir şekilde tasarlanmaktadır. Bu çalışmada ise konuşmacıdan bağımsız konforlu ve güvenli bir sürüş sağlayacak özelliklerde araç içi konuşma tanıma sisteminin laboratuvar ortamında gerçek zamanlı uygulaması yapılmıştır. Otomobil üreticilerinin maliyet ve güvenilirlik endişeleri giderilirse, önümüzdeki yıllarda bu tür sistemlerin ticari amaçlı yaygınlaşmasının mümkün olacağı düşünülmektedir.

5. KAYNAKLAR

1. Furui, S., "Speech Recognition Technology in the Ubiquitous/Wearable Computing Environment", Acoustics-Speech and Signal Processing, ICASSP '00. Proceedings. 2000 IEEE International Conference, 6, 3735-3738, 2000.
2. Thong, J-M, vd., "SpeechBot: a Speech Recognition based Audio Indexing System for the Web", International Conference on Computer-Assisted Information Retrieval, RIAO, 106-115, 2000.
3. Davis, C.D., "Automatic Speech Recognition and Access: 20 years, 20 Months, or Tomorrow?", Issue of Hearing Loss: The Journal of Self Help for Hard of Hearing People, 22, No. 4, 11-14, 2001.
4. Wald, M., "Using Automatic Speech Recognition to Enhance Education for All Students: Turning a Vision into Reality", 34th ASEE/IEEE Frontiers in Education Conference, Savannah, GA, 2004.

5. Li, M., and Okamura, A.M., "Recognition of Operator Motions for Real Time Assistance using Virtual Fixtures", Proc. of the 11th Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, (HAPTICS'03), 2003.
6. Russell, M., vd., "Applications of Automatic Speech Recognition of Speech and Language Development in Young Children", British Crown Copyright/DERA, 1996.
7. Forlines, C., Nielsen, B.S., Raj, B., Wittenburg, K., Wolf, P., "A Comparison Between Spoken Queries and Menu-based Interfaces for in car Digital Music Selection", Mitsubishi Electric Research Laboratories Inc., Cambridge, Massachusetts, 2005.
8. Cavedon, L., vd., "Developing a Conversational In-Car Dialog System", The 12th International Congress on Intelligent Transportation Systems, San Francisco CA, USA, 2004.
9. Weng, F., vd., "A Flexible Conversational Dialog System for MP3 Player" Proceedings of HLT/EMNLP 2005 Demonstration Abstracts, 24-25, 2005.
10. Westphal, M., and Waibel, A., "Towards Spontaneous Speech Recognition for On-Board Car Navigation and Information Systems", Proceedings of the Eurospeech, 99, 1955-1958, 1999.
11. Yapanel, U., Zhang, X., Hansen, J.H.L., "High Performance Digit Recognition in Real Car Environments", ICSLP-2002: Inter. Conf. on Spoken Language Processing, 2, 793-796, 2002.
12. Watika, T., "Speech Based Interfaces in Vehicles", R&D Review of Toyota CRDL, 39, No. 1, 2004.
13. Heisterkamp, P. "Linguatronic Product-Level Speech System for Mercedes-Benz Cars", DaimlerChrysler AG Research and Technology Speech Understanding, Germany, 2001.
14. Kelebekler, E., İnal, M., "White and Color Noise Cancellation of Speech Signal by Adaptive Filtering and Soft Computing Algorithms", Springer, Lecture Notes in Computer Science, AI 2006: Advances in Artificial Intelligence, 4304, 970-975, 2006.
15. Rabiner, L.R., Rosenberg, A.E., Levinson, S.E., "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition", IEEE Transaction on Acoustic, Speech, and Signal Processing, 26, No 6, 1978.
16. Myers, C.S., Rabiner, L.R., Rosenberg, A.E., "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition", IEEE Transaction on Acoustic, Speech, and Signal Processing, 28, No 6, 1980.
17. Furui, S., "Speaker-independent isolated word recognition using dynamic features of speech spectrum", IEEE Transaction on Acoustic, Speech, and Signal Processing, 34, No. 1, 1986.
18. Rabiner, L.R., Juang, B.H., "Fundamentals of Speech Recognition", PTR Prentice-Hall Inc., Englewood Cliffs NJ, 1993.
19. İnal, M., vd., "Comparison of Linear Predictive Analysis Methods for ANN-Based Speaker Identification", Neurel-2000, 5. Seminar on Neural Network Application in Electrical Engineering, Faculty of Elec. Eng. Univ. of Belgrade, YU, 109-112, 2000.
20. Koç, A., "Acoustic Feature Analysis for Robust Speech Recognition", Yüksek Lisans Tezi, Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü, 2002.
21. "Data Acquisition Toolbox User's Guide", The Mathworks Inc. 2002.