

Ses Komut Tanıma ile Gezgin Araç Kontrolü

*¹Muharrem ÇELEBİ, ²Ali BULDU

¹ Elektronik ve Haberleşme Mühendisliği, Fen Bilimleri Enstitüsü, Kocaeli Üniversitesi, Türkiye
²Bilgisayar Mühendisliği, Teknoloji Fakültesi, Marmara Üniversitesi, Türkiye

Özet :

Bu çalışmada, ses komut tanıma ile gezgin araç kontrolü gerçekleştirmek istenmiştir. Bu sayede, kullanıcılara herhangi bir cihazı kontrol etmek için kullanım kolaylığı, hareket serbestliği, uzaktan veri girişi imkanı sağlamaktadır. MATLAB paket programı kullanılarak ses komutları bilgisayar ortamına kaydedilmiştir. Ses komut tanıma sisteminde yalıtık kelime tanıma sistemi tercih edilmiştir. Öznitelik çıkarma yöntemi olarak Mel frekans kepstrum katsayıları kullanılmış, öznitelik eşleştirme yöntemi olarak ise vektör nicemleme yöntemi kullanılmıştır. Gezgin araç kontrolü için “ileri git”, “geri git”, “sağa dön”, “sola dön” ve “dur” komutları kullanılmaktadır. Bilgisayar ortamında ses komutlarını tanıma işlemi yapıldıktan sonra radyo frekansları ile komut bilgisi gezgin araca iletilmektedir. Gürültüsüz ve gürültülü olmak üzere iki ayrı ortamda yapılan çalışmalarda sırasıyla, %96, %68 başarı oranları elde edilmiştir.

Anahtar Kelimeler : Ses komut tanıma, öznitelik çıkarma, öznitelik eşleştirme, gezgin araç.

Mobile Vehicle Control With Voice Command Recognition

Abstract :

In this study, it is aimed to perform voice command recognition and mobile vehicle control. Thus provides ease of use to control any device, mobility and remote data entry to users. Voice commands are recorded to computer by using MATLAB. The isolated word recognition system is preferred in voice command recognition system. Mel Cepstrum coefficients have been used as the method of feature extraction. In addition, feature matching method has been used as the method of the vector quantization. The commands, “go forward”, “go back”, “turn right”, “turn left” and “stop” are used to control the mobile vehicle. After having the recognition process by using computer voice commands, radio frequencies have been transmitted to the mobile vehicle. In studies which have been made in two separate environment as noiseless and noisy, success rates are determined as %96, %68 respectively.

Keywords: Voice command recognition, feature extraction, feature matching, mobile vehicle.

1. Giriş

Uzun yıllardır kullanılan parmak izi ve retina gibi kişiye has, kişinin kimliğini tanımlayıcı biyometrik özelliklere son yıllarda ses de eklenmiştir [1]. Ses tanıma sistemlerinin, konuşmacı tanıma, konuşma tanıma, konuşma sentezleme, komut – kontrol şeklinde uygulamaları mevcuttur [2]. Sesli komut – kontrol uygulamaları ile insan ve makine arasında doğal bir arabirim elde edilmiş olur. Bu uygulamalara örnek olarak ses ile kontrol edilebilen motorlu tekerlekli sandalye, robotların sesle kontrolü, araç içerisindeki cihazların sesle kontrolü, bilgisayarın ses ile kontrolü şeklinde verilebilir [3].

*Corresponding author: Address: Faculty of Engineering, Department of Electronic and Communication Engineering Kocaeli University, 41380, Kocaeli TURKEY. E-mail address: mhrmclb@gmail.com, Phone: 0 507 705 56 57

1.1 Konuşma Tanıma Sistemlerinin Sınıflandırılması

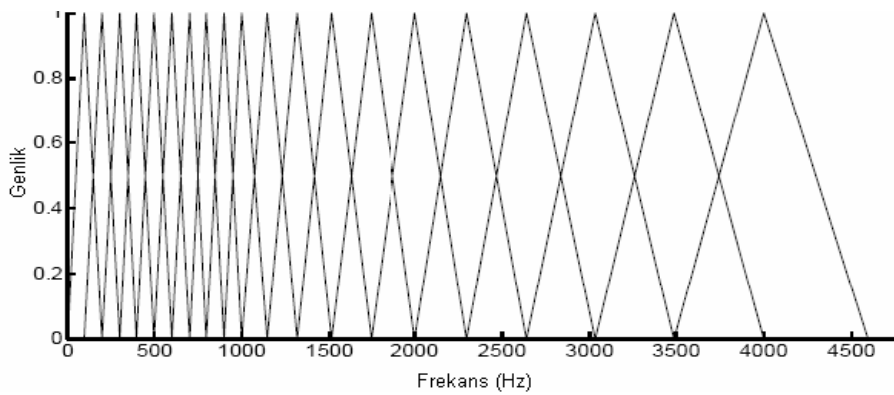
Konuşma yöntemine göre, yalıtık kelime tanıma, bağlantılı kelime tanıma ve sürekli konuşma tanıma olmak üzere üçe ayrılır. Yalıtık kelime tanıma, konuşmacının bireysel kelimelerinden, özel kelimelerin çıkarılması işlemidir. Bağlantılı kelime tanıma, konuşmacının akıcı konuşması esnasında özel kelimelerin çıkarılması işlemidir. Sürekli konuşma tanıma, konuşmacının akıcı bir şekilde konuşabildiği bir sistemdir [4]. Konuşmacı durumuna göre, konuşmacı bağımlı sistemler ve konuşmacı bağımsız sistemler olmak üzere ikiye ayrılır. Konuşmacı bağımlı sistemler, sadece bir konuşmacı için özel olarak tasarlanmış olan sistemlerdir. Konuşmacı bağımsız sistemler birden çok konuşmacı tarafından kullanılabilen sistemlerdir [4]. Metin yapısına göre, metin bağımlı sistemler ve metin bağımsız sistemler olmak üzere ikiye ayrılır. Metin bağımlı sistemlerde konuşmacının önceden belirlenmiş bir kelime veya cümleyi telaffuz etmesi istenir. Metin bağımsız sistemlerde konuşmacının söylemesi istenen kelime veya cümlede sınırlandırma yoktur [5].

2. Konuşma Bölgesini Belirleme

Girişten gelen konuşma işaretinin başlangıç ve bitişinin belirlenmesi için işaretin enerji seviyesine ve sıfır geçiş sayısına bakılır [6].

2.1. Mel Frekansı Kepstrum Katsayıları

Mel frekansı, iç kulak içinde gerçekleşen frekans analizine dayanmaktadır. İnsanın duyma algısı 1 kHz 'e kadar doğrusal, 1 kHz'nin üzerinde ise logaritmiktir [7]. Bu model Davis ve Mermelstein tarafından formüle edilmiştir [8].

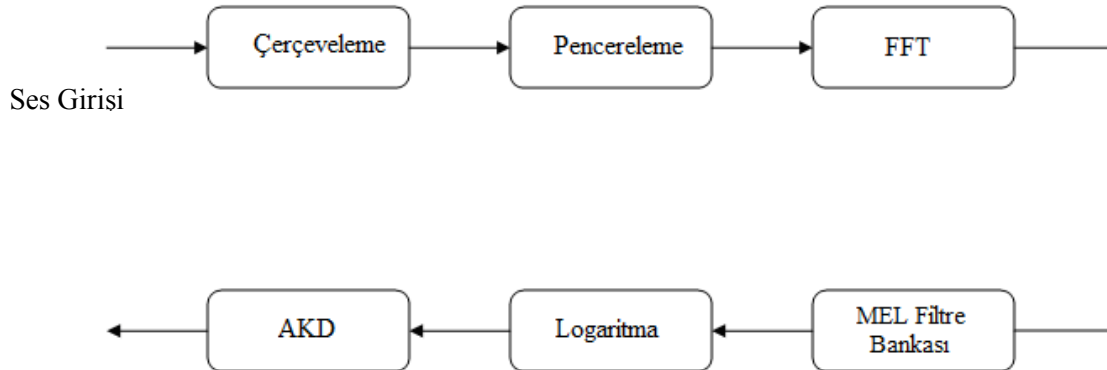


Şekil 1. Mel Filtre Bankası

Mel ölçeği için eşit aralıklı bir filtre bankası kullanılmaktadır. Bu filtre bankası, üçgen bant geçiren filtre cevabı olan ve bant genişliği sabit olan filtreler ile sağlanır [40]. Bu işlem ile lineer frekans ölçeği, mel ölçeğine dönüştürülmüş olur. Mel frekansı, denklem (1)'deki formül ile hesaplanır. Gerçek frekansın birimi hertz, Mel frekansının birimi ise Mel'dir [9].

$$mel(f) = 2595(\log(1 + f / 700)) \quad (1)$$

Mel frekans keppstrum katsayıları (MFKK) işleminin aşamaları Şekil 2'de gösterilmektedir. MFKK elde edilmesinde ilk olarak, sürekli konuşma işareti, N örnekten oluşan çerçevelere ayrılıp takip eden çerçeve M örnekten itibaren alınır ($M < N$). Her bir çerçevenin başından sonuna kadar işaret süresizlikleri minimuma indirmek için her bir çerçeve pencereleme işlemine tabi tutulur. Hızlı fourier dönüşümü ile N örnekten oluşan zaman alanındaki her bir çerçeve frekans alanına çevrilir. Bu işaret, Mel frekans ölçeğine göre dizilmiş süzgeç dizilerinden geçirilip logaritması alınır. Son olarak, logaritmik mel spektrumundan ayırık kosinüs dönüşümü kullanılarak zaman alanına geri dönülür. Sonuç olarak elde edilen katsayılara mel frekans keppstrum katsayıları denir.

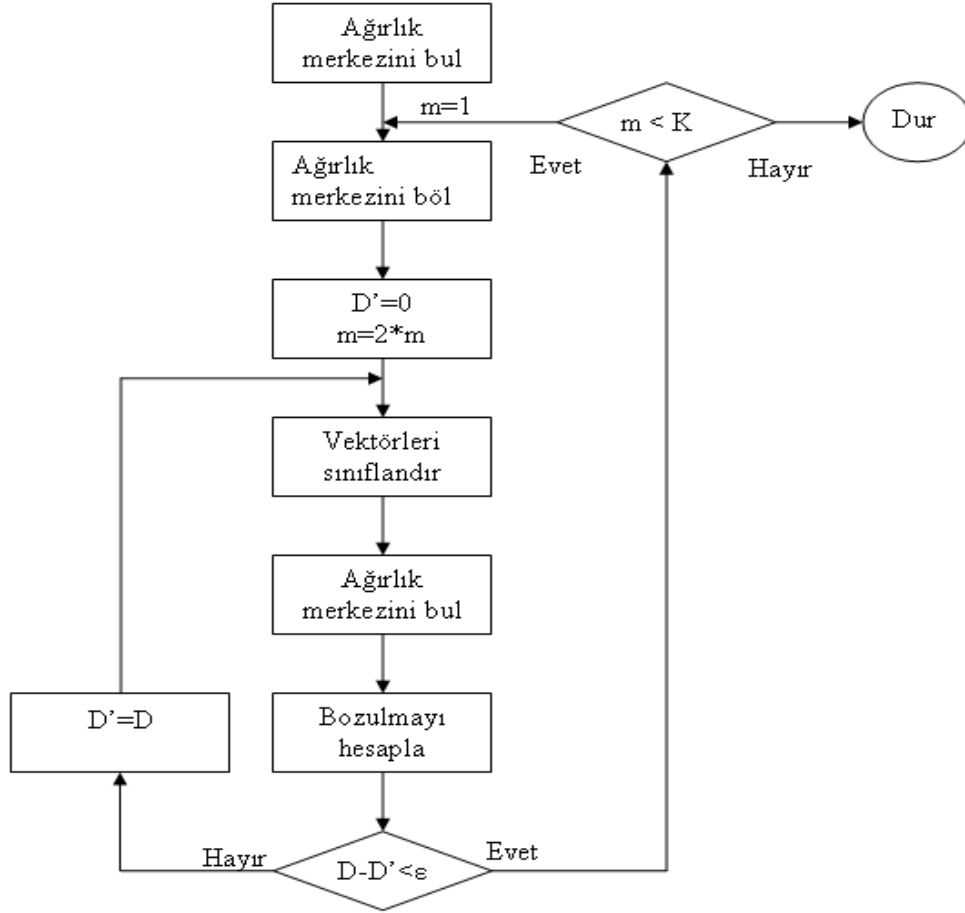


Şekil 2. MKKK Aşamaları

2.2. Vektör Nicemleme

Vektör nicemleme, verideki bilgi miktarını azaltmaya yarayan bir tür kümelendirme yöntemidir. Kümelendirme işlemine, günlük hayatın her alanında karşılaşmaktayız. Örneğin yiyecek satılan marketlerde aynı gruba ait et veya meyveler aynı yerlerde olacak şekilde yerleştirilir. Vektör nicemleme işleminde verimli sonuçlar veren LBG algoritması (Linde, Buzo and Gray Algoritma) kullanılmıştır [10, 11].

Konuşma işareti, 1 boyutlu vektör kod kitabı olmasına rağmen kod kelimeleri bölündükten sonra 2 boyutlu kod kitabı oluşturulur. L boyutlu vektör bilgisi K boyutlu kod kitabı değerine kadar bölünür. LBG algoritması Şekil 3’de gösterilmiştir [10, 11].



Şekil 3. LBG Algoritmasının Akış Diyagramı

L boyutlu eğitim vektörünün, K boyutlu kod kitabı haline dönüştürülmesi için işlem basamakları aşağıda açıklanmıştır :

1. Bir boyutlu vektör kod kitabı oluştur. Bu eğitim vektörlerin tamamının ağırlık merkezi hesaplanır.
2. Her bir kod kitabı aşağıdaki kurala göre ikiye ayrılır. Her bir eğitim vektörü, öklid uzaklığı kullanılarak mevcut kod kitabı içindeki en yakın olduğu kod vektörü bulunup onun etrafında kümelendirilir.

$$\begin{aligned} y_n^+ &= y_n(1 + \varepsilon) \\ y_n^- &= y_n(1 - \varepsilon) \end{aligned} \quad (2)$$

n : 1'den K 'ya kadar

ε : bölme parametresi olarak adlandırılır, 0,01 – 0,05 arasında değer alır.

3. Yeni ağırlık merkezi güncellemesi yapılır. Atanan yeni eğitim vektörleri göz önüne alınarak ağırlık merkezi güncellenir.
4. İstenen sayıda (K) kod vektörü elde edilinceye kadar ve ağırlık merkezi değişmeyinceye kadar 2. ve 3. adımlar tekrar edilir [10].

2.3. Karar Verme Aşaması

Eğitim aşamasında her bir konuşma örneği için kod kelimeleri elde edilir. Test aşamasında ise girişe uygulan konuşma bilgisinin öznitelik değerleri ile kod kelimeleri karşılaştırılır. Bu karşılaştırma ile uzaklık değeri hesaplanır ve en düşük uzaklık olan konuşma bilgisi seçilir. İki vektör arasındaki mesafeyi ölçmek için Öklid Uzaklık ölçümü kullanılır. Öklid uzaklık ölçümü en çok en çok kullanılan yöntemdir. Dik koordinat sisteminde iki nokta arasındaki geometrik uzaklığın hesaplanmasında kullanılır. Öklid uzaklık ölçümü aşağıdaki formül ile tanımlanır :

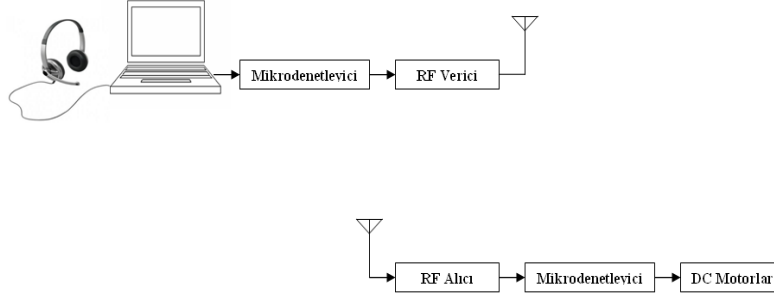
$$d_{\text{ö}}(x, y) = \sum_{i=1}^K \sqrt{(x_i - y_i)^2} \quad (3)$$

K : Vektör uzunluğunu, $d_{\text{ö}}$: Öklid uzaklık ölçüm değerini, x : kod kitabını, y : test işaretinin öznitelik verisini ifade eder.

3. Ses Komut Tanıma ile Gezgin Araç Kontrolü

Uygulamanın gerçekleştirilmesi için MATLAB 7.10 yazılımı kullanılmıştır. MATLAB, temel olarak sayısal hesaplama, grafiksel veri gösterimi ve programlamayı içeren teknik ve bilimsel

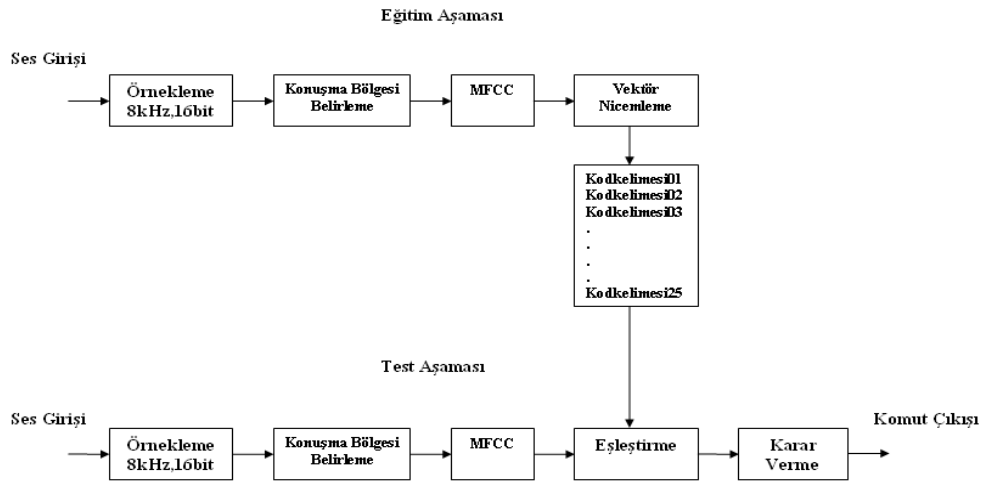
hesaplamalar için kullanılan bir yazılımdır. Bu aracın kullanılmasının sebebi, vektör ve matrislerle çalışmak için oldukça uygun bir programlama ortamı sunmasıdır. Ayrıca vektörler ve matrislerin matematiksel işlemleri birer komutla yapılabilmektedir. Bunun yanında kullanıcılar için arayüz tasarımı olanağı sunmaktadır. Gerçekleştirilen çalışmayı iki kısma ayırabiliriz. Birinci kısım ses komut tanıma, ikincisi ise gezgin araç kontrolüdür. Ses komut tanıma kısmı bilgisayar ortamında gerçekleştirilmektedir. İkinci kısım ise mikrodenetleyiciler, RF alıcı-verici, DC motorlardan oluşmaktadır.



Şekil 4. Gerçekleştirilen sistemin blok diyagramı

3.1. Ses Komut Tanıma Sistemi

Ses komut tanıma kısmında öznitelik çıkarma, öznitelik eşleştirme işlemi ve komutun gönderilmesi işlemleri yapılmaktadır. Ses komut tanıma sistemi kendi içinde eğitim aşaması ve test aşaması olarak iki kısma ayrılır. Eğitim aşamasında kullanıcının kod kitabı tasarımı gerçekleştirilir. Test aşamasında ise girişe uygulanan ses işareti ile kod kitabı karşılaştırılır. Sonuç değeri gezgin araca iletilir.



Şekil 5. Ses komut tanıma sistemi

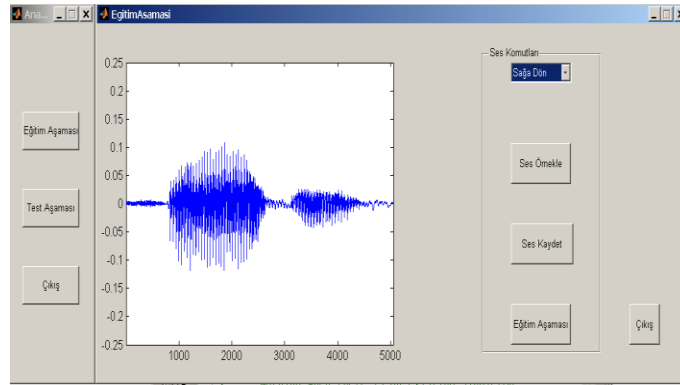
Eğitim aşamasında; ses sinyali örneklenir. Bu örneklerden konuşma bölgesi belirlenir. Konuşma bölgesi çıkarılmış işarete sırasıyla çerçeveleme, pencereleme, FFT, logaritma ve Ayrık Kosinüs Dönüşümü (AKD) işlemleri uygulanır. Elde edilen MFKK değerlerine vektör nicemleme işlemi uygulanır ve her konuşma işareti için 25 adet vektör elde edilir. Bu değerlerin ortalamaları alınarak kod kitabı oluşturulmuş olur.

Test aşamasında, mikrofon aracılığıyla ses işareti bilgisayar ortamında alınmakta, konuşma bölgesi belirlenmektedir. Konuşma işaretinin öznitelik verileri elde edilmektedir. Eşleştirme kısmında ise kod kitabı verileri ile test aşamasında girişe uygulanan işaretinin uzaklık ölçümleri yapılarak en yakın uzaklık belirlenmektedir. Belirlenen sesli komuta karşılık, gezgin araç hareket bilgisi seri port üzerinden 9600pbs hızında, 8,N,1 biçiminde gönderilir.

3.2. Gezgin Araç Kontrolü

Ses komut tanıma ile gezgin araç kontrolünde kişisel bilgisayar, RS232 dönüştürücü devresi, RF alıcı verici modülleri, gezgin araç kullanılmıştır. Kişisel bilgisayar üzerinde bulunan MATLAB ortamında koşturulan yazılım ile ses komutları tanınmaktadır. Komut tanıma ile ilgili hareket bilgisi önce seriport ile mikrodenetleyiciye, oradan RF verici ile gezgin araca iletilmektedir. RF alıcı ile hareket bilgisi çözümlenmekte ve araç üzerindeki motora gerekli uyarım işaretleri uygulanmaktadır.

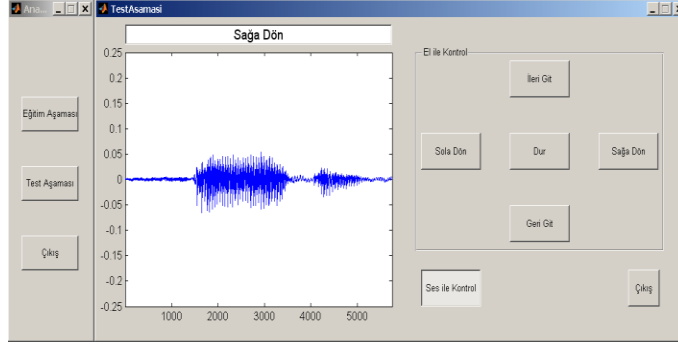
Bilgisayar ortamında hazırlanan yazılımın, Ana Menüde Eğitim Aşaması, Test Aşaması, Çıkış fonksiyonları bulunur.



Şekil 6. Sağa Dön komutunun kayıt işlemi

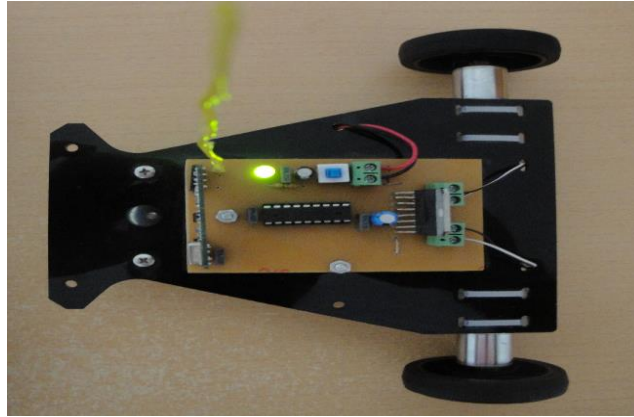
Eğitim aşaması butonu ile ses girişlerinin örneklenmesi, ses işaretinin kaydı, eğitim için gerekli olan kod kitabı oluşturma işlemleri gerçekleştirilir. Açılır menü seçeneğinden kayıt edilmek

istenen komut seçilmektedir. Ses Örnekle komutu ile ses kaydı yapılmakta ve kaydedilen ses tekrar konuşmacıya dinletilir. Örneklenen işaret uygun bulunursa Ses Kaydet butonu ile bilgisayara kayıt işlemi gerçekleştirilir. 25 adet ses kaydı bittikten sonra Eğitim Aşaması butonu ile kod kitabı oluşturulur.



Şekil 7. Ses ile kontrol işleminde ses komut örneği

Şekil 7’de girişe uygulanan ses işaretinden, konuşma bölgesi alınmış, grafik ekranda çizdirilmiştir. Öznitelik çıkarma ve eşleştirme işlemlerinden sonra Sağa Dön komutu olduğuna karar verilmiştir. Bu komuta göre, gezgin araç sağa dönme işlemini gerçekleştirir



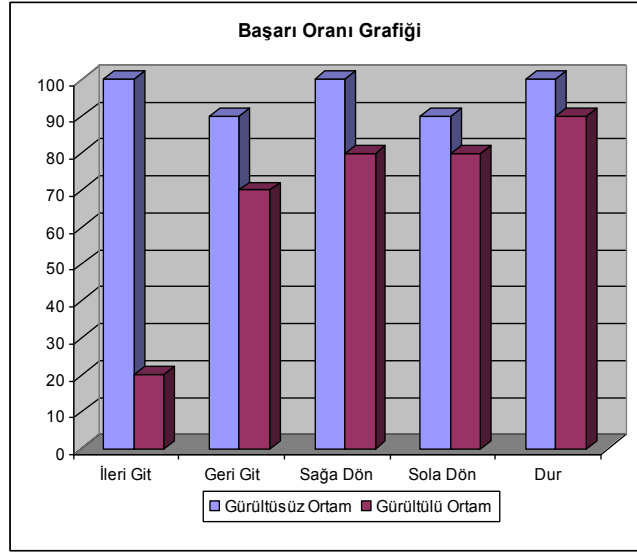
Şekil 8. Gezgin aracın görüntüsü

4. Sonuçlar

Bu tez çalışmasında, Türkçe komutları tanıyan ve komutlara uygun olarak gezgin aracın kontrolü gerçek zamanlı olarak gerçekleştirilmiştir. Yazılım kısmında kişiye bağımlı, yalıtık kelime tabanlı, komut – kontrol sistemi geliştirilmiştir. Donanım kısmında ise gezgin bir araç

oluşturulmuştur. Sistemin başarı oranı kişiye bağımlı tanıma için gürültüsüz ortamda, %96 gürültülü ortamda ise %68 olarak elde edilmiştir.

MFKK özellikleri konuşmacı bağımsız olarak ses verisine ilişkin özelliklerin ortaya konmasında etkili bir yöntemdir. VN algoritması, işlem sayısı ve uygulamadaki kolaylığı bakımından avantajlı bir yöntemdir. Yapılan bir çok deneme sonucunda geliştirilen yazılımın, sesli ifade tanıma doğruluk oranı ve hız değerleri yüksek çıkmıştır. Ancak sesli ifadeyi tanımayı olumsuz etkileyen bazı faktörlerin olduğu yapılan denemelerde ortaya çıkmıştır. Bunlardan en önemlisi görüldüğüdür.



Şekil 9. Başarı Oranı Grafiği

Gürültü özellikle sesli ifadenin başlangıç, ve bitişinin doğru şekilde belirlenmesini engellemektedir ve söylenen kelime doğru bir şekilde yakalanamamaktadır. Bu durum, doğal olarak tanıma performansını etkilemektedir. Sesli ifade tanımayı etkileyen diğer bir unsur da telaffuz farklılığıdır. Konuşmacının her iki telaffuz arasındaki benzer olması durumunda test sonuçları yüksek çıkmakta ve kelimelerin benzeme oranı artmaktadır. Eğer her iki telaffuz arasında fark olması durumunda ise test sonuçları düşük çıkmamakta ve yanlış eşleşmeler meydana gelmektedir.

Yapılan bu çalışma ile kişiye bağımlı yalıtık kelime tanıma yapabilen bir sistem tasarlanmıştır. Bu sistemin yerine fonem tabanlı bir sistem gerçekleştirilebilir. Bu sayede sistem kişiden bağımsız çalışabilir. Gerçekleştirilen bu sistem daha küçük ölçekli gömülü sistemler üzerinde başarılı bir şekilde çalıştırılabilir. Bu çalışma ile Türkçe sesli komutlar ile bir cihazın kontrolünü gerçekleştirmiştir. Bu çalışmanın sonraki adımlarında, otomobil içindeki cihazların kontrolü, bilgisayar fonksiyonlarının kontrolü, akıllı ev uygulamaları, engelli kişilerin sesli olarak cihazları kontrol etmeleri sağlanabilir.

Bu konunun oldukça yeni ve gelişmeye açık bir dal olduğu görülmüştür. Yeni geliştirilecek algoritmalarla hiç şüpheye yer bırakmayacak şekilde %100'e yakın başarı elde edilebilir.

5. Kaynaklar

- [1] Eskidere, Ö., İstatiksel Modelleme İle Konuşmacı Tanıma, Doktora Tezi, Uludağ Üniversitesi, Fen Bilimleri Enstitüsü, Bursa, Türkiye, 2007.
- [2] Çelebi, M., Ses Komut Tanıma ile Gezgin Araç Kontrolü, Yüksek Lisans Tezi, Marmara Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, Türkiye, 2012.
- [3] Kelebekler, E. ve İnal, M., "Otomobil İçindeki Cihazların Sesle Kontrolüne Yönelik Konuşma Tanıma Sisteminin Gerçek Zamanlı Laboratuar Uygulaması", Politeknik Dergisi, Cilt 11, Sayı 2, 109-114, 2008.
- [4] Madisetti, V.K., The Digital Signal Processing Handbook, CRC Pres, Second Edition, United States of America, 2010.
- [5] Meane, H.P., Advances in Audio And Speech Signal Processing, Idea Group Publishing, Mexico, 2007.
- [6] Rabiner, L.R. ve Sambur, M.R., "An Algorithm for Determining the Endpoints of Isolated Utterances", The Bell System Technical Journal, 1974.
- [7] Shaneh, M. ve Taheri, A., "Voice Command Recognition System Based on MFCC and VQ Algorithms", World Academy of Science, Engineering and Technology, 2009.
- [8] Rabiner, L.R. ve Schafer, R.W., Introduction to Digital Speech Processing, NOW The Essence of Knowledge, Hanover, USA, 2007.
- [9] Furui, S., Digital Speech Processing, Synthesis and Recognition Second Edition, Marcel Dekker INC, New York, USA, 2001.
- [10] Rabiner, L. ve Juang, B.H., Fundamentals of Speech Recognition, Prentice-Hall International, Inc., USA, 1993.
- [11] Linde, Y. ve ark., "An Algorithm for Vector Quantizer Design", IEEE Transactions on Communications, Volume 29, No 1, 1980.