

Altuzay Sınıflama ile Sesbirim Temelli Türkçe Yalıtık Kelime Tanıma Phonem-Based Isolated Turkish Word Recognition With Subspace Classifier

Serkan Keser¹, Rifat Edizkan²

¹Endüstriyel Elektronik Programı Kırşehir Ahi Evran Üniversitesi
skeser@ahievran.edu.tr

²Elektrik Elektronik Mühendisliği Bölümü Eskişehir Osmangazi Üniversitesi
redizkan@ogu.edu.tr

Özetçe

Bu çalışmada, ortak vektör yaklaşımı (OVY) ile sesbirim temelli Türkçe yalıtık kelime tanıma gerçekleştirilmiştir. OVY sesbirim sınıflamada kullanılmıştır. Sınıflama sonucu ortaya çıkan sesbirim dizisi için kelime çözümlemesi gereksiz karma adresleme (GKA) ile yapılmıştır. Sözlüklerinde farklı kelimeler kullanılan uygulamalar geliştirmek için kelime temelli ses tanıma yöntemleri yerine sesbirim temelli sistemler daha uygundur. Bu nedenle, bu çalışmada OVY'nin sesbirim temelli kelime tanımadaki başarımını değerlendirilmiştir. Deneysel çalışmada METU veri tabanından rastgele seçilmiş kelimelerden %70-80 kelime tanıma başarımı elde edilmiştir. OVY'de iyileştirmeler yaparak ve farklı dil çözümleme yöntemleri kullanılarak yüksek başarımli sınırlı sayıda kelime tanıma yapmak mümkün olabilir.

Abstract

In this study, phoneme-based isolated Turkish word recognition with Common Vector Approach (CVA) has been performed. CVA has been used to classify phonemes. The phoneme sequence obtained from the classification is decoded into the word using redundant hash addressing (RHA). The phoneme-based speech recognition is more suitable than the word-based speech recognition for implementing applications that use different words in their dictionaries. For that reason, in this study the CVA is evaluated to see whether it could be used in phoneme-based word recognition or not. In the experimental study we obtained the word recognition rates 70-80% from randomly selected words in METU database. It might be possible to obtain higher recognition rates by improving the CVA and by using different word decoding techniques.

1. Giriş

Ses tanıma; sesli komut (robot, taşıt ve bilgisayar kontrolü), IVR (Interactive Voice Responce), sesi yazıya dönüştürme, konuşmacı tanıma ve onaylama ve anahtar kelime tanıma gibi uygulamalarda kullanılmaktadır.

Ses tanıma sistemlerinde, Dynamic Time Warping (DTW), Hidden Markov Model (HMM), Neural Networks (NN) veya altuzay sınıflandırıcılar (FLDA, CLAFIC, OVY) kullanılmaktadır. Bunlar içinde "state of art" olarak bilenen

HMM, ticari ses tanıma sistemlerinde tercih edilmektedir. HMM ile karşılaştırılabilir başarımlar veren yöntemlerin ses tanımadaki kullanımı araştırmacıların ilgisini çekmektedir. Bunlar içinde bir altuzay sınıflandırma yöntemi olan OVY, sınırlı sayıda yalıtık kelime tanıma uygulamalarında HMM'le karşılaştırılabilir başarımlar vermektedir [1]. OVY ayrıca, sınıflandırma yöntemindeki basitlikten dolayı HMM göre daha hızlı çalışmaktadır. OVY yöntemi ses tanıma yanında, konuşmacı tanıma ve anahtar kelime tanıma uygulamalarında da kullanılmıştır [2,3].

OVY'nin ses tanıma uygulamalarında şimdiye kadar kelime temelli sınıflandırma yapılmış ve sınırlı sayıda kelime kullanılarak % 95'in üzerinde başarımlar elde edilmiştir[1].

Kelime temelli ses tanımadaki her kelime için veri tabanı ihtiyacı vardır. Bu nedenle kelime temelli tanıma sistemlerine yeni kelime eklemek her zaman kolay olmayabilir. Çünkü istenilen kelime için veri tabanı hemen elde edilemeyebilir. Bu da ses tanıma sisteminin ticari olarak kullanılmasını güçleştirir. Fakat sesbirim temelli ses tanıma yapılan sistemlerde sesbirim dizileri kullanılarak tanıyıcı sisteme yeni bir kelime eklemek mümkündür. Bu da tanıyıcı sistemin kullanılabilirliği arttırmaktadır.

Sesbirim temelli ses tanıma sistemlerinde dildeki her sesbirim sınıfı için bir model oluşturulur. Bu sistemlerde yüksek başarımlar elde etmek için kullanılan veri tabanının dildeki bütün farklı söyleyişleri kapsamalıdır. Veri tabanında sesbirim sınırlarını kesin olarak belirlemek çok zordur. Bunun yanında ses yolunun farklı şekil almasında dolayı sesbirimler farklı şekilde üretilebilmektedir. Bütün bunlar öznelik uzayında sesbirim dağılımlarının iç içe geçmesine neden olmaktadır ve sınıflandırma başarımını düşürmektedir. Ayrıca kelimelerin fonetik gösterimlerindeki bazı sesbirimler ortak eklenmeden dolayı konuşmacı tarafından yutulabilmektedir. Bu da sesbirim dizilerinden kelime tanımayı etkileyen başka bir etkidir. Sesbirimler kısa süreli olmalarından dolayı çok fazla bilgi içermemektedirler. Bu nedenlerden dolayı sesbirim sınıflandırma başarımları çok yüksek değildir [4]. Fakat sesbirim tanıma sonuçları dil modeliyle birleştirilerek çok yüksek kelime tanıma oranları elde edilebilmektedir. Literatürde daha çok üçlü sesbirim (triphone) kullanılarak ses tanıma yapılmaktadır. Üçlü sesbirimleriyle dildeki farklı söyleyişler modellenilebilir. Bir dilde sayıları binlerle ifade edilen üçlü sesbirimleri bulunmaktadır.

Bu çalışmada OVY ile sesbirim temelli Türkçe yalıtık kelime tanıma konusu üzerinde çalışılmıştır. OVY altuzay sınıflandırıcısıyla kelime söyleyişi içindeki sesbirim dizisi elde edilmektedir. Bu dizinin sözlükteki hangi kelimeyi oluşturduğunu çözümlmek için gereksiz karma adresleme (GKA) yönteminden yararlanılmıştır[5]. Çalışmada kelime tanıma başarımları METU veri tabanı üzerinde elde edilmiştir.

2. Sesbirim Temelli Yalıtık Kelime Tanıma

Sesbirim temelli yalıtık kelime tanıma sisteminde sesbirim sınıflama ve kelime çözümlene yapılmaktadır. Bu sistemlerde sesbirim sınıflama için değişik sınıflandırıcılar kullanılabilir. Bu çalışmada OVY altuzay sınıflandırma yöntemi kullanılmıştır. OVY'den elde edilen sesbirim dizisi RHA kullanılarak çözümlenmiş ve bu dizinin sözlükte hangi kelime olabileceği bulunmuştur. GKA ile kelime çözümlene yapılırken en basit seviyede dil modellemesi yapılmaktadır.

2.1. Ortak Vektör Yaklaşımı

OVY, ses ve görüntü tanıma uygulamalarında kullanılan bir altuzay sınıflama yöntemidir[6,7]. Bu yöntem ile her sınıfa ait değişmez özellikleri taşıyan ortak vektör elde edilir. OVY hem yeterli veri durumu ($m \geq n$), hem de yetersiz veri durumları ($m < n$) için uygulanabilir [9,11,12]. Burada m ses komut sınıfına ait vektör sayısını, n ise her vektördeki eleman sayısını göstermektedir. Bu çalışmada OVY yetersiz veri durumu için ses komut tanıma yapılmaktadır.

OVY' de eğitim ile bir sınıfa ait ortak vektör ve o sınıfın farksızlık altuzayını veren özvektörler elde edilir. Yetersiz veri durumunda ($m < n$) ortak vektör ve farksızlık altuzayı Gram-Schmidt dikleştirme yöntemi veya altuzay teknikleri kullanılarak elde edilebilir. Altuzay tekniğinde ortak vektör ile ortak değişinti matrisinin özvektörleri arasındaki ilişkiyi yararlanılır[8].

Eğitim setinde bir sınıfa ait öznelik vektörleri $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathcal{R}^n$ ile gösterilsin Bu sınıfa ait ortak değişinti matrisi Φ , Denklem (1)'den elde edilir.

$$\Phi = \sum_{i=1}^m (\mathbf{a}_i - \mathbf{a}_{ort})(\mathbf{a}_i - \mathbf{a}_{ort})^T \quad (1)$$

Burada \mathbf{a}_{ort} , sınıf ortalama vektörünü göstermektedir. OVY' de öznelik uzayı, farklılık altuzayı B ve farksızlık altuzayı B^\perp olmak üzere birbirine dik iki altuzaya ayrılır. Farksızlık altuzayı B^\perp , ortak değişinti matrisinin sıfır özdeğerlerine karşılık gelen özvektörler tarafından gerilir. Yetersiz veri durumu ($m < n$) için ortak değişinti matrisi Φ , $n-m+1$ adet sıfır özdeğere sahip olacaktır. Bu özdeğerler büyükten küçüğe sıralandığında fark ve farksızlık altuzayları için özdeğerler aşağıdaki şekilde seçilir:

$$\underbrace{\lambda_n \geq \lambda_{n-1} \geq \dots \geq \lambda_{n-m+2}}_{m-1} \geq \underbrace{\lambda_{n-m+1} \geq \dots \geq \lambda_2 \geq \lambda_1}_{n-m+1}$$

$$\lambda_i \neq 0 \quad \lambda_i = 0$$

Fark Altuzayı Farksızlık Altuzayı

λ_i özdeğerine karşılık gelen özvektör \mathbf{u}_i ile gösterilsin.

Bu durumda farksızlık altuzay izdüşüm matrisi aşağıdaki şekilde bulunur.

$$P^\perp = \sum_{j=1}^{n-m+1} \mathbf{u}_j \mathbf{u}_j^T \quad (2)$$

Bir sınıfa ait ortak vektör \mathbf{a}_{ortak} , o sınıfa ait her hangi bir öznelik vektörünün farksızlık altuzayına izdüşümünden elde edilir:

$$\mathbf{a}_{ortak} = P^\perp \mathbf{a}_i \quad \forall i = 1, 2, \dots, m \quad (3)$$

Yeterli veri durumunda ($m \geq n$), fark ve farksızlık alt uzayları tahmini olarak belirlenebilir. Ortak değişinti matrisinin sıfırdan farklı n adet özdeğeri olur. Bu özdeğerler küçükten büyüğe sıralandığında ilk k adet sıfıra yakın özdeğere karşılık gelen özvektörler farksızlık alt uzayını gerer. Yeterli veri durumunda ortak vektör, sınıf ortalama vektörünün farksızlık alt uzayına izdüşümünden elde edilir:

$$\mathbf{a}_{ortak} = \sum_{i=1}^k (\mathbf{a}_{ort}^T \mathbf{u}_i) \mathbf{u}_i \quad (4)$$

Burada \mathbf{a}_{ort} , öznelik vektörlerinin ortalamasını göstermektedir:

$$\mathbf{a}_{ort} = \frac{1}{m} \sum_{j=1}^m \mathbf{a}_j \quad (5)$$

Farksızlık altuzayını veren en küçük k özdeğer sayısı, özdeğerlerin toplamının belli L yüzdesine karşılık gelecek şekilde seçilebilir [1].

OVY'de tanıma yapılırken en küçük Öklid uzaklık ölçütü kullanılır. Bu ölçüte göre, öznelik vektörü \mathbf{a}_x 'in her sınıfın farksızlık altuzayındaki izdüşümünün, o sınıfın ortak vektörüne olan uzaklığı bulunur. Bu vektör, en küçük uzaklığı veren sınıfa atanır.

$$c^* = \underset{1 \leq c \leq C}{\text{enküçük}} \left\| \mathbf{P}^{c^\perp} \mathbf{a}_c - \mathbf{a}_{ortak}^c \right\| \quad (6)$$

Denklem (6)'de C , sınıf sayısını göstermektedir.

2.2. Gereksiz Karma Adresleme

Bu çalışmada, GKA kelime çözümlenmede kullanılmıştır. Altuzay sınıflandırıcı kullanılarak bir kelimeye ait söyleyiş içindeki sesbirimler elde edilir. Bu sesbirimlerin sözlükteki hangi kelimeye ait olduğu GKA tarafından belirlenir [5,9,10].

Karma yönteminde, veri arama işlemini bir çırpıda yapmak ve aranana doğrudan ulaşmak için karma fonksiyonu kullanılır. Bu fonksiyona, arama işleminde kullanılacak anahtar değeri girilir ve karşılığında bir tamsayı (indis) değeri elde edilir. Verilerin tablodaki yerleri buradan elde edilen indis değeriyle belirlenir.

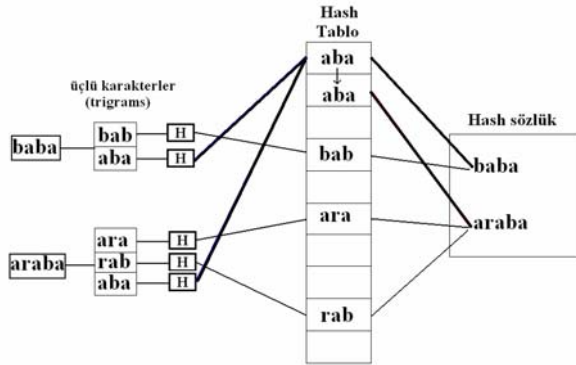
Karma tablosu veri dizisinden oluşur. Bu tabloda her veriye, karma fonksiyondan anahtar değerine göre üretilen indisle erişilebilir.

Karma fonksiyonu iki farklı anahtar değeri için aynı değeri üretiyorsa çatışma oluşur. Çatışmayı en aza indirmek

için karma fonksiyonunu iyi seçmek gerekir. Fakat bazı durumlarda çatışmayı önleyecek karma fonksiyon bulunamaz. Çatışmayı önlemek için en çok kullanılanları açık adresleme ve bağlı liste yöntemi kullanılmaktadır [11].

2.2.1. Açık Adresleme

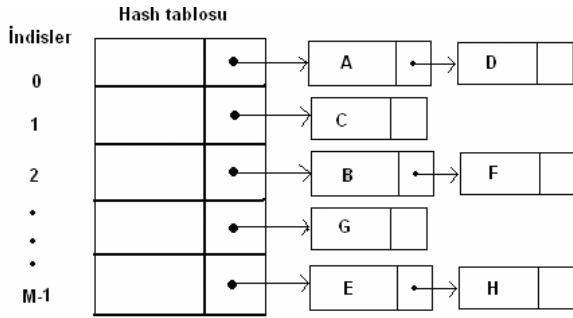
Çatışma durumunda veri, çatışmanın olduğu indisten sonraki boş kısma (slot) yerleştirilir. Bu işlem için tablo boyutu kadar farklı değerler üreten karma fonksiyonları $(h_0, h_1, h_2, \dots, h_{M-1})$ oluşturulur. Bu fonksiyonlar çatışmanın olduğu indisten itibaren indis değeri üretirler. Fonksiyonlar kullanılarak boş bir kısım bulunana kadar arama yapılır. Bu şekilde çatışma oluşturan veriler tablodaki boş kısımlara yerleştirilir. Şekil 1'de "baba" ve "araba" kelimeleri için çatışmanın açık adreslemeyle nasıl çözümlendiği gösterilmektedir.



Şekil 1: Açık adresleme.

2.2.2. Bağlı liste kullanımı

Çatışma durumunda, çatışan veriler bir liste halinde bağlanırlar [12]. Karma fonksiyonu bu listelerden birine ait indis ürettiğinde bu bağlı liste üzerinden dolaşarak karar kuralı temelinde doğru kelimeye ulaşmaya çalışır (Şekil 2).



Şekil 2: Bağlı liste kullanımı.

2.2.3. GKA ile Kelime Çözümleme

Karma sözlüğün, tanınması istenilen tüm kelimeleri kapsamaması gerekmektedir. Bilinmeyen bir kelime için karma fonksiyonu ile indis değeri üretilir. Indis değerinin gösterdiği bellek gözündeki adres bilgisi bir işaretçide tutulur [5,10,12]. İşaretçi karma tablosundaki N-gram'lara karşılık gelen sözlük kelimelerini gösterir. Bu sayede bilinmeyen kelime için

sözlükte kelime adayları bulunur. Aday kelimeler içinde özellik uzaklığı en küçük veren kelime aranan kelimelidir.

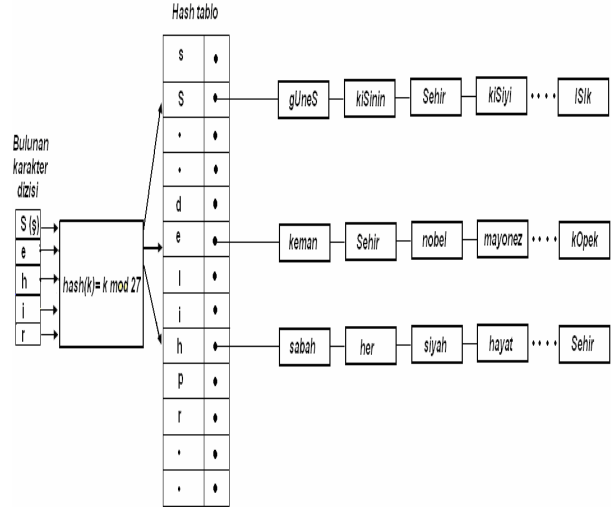
$$FD(A,B) = \text{enbüyük}(n_A, n_B) - n_B \quad (7)$$

Denklem (7)'de özellik uzaklığı (feature distance), bilinmeyen kelimeye ait N-gramlar ile karma tablodaki N-gramlar arasındaki uzaklık ölçütünü göstermektedir. Denklemde n_A ve n_B , A ve B dizilerindeki N-gram'ların sayısını ve n_E ise iki dizi arasındaki benzer N-gram sayısını göstermektedir. N-gram'lar, sözcüklere ait karakter dizisinin N uzunluklu alt dizisi şeklinde tanımlanır. N-gram'ın boyutu bir ise unigram, iki ise bigram ve üç ise trigram olarak adlandırılmaktadır.

GKA'nın uygulamasında önce sesbirimlerin alfabetik karşılıkları aşağıda verilen karma fonksiyonu ile karma tablosuna yerleştirildi.

$$\text{hash}(k) = k \text{ mod } 27$$

Burada k , sesbirimin alfabetik karşılığının ASCII değeridir. Daha sonra, sesbirim dizisinde elde edilen kelimeye ait tekli karakter (unigram) dizileri kullanılarak kelime çözümleme yapılmıştır. Çatışmanın çözülmesi için bağlı liste yöntemi kullanılmıştır. Şekil 3'de "şehir" kelimesi için GKA kullanılarak kelime çözümleme gösterilmiştir.



Şekil 3: GKA kullanarak kelime çözümleme.

3. Deneysel Çalışma

Deneysel çalışmada kullanılan METU veritabanı, 120 konuşmacıdan alınan 40'ar adet üçlü ses dengeli cümlelerin seslendirilmesi ile Türkçede 29 harfe karşılık gelen 38 sesbirim göz önünde bulundurularak oluşturulmuştur[13]. METUbet fonetik alfabesinde 38 sesbirim sınıfı bulunmaktadır. Deneysel çalışmada kelime başarımları kişi bağımlı ve kişi bağımsız durumlar için elde edilmiştir[14].

METU veritabanında ses kayıtları 16 kHz'de örneklendirilmiştir ve her örnek 16 bit'le gösterilmektedir. OYV'de sesbirimler için vektör boyutu, veri tabanındaki

ortalama sesbirim uzunluğu olarak alınmıştır. Öznitelik vektörleri 20 ms'lik çerçeveler üzerinden elde edilmiştir. Bu çerçeveler 12 MFCC parametresi ve çerçeve enerjisinden oluşan 13 parametreyle temsil edilmektedir. Çerçeveler arası %50 üst üste binme yapılmıştır. Ardışık 7 çerçeveden elde edilen öznitelikler arka arkaya eklenerek 80 ms çerçeveyi temsil eden 91 boyutlu (13x7) öznitelik vektörü elde edilir.

Kişi bağımlı kelime tanıma çalışmasında seçilen bir kişiden alınan 15 örnekten 10'u eğitim, 5'i test için kullanılmıştır. Kişi bağımsız kelime tanımda ise her sesbirim sınıfı için 300 örnek eğitim ve 100 örnek test için kullanılmıştır. Kişi bağımlı ve bağımsız tanımda sırasıyla yetersiz ve yeterli veri durumu için OVY uygulanmıştır. Sesbirim temelli yalıtık kelime tanımda, veri tabanından kişi bağımsız ve kişi bağımlı tanıma için sırasıyla rastgele seçilen 90 ve 40 kelime kullanılmıştır.

METU veri tabanında bazı sesbirimler çok sık geçmemektedir. Bu nedenle allofonlar birleştirilerek 27 sesbirim kullanılmıştır.

Ünlü sesbirimler (8) : A, E, O, OE, U, UE, I, IY
 Ünsüz sesbirimler (19) : B, C, CH, D, F, G, H, K, L, M, N, P, R, S, SH, T, VV, Y, Z

Ünlü ve ünsüz sesbirimlerin OVY ile sınıflama başarımları Tablo 1'de verilmektedir.

Tablo 1: Sesbirim tanıma başarımları (%).

	Eğitim		Test	
	Ünlü	Ünsüz	Ünlü	Ünsüz
Kişi Bağımlı	100	100	63.30	63.06
Kişi Bağımsız	85.7	90	48.75	53.02

OVY ile bir kelime söyleyişi içindeki sesbirimler belirlenirken ünlü ve ünsüz sınıfları, ünlü ve ünsüz ses bölgelerine ayrı ayrı uygulanmıştır. Bunu yapabilmek için ses, sıfır geçiş ve enerji eşik değerine göre ünlü ve ünsüz bölgelere ayrılmıştır.

Yapılan çalışmada, ses bölgesi üzerinde 20 ms'lik çerçeve 10ms'lik kaydırılarak OVY sınıflandırıcısıyla sesbirim dizisi elde edilir. Her ünlü/ünsüz bölgesine bir sesbirim atanır[14]. Bu şekilde elde edilen sesbirim dizisinin çözümlenmesi GKA tarafından yapılır.

Deneysel çalışmada kişi bağımsız kelime tanıma başarımları %72.22 olarak elde edilmiştir. Kişi bağımlı kelime tanıma başarımları ise % 82.5'dir.

4. Sonuçlar ve Öneriler

Bu çalışmada OVY kullanılarak sesbirim temelli yalıtık kelime tanıma gerçekleştirilmiştir. Sesbirim ve kelime tanıma başarımları METU veri tabanı üzerinde elde edilmiştir. Kelime çözümlenme için GKA kullanılmıştır. Çalışmada sadece klasik OVY yöntemi kullanılmış ve altuzaylar üzerinde başarımları artıracak çalışmalar yapılmamıştır. Tanıma başarımlarını artırmak için yeni sınıflandırıcılar veya sınıflandırıcı birleştirme yöntemleri kullanılabilir. GKA yerine dil yapısındaki sesbirimlerin çıkma olasılıkları ile bunların birbirini takip etme olasılıkları incelenerek oluşturulacak dil modeliyle yüksek kelime tanıma başarımları elde edilebilir.

5. Teşekkür

Bu çalışmaya desteklerinden dolayı Prof.Dr. Tolga Çiloğlu'na teşekkür ederiz.

6. Kaynakça

- [1] Gülmezoğlu, M. B. , Dzhafarov, V. , Edizkan R. and Barkana, A., "The Common Vector Approach and Its Comparison with Other Subspace Methods in Case of Sufficient Data", *Computer Speech and Language*, (21:2) 266-281, 2007.
- [2] Gülmezoğlu, M. B. and Barkana, A., "Text-Dependent Speaker Recognition by Using Gram-Schmidt Orthogonalization Method", *IASTED International Conference Signal Processing and Communications*, , 1998, p. 438-440.
- [3] Bayrakçı, M.K., Çay, M.A. ve Barkana, A., "Ortak Yöney Yaklaşımı ile Anahtar Sözcük Yakalama", *15. Sinyal İşleme ve İletişim Uygulamaları Kurultayı*, 2007.
- [4] Arısoy, E. ve Arslan, L.M., "Turkish Dictation System for Broadcast News Applications", *13. Sinyal İşleme ve İletişim Uygulamaları Kurultayı*, 2005, p. 629 – 632.
- [5] Kohonen, T., Rittinen, H., Jalanko, M., Reuhkala, E. and Haltsonen, S., "A thousand-word recognition system based on the learning subspace method and redundant hash addressing", *Proceedings of the Fifth International Conference on Pattern Recognition*, 1980, p. 158-165.
- [6] Gülmezoglu, M.B, Dzhafarov, V., Keskin, M. and Barkana, A., "A Novel Approach to Isolated Word Recognition", *Speech and Audio Processing, IEEE Transactions on* (7: 6) 620 – 628, 1999.
- [7] Edizkan, R., Tiryaki, B., Büyükcın, T. ve Uzun, İ., 2007, "Ses Komut Tanıma İle Gezgın Araç Kontrolü", *Akademik Bilişim*, 2007.
- [8] Gülmezoglu, M.B., Dzhafarov, V. and Barkana, A., "The Common Vector Approach and its Relation to Principal Component Analysis", *IEEE Transactions on Speech and Audio Processing*, 9(6):655 – 662, 2001.
- [9] Erickson, J., *Hash Tables*, University of Illinois, Lecture Notes, 2002.
- [10] Reuhkala, E., Jalanko, M. and Kohonen, T., "A Redundant Hash Addressing Method Adapted For The Postprocessing and Error-Correction of Computer-Recognized Speech", *Helsinki University of Technology* 1979, p. 591-594.
- [11] Yağmur, O., *Hashing Fonksiyonları, Hash Tablosu ve Collision*. <http://www.csharpnedir.com>, 11 Mart 2007.
- [12] Çölkesen, R., *Veri Yapıları ve Algoritmalar*, Papatya Yayıncılık, 2004.
- [13] Salor, Ö., Pellom, B., Çiloğlu, T., Hacıoğlu, K. and Demirekler, M., "New Corpora and Tools for Turkish Speech Research", *10. Sinyal İşleme ve İletişim Uygulamaları Kurultayı*, 2002, s. 476-481.
- [14] Keser S., *Ortak vektör Yaklaşımı ile Fonem Tabanlı Türkçe Yalıtık Kelime Tanıma*, Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, 2008.