

# Canlı İnternet Yayınları İçin Otomatik Konuşma Tanıma Tekniği Kullanılarak Alt Yazı Oluşturulması

Kutan Koruyan

Yönetim Bilişim Sistemleri Bölümü, İİBF, Dokuz Eylül Üniversitesi, İzmir, Türkiye

[kutan.koruyan@deu.edu.tr](mailto:kutan.koruyan@deu.edu.tr)

(Geliş/Received: 11.02.2015; Kabul/Accepted: 10.04.2015)

DOI: 10.17671/btd.31441

**Özet**— Televizyon, sinema veya diğer farklı görüntülerde konuşulan dili, çevirisini gösterme veya görüntü ve ses hakkında metin şeklinde bilgi verme amaçlı uygulanan alt yazı tekniği, 1900'lerin başlarında kullanılmaya başlanmış ve gelişerek günümüzdeki haline gelmiştir. Bilişimdeki gelişmeler ise alt yazı tekniklerinin ilerlemesine büyük katkı sağlamış, özellikle konuşmaların metne dönüştürülmesi konuşma tanıma teknikleri ile daha kolay bir hale gelmiştir. Bunun yanında, işitme engelliler için kullanılan alt yazılar ise özellikle canlı yayınlarda konuşma tanıma tekniği ile işaret diline alternatif olarak kullanılmaktadır. Bu teknik daha çok ticari amaçlı özel donanım ve yazılımlarla beraber kullanılmakta, bireysel kullanım veya küçük ölçekli kurumlar için maliyet oluşturmaktadır. 2011'de Google Chrome'un Türkçe'yi de destekleyen sesle aramayı dünyaya duyurması ise bu çalışmanın çıkış noktasını oluşturmuştur. Bu çalışmada, bir medya sunucusu yardımıyla internet sayfasında canlı yayınlanan bir videodaki konuşmaların Google'ın desteklediği açık kaynak kodlu Web Speech API kullanılarak metne dönüştürülmesi ve anlık alt yazı haline getirilmesi uygulaması anlatılmaktadır. Çalışmada, web sayfasında video yayını HTML5 dilinin getirdiği video elementi ile sağlanmakta, web uygulaması JavaScript ve PHP programlama dilleri ve jQuery kütüphanesi kullanılarak yazılmıştır.

**Anahtar Kelimeler**— Otomatik Konuşma Tanıma, Canlı İnternet Yayını, Canlı Altyazı, HTML5, İnternet

## Generating Captions Using Automatic Speech Recognition Technique for Live Webcasts

**Abstract**— Captioning technique used to display speaking language or its translation or to give information about images or sounds on television, cinema or other images as text has been used since the beginning of the 1900s and has developed to take its contemporary form. The development of informatics has greatly contributed to the progress of captioning techniques; it has especially become easier to convert speech to text with the aid of speech recognition. Furthermore, the captions for the hearing impaired, especially with speech recognition technique, is an alternative to sign language on live events. This technique is commercial and predominantly used with special hardware and software, and increases costs for the individual usage or small-sized companies. The announcement of voice search of Google Chrome in 2011 has been the start of this work. In this study, an application converting the speech to the text and displaying the live captions on a video broadcasted live on a web page using Google supported open source Web Speech API with the help of a media server is represented. The broadcast of a video on a web page is performed by the HTML5 video element, and the web application is coded using JavaScript and PHP programming languages and jQuery library.

**Keywords**— Automatic Speech Recognition, Live Webcast, Live Captions, HTML5, İnternet

## 1. GİRİŞ (INTRODUCTION)

1900'lü yılların başlarında ilk olarak sinema filmlerinde kullanılmaya başlayan alt yazı tekniği, görüntülerdeki yabancı dilde konuşmaların çeviri metin şeklinde gösterilmesi, işitme engelliler için konuşma metinleri ve/veya konuşma dışı diğer ortam seslerinin açıklayıcı yazı olarak verilmesi veya gürültülü ortamlarda iletilmesi gereken bilgilerin alt yazı biçiminde sunumu gibi daha birçok farklı şekilde kullanılmaktadır.

Günümüzde görsel medyanın toplumdaki yeri tartışılmazken, alt yazı tekniğinin işitme engelliler için kullanımı engelli bireylerin toplumla bütünleşik olması açısından büyük önem taşımaktadır. Örneğin, ABD'de 1990'lardan itibaren televizyon programlarının bazı istisnalar dışında alt yazılı verilmesi zorunlu hale getirilmiştir [1]. Ayrıca, 2012'den itibaren internetten video akışı şeklinde tekrardan yayınlanan televizyon programlarına da alt yazı zorunluluğu getirilmiştir [2]. CISCO'nun bir araştırmasına göre 2018'de dünya internet trafiğinin %84'ünü internet video akışının oluşturacağı düşünüldüğünde, internet video yayını ve bununla birlikte kullanılan alt yazıların önemi büyüktür [3]. Buna örnek olarak, YouTube 2006'da barındırdığı çevrimiçi videolara alt yazı ekleme özelliği sunmuştur [4]. Bununla birlikte, 2009'da Google'ın otomatik konuşma tanıma (OKT) teknolojisi kullanılarak YouTube'a yüklenen videolar için otomatik alt yazı oluşturma özelliğini duyurulmuştur [5]. Şu anda YouTube'un OKT kullanılarak otomatik alt yazı oluşturma özelliği ne yazık ki Türkçe'nin aralarında bulunmadığı on dili kapsamaktadır [6]. Son olarak, 2011'de YouTube'dan videoların canlı (gerçek zamanlı) yayınlanabileceği duyurulmuş, videolara canlı alt yazı eklenip bunun da desteklenen farklı yazılımlarla yapılabileceği söylenmektedir ve Google'ın desteklediği bu yazılım veya hizmetler ücretlidir [7], [8].

Son yıllarda dünya genelinde kullanıcı başına internet hızlarındaki artış ve HTML5'in bir eklenti (örneğin Adobe Flash) kullanmadan web sayfalarında video gösterebilme yeteneği, internetten video yayınına daha cazip hale getirmiştir. İnternetten yayınlanan videolara rağbetin artmasının yanında, günümüz internet çağında işitme engelliler için önem arz eden internetten canlı yayınların (Televizyon programları, spor karşılaşmaları, konserler, uzaktan eğitim, vb.) bireylere alt yazı ile birlikte ulaşması ve bunun az bir maliyetle veya maliyetsiz açık kaynak kodları kullanılarak oluşturabilme düşüncesi, bu çalışmanın ana çıkış noktasını oluşturmuştur.

Bu çalışmada, bir medya sunucusu kullanılarak internetten canlı video yayını yapılırken konuşmaların canlı alt yazı olarak web sayfasında video ile beraber eş zamanlı gösterimi uygulaması anlatılmaktadır. Web sayfasında video sunumu HTML5 dilinin olanak sağladığı

video elementi ile yapılmaktadır. Konuşmaların yazıya dökülmesi için Google'ın desteklediği Web Speech API'den (WSA), alt yazıya dönüştürülmesi için ise jQuery kütüphanesi ve JavaScript dilinden, metinlerin veri tabanı sunucusuna gönderilmesinde PHP dilinden yararlanılmıştır.

Bu çalışma konusuna örnek olacak çalışmalar daha çok canlı sunumlar ve ders anlatımlarına yönelik olup, görüntülü internet yayınlarına eş zamanlı alt yazı eklenmesi uygulamalarının halen geliştirilmekte olduğu görülmektedir. Yapılan bir çalışmada internetten canlı yayınlarda kullanılan OKT test edilerek, uygulamada kelime hata oranının (KHO) yaklaşık %45 dolaylarında olduğunu, Türkçe için yapılan diğer bir çalışmada ise %37'ye ulaşıldığı söylenmiştir [9], [10]. Webde OKT kullanımına yönelik olarak işitme engelliler için yapılan başka bir çalışmada halen OKT teknolojisinin tam verimle kullanılmadığı ve %2-4 oranında insan hatasından kaynaklanan transkripsiyon (çeviri yazı) KHO'na ulaşamadığı belirtilmiştir [11]. İşitme engelli ve yabancı öğrenciler için geliştirilen web tabanlı eğitim projesinde ise OKT teknolojisi kullanarak canlı web videolarına eşzamanlı alt yazı eklemek suretiyle öğrencilere mesafeden bağımsız şekilde ulaşabilmektedir [12]. Ayrıca OKT teknolojisinin eğitimde öğrenciler üzerinde daha verimli olduğu belirtilmiştir [13].

## 2. HTML5 VIDEO (HTML5 VIDEO)

Bir web sitesinde herhangi bir eklenti kullanmadan video sunumu yapabilmeye olanak tanıyan HTML5 video elementi ilk olarak 2007'de Opera tarafından duyurulmuştur. Apple'ın Safari tarayıcısı video elementini destekleyen ilk tarayıcıdır. Video akış sitelerinden ise ilk olarak destekleyen Dailymotion'dır [14].

Sayılaştırılmış görüntü veya sesin sıkıştırılması için kullanılan çözücüler (dekoder) videonun web tarayıcısında oynatılabilmesi için önem arz etmektedir. Her tarayıcı her çözücüye desteklememekle birlikte, Google Chrome, sıklıkla kullanılan Theora (.ogg), H.264 (.mp4), VP8 (.webm) ve VP9 (.webm) çözücülerini ile sıkıştırılmış videoları oynatabilmektedir.

Video akışını sağlayan bir web sayfasında, video elementi içinde oynatılacak videolar için medya kontrolleri bulunmaktadır (oynat, durdur, sessiz, otomatik oynat, oynatıcı boyutları, vb). Video elementinin ast elementi olan source, video ile ilgili detayların verilmesi için kullanılmaktadır (video kaynağı, çözücü, vb). Yine ast element olan track elementi ise alt yazı dosyası kaynağı, dili, formatı gibi özellikleri barındırmaktadır (Şekil 1). Track elementinin desteklediği WEBVTT (The Web Video Text Tracks Format, .vtt) alt yazı formatı ise alt yazı metninin SS:DD:SN,SLS zaman aralıklarında

tanımlanmasını ve bu zaman aralıklarında altyazıların video üzerinde gösterimini sağlamaktadır (Şekil 2).

```
<video width="320" height="240" controls>
  <source src="MerhabaDunya.mp4" type="video/mp4">
  <track src="MerhabaDunya.vtt" kind="subtitles"
    srclang="tr" label="Türkçe">
</video>
```

Şekil 1. HTML5 video örneği  
(Example of HTML5 video)

```
WEBVTT
00:01.256 --> 00:04.412
Merhaba Dünya!

00:05.114 --> 00:09.854
Bugün çok güzel bir gün.
```

Şekil 2. WEBVTT örneği  
(Example of WEBVTT)

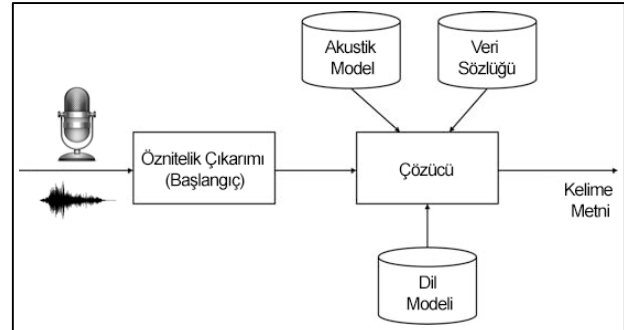
### 3. OTOMATİK KONUŞMA TANIMA TEKNOLOJİSİ VE OTOMATİK KONUŞMA TANIMA İLE ALTYAZI OLUŞTURMA (AUTOMATIC SPEECH RECOGNITION TECHNOLOGY AND GENERATING CAPTIONS USING AUTOMATIC SPEECH RECOGNITION)

İnsanlığın bilinen tarihinden beri bireyler arası en önemli iletişim aracı konuşmaktır. Nesilden nesile bilginin aktarımı önceleri konuşma, daha sonra ise yazı ile yapılmıştır. Günümüz teknolojisinde bilgisayar ve internetin gelişimi ile birlikte bilgi aktarımının büyük bir bölümü ise bu gelişen teknolojiler ile yapılmaktadır. Bilgisayarlarda bilgi girişinin çoğunluğu kullandığımız klavyeler ile yapılıyor olsa da OKT teknolojisi ile birlikte sadece konuşarak söylediklerimiz metne dönüştürülebilmektedir. Bir anlamda, OKT teknolojisi ile bilgisayarlar "söylediklerimizi" anlayabilmekte ve ayrıca bu söylediklerimiz de birer bilgisayar komutuna dönüştürülebilmektedir.

Konuşma tanıma teknolojisi ilk olarak 1920'lerde ismi söylendiğinde tepki veren oyuncak bir köpek olan "Radio Rex" ile başlamış ve büyük bir ilerleme göstererek günümüzdeki haline gelmiştir [15]. Bu ilerlemenin kanıtı olan OKT'nin yaygın uygulama alanları aşağıda verilmektedir [16], [17], [18]:

- Akıllı telefonlar,
- Bilgisayar yönetimi,
- Ses denetimli cihazlar ve robotlar,
- Oyuncaklar,
- Araç navigasyonu,
- Çağrı merkezleri,
- Engelliler için destek,
- Gazetecilik,
- Eğitim,
- Askeri uygulamalar,
- Görsel medya,
- Sağlık, vb.

OKT teknolojisinin genel prensibi, kullanıcının mikrofon veya telefon gibi bir cihaza konuşarak, cihaza bağlı bilgisayarın belirli bir görev için konuşmayı metne çevirmesine dayanmaktadır [19]. Temel olarak OKT'nin çalışma prensibi ise başlangıç aşaması ve çözücü olarak iki ana bileşenden oluşmaktadır (Şekil 3) [20]. Başlangıç aşaması, konuşma ses sinyalini temsil eden akustik özellikleri çıkarmakta ve konuşma sinyallerini bir sonraki adıma taşımaktadır. Çözücü ise akustik, veri sözlüğü (lexicon) ve dil modeline dayanarak giriş akustik özellikleri için kelime dizilerinde en iyi eşleşmeleri aramaktadır. Akustik model konuşanın konuşma şeklini ve çevre sesini tanımlamak için kullanılmaktadır. Veri sözlüğü, sözlük içindeki kelimelerin nasıl telaffuz edildiğini barındırmaktadır. Dil modeli ise konuşma dizisinde bir sonraki kelimeyi tahmin etmek için kullanılmaktadır [21].



Şekil 3. Otomatik konuşma tanıma sistemi diyagramı [20]  
(Diagram of automatic speech recognition system)

OKT'nin her ne kadar fazla uygulama alanı olsa da bu teknolojinin verimli kullanımını aksatacak bazı engeller de bulunmaktadır. Genel olarak bu dezavantajlar; konuşma farklılıkları (vurgu, ses, cinsiyet, aksan), ortam gürültüsü, söyleniş benzer kelimeler olarak sıralanmaktadır [22].

OKT'nin yaygın kullanım alanlarından biri görüntüler için alt yazı oluşturmaktır. Önceleri televizyonlarda -yayın canlı olsun olmasın- bir stenograf yardımı ile konuşmalar metne, sonra da alt yazıya dönüştürülmekteydi. OKT teknolojisi kullanımının yaygınlaşmasıyla artık günümüzde alt yazı oluşturmada bu teknolojiye yararlanılmaktadır. Gerçek zamanlı alt yazılar ise işitsel konuşmanın okul, toplantı, gündelik sohbet ve diğer gerçek zamanlı durumlarda işitme engelliler için görsel metin oluşturulması için kullanılmaktadır [23].

Bu çalışmada konuşmaların metne dönüştürülmesi için W3C Speech API Community Group (W3C-SACG) tarafından 2012 sonlarında tanıtılan, kullanıcılara açık WSA'den yararlanılmıştır. Google Chrome'un 25'inci sürümünden sonra kullanılabilen bu API, Türkçe'nin de aralarında bulunduğu 62 dil ve lehçeyi desteklemektedir.

WSA işlem süreci genel olarak, bir mikrofondan konuşulan sesin FLAC (Free Lossless Audio Codec)

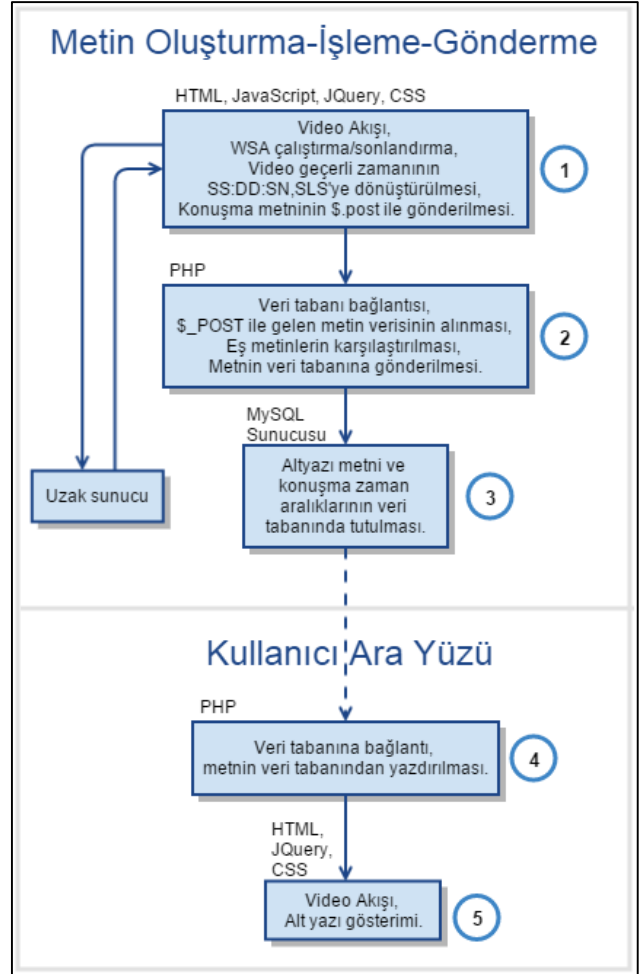
formatında kaydı alınarak, HTTPS POST isteğiyle konuşma tanıma web servisine (uzak sunucu) göndermesiyle başlamakta ve buradan söylenen sözcükleri metin olarak geri göndermesiyle noktalanmaktadır [24]. API, sunucu veya istemci tabanlı olarak yapılan işlemleri desteklemektedir. Aynı zamanda, tek bir kelime veya kesintisiz konuşmaya da olanak sağlamaktadır [25].

WSA'yi çalıştırmak için ise kod kısmında ilk olarak Google Chrome kütüphanesinde yer alan `webkitSpeechRecognition()` nesnesini oluşturmak gerekmektedir. Bu nesne; konuşma arabirimi, konuşma tanıma başlama-sonlanma (`onstart-onend`), hata (`onerror`), sonuçların alınması (`onresult`) veya konuşma devamlı olup olmadığı (`continuous`), mikrofon ve dil denetimi gibi bazı özellik ve olay tetikleyicilerini kullanabilmeyi sağlamaktadır. Ayrıca, konuşma tanıma sonuçlarından geri dönen geçici (`interim_transcript`) ve nihai (`final_transcript`) metin olmak üzere iki farklı sonuç bulunmaktadır. Geçici sonuçlar çözücü yardımıyla (veri sözlüğü, dil modeli ve akustik model) en doğru sonuca (nihai sonuç) dönüşmektedir. Bunun anlamı, çözücünün en doğru sonuca ulaşma gayesinden dolayı, bazen geçici sonuçlar nihai sonuçlardan farklılıklar gösterebilmektedir.

#### 4. METODOLOJİ (METHODOLOGY)

Bu çalışmada, internetten canlı yayınlanan videolara OKT tekniği kullanılarak canlı alt yazı oluşturma adımları anlatılmıştır. Çalışmada yapılan adımlar Şekil 4'de verilmektedir.

Sistem, metin oluşturma-işleme-gönderme (MOİG) ve kullanıcı ara yüzü (KAY) olmak üzere iki farklı kısımda tasarlanmıştır. Genel olarak, MOİG konuşmaların metin olarak oluşturulması ve veri tabanına gönderilmesi, KAY ise video ve alt yazının izleyici tarafından eşzamanlı görüntülenmesini sağlamaktadır.



Şekil 4. Otomatik konuşma tanıma tekniği kullanılarak canlı internet yayınlarına alt yazı oluşturma akış şeması (Flowchart of generating captions for live webcasts using automatic speech recognition technique)

MOİG kısmının ilk aşamasında (Şekil 4, (1)) komut dizisinde bir tetikleyici yardımıyla `webkitSpeechRecognition` nesnesi başlatılmakta, konuşulanlar uzak sunucuya gönderilmekte ve uzak sunucudan dönen geçici ve nihai konuşma metinleri kontrol amaçlı bir blok (`div`) veya satır içi (`span`) etiketi içine yazdırılmaktadır.

Sayfada aynı zamanda video akışı da izlenebilmektedir. Bu video akışı hem görüntüyü takip etmek, hem de HTML5 video elementi özelliklerinden biri olan ve o anki video akışındaki saniye cinsinden geçerli zamanı elde etmede kullanılan `currentTime` için gereklidir. Videolar daha sonra banttan yayınlanmak istendiğinde işte bu kaydedilen metin ve zaman aralıkları, veri tabanından dışarı aktarılıp WEBVTT formatına dönüştürüldükten sonra alt yazı olarak kullanılabilir. Bu yüzden, `currentTime` özelliği biriminin saniye cinsinden olmasından dolayı, zaman değişkeni web uygulamasında WEBVTT'nin standardı olan SS:DD:SN,SLS formatına dönüştürülmektedir.

Uzak sunucudan dönen konuşma metin ve zaman değişkenleri ise jQuery.post() işlemi ile PHP komut dizisine aktarılmakta (2) ve buradan da veri tabanına (3) gönderilmektedir. Ek olarak, PHP’de veri tabanını çok fazla meşgul etmemek amacıyla yeni gelen metinle veri tabanı tablosundaki en son satırdaki metnin aynı olup olmadığı karşılaştırılıp, aynı değilse veri tabanına gönderilmektedir. Böylelikle, satırlarca eş olan metinlerden kaçınılmaktadır (Şekil 5).

```
INSERT INTO `live_caps` (`_id`, `ct1`, `ct2`, `text`) VALUES
(1, '00:00:24,934', '00:00:24,940', 'merhaba'),
(2, '00:00:24,940', '00:00:25,209', 'merhabalar'),
(3, '00:00:25,209', '00:00:25,273', 'merhabalar bugün'),
(4, '00:00:25,273', '00:00:25,347', 'merhabalar bugün farklı'),
(5, '00:00:25,347', '00:00:25,468', 'merhabalar bugün farklı'),
(6, '00:00:25,468', '00:00:25,610', 'merhabalar bugün farklı konuda'),
(7, '00:00:25,610', '00:00:26,003', 'merhabalar bugün farklı konudan bahsedecem);
```

Şekil 5. Zaman aralıkları ve konuşma verileri  
(Time intervals and speech data)

KAY’da ise video yayınına ek olarak, jQuery.load() fonksiyonu yardımıyla veri tabanı tablosundaki son satırın istenmesine olanak sağlayan başka bir PHP komut dizisi (4) bir blok etiketi içine yüklenmektedir (5). Böylece, MOİG’den veri tabanına gönderilen en son metin videonun altında gösterilebilmektedir. Uygulamanın hem MOİG hem de KAY ekranları Şekil 6’de verilmektedir.



Şekil 6. Metin oluşturma-işleme-gönderme ve kullanıcı ara yüzü web sayfaları.

(Web pages of text generating-processing-posting of text and user interface)

## 5. SONUÇLAR VE TARTIŞMALAR (CONCLUSIONS AND DISCUSSION)

Günümüzde görsel medya insanların temel iletişim araçlarından biri olmuştur. Yayınlar ister televizyondan, ister internetten olsun birçok insan için önemli bir bilgi kaynağı haline gelmiştir. 1990’lardan itibaren yayınları alt yazı ile birlikte verebilme çabası engelli bireylerin de bu bilgi kaynağından yararlanabilmesini mümkün kılmıştır. İnternette video akışlarına yıllar geçtikçe rağbetin artması, alt yazılı video yayın teknolojilerini de geliştirmiş ve yayın canlı olsun olmasın görüntülere alt yazı eklenebilmesi olanaklı hale gelmiştir. Belirli bir ücret karşılığında verilen bu hizmetlere karşılık, bu çalışmada özellikle açık kaynak kodlu uygulamalar ve ücretsiz hizmetlerden yararlanılmıştır.

Bu çalışmada, OKT tekniği kullanılarak internetten canlı yayınlanan videolardaki konuşmaların metne dönüştürülmesi sayesinde videonun alt yazı ile birlikte eşzamanlı sunumu anlatılmıştır. Konuşma tanınmanın yapılabilmesi için W3C-SACG tarafından geliştirilen ve tüm kullanıcılara açık WSA’dan yararlanılmıştır. Web uygulamasında, canlı alt yazıların işlenmesi, sunucuya gönderilmesi ve kullanıcının video ile alt yazıyı eşzamanlı seyredebilmesi için iki farklı sistem tasarlanmıştır. Uygulamada, canlı video yayını bir medya sunucusu yardımı ile yapılarak, web sayfasında gösterimi HTML5 dilinin olarak sağladığı video elementi ile sağlanmaktadır. Alt yazı ve currentTime’in oluşturulması aşaması jQuery kütüphanesi ve JavaScript diliyle, veri tabanına gönderilmesi ise PHP dili ile gerçekleştirilmiştir. Metin ve zaman aralıklarının veri tabanına yazdırma işlemi, sonradan tekrar yayınlanacak videolar için alt yazıların kullanıma hazır halde olmasını sağlamaktadır.

OKT teknolojisi ile yapılan canlı videoları eşzamanlı alt yazılı sunma uygulamasında yapılan deneme yayınları istenilen seviyede olmuştur. Fakat bu uygulamanın veriminin insan ve teknik bazda incelenmesi uygulamanın geliştirilmesi için önem taşımaktadır.

İnsan kaynaklı olarak; konuşma tanıma teknolojisinin dezavantajlarından olan, söylenen kelimelerin vurgu ve ses şiddeti, ayrıca konuşanın mikrofona olan uzaklığı veya ortam gürültüsü sonuçların hatalı dönüşüyle sonuçlanabilmektedir. Bunun yanında, hızlı veya kesintisiz konuşma yine sonuçları etkilemektedir. Dünyada OKT teknolojisinin halen gelişmekte olduğu ve diğer OKT uygulamalarında da aynı sorunların olduğu düşünüldüğünde, WSA kullanımında sayılan faktörlerin en aza indirilmesiyle en verimli sonucun elde edileceği düşünülmektedir.

Teknik bazda ise tecrübeler, internet ve sunucu işlemci hızının alt yazıların gösterimini geciktirdiğini göstermiştir. Ayrıca, konuşmaların metne çevriminin uzak sunucu tarafından yapılıyor olmasından dolayı, uzak sunucu etkenli bazı gecikmelerin de olduğu görülmüştür. Bunun yanında, Google Chrome tarayıcısının uzun zaman diliminde kullanıldığında, tarayıcıda önbellek kaynaklı yavaşlamalar gözlenmiştir. Ek olarak, canlı alt yazı oluşturma işlemi sadece Chrome’da yapılabiliyor olmasına rağmen, oluşturulan web uygulaması kullanıcı ara yüzündeki video ve alt yazı akışı hemen hemen tüm tarayıcılarda görüntülenebilmektedir.

Bu web uygulaması halen geliştirme safhasında olduğundan, OKT’yi etkileyen faktörlerin asgari seviyeye indirilmesi için ileriki çalışmalarda araştırma ve testlerin farklı kullanıcılar ile birlikte yapılması (konuşma şekli, lehçe, ses tonu gibi testler), uygulamanın farklı ortamlar (ortam gürültüsü) veya senaryolarda (farklı bilgisayar sistemleri, internet kapasitesi, sunucu konfigürasyonları) denenmesi ve bunların irdelenmesi gerektiği düşünülmektedir.



**KAYNAKLAR (REFERENCES)**

- [1] Internet: Federal Communications Commission, Closed Captioning on Television, <http://www.fcc.gov/guides/closed-captioning>, 05.05.2014.
- [2] Internet: Federal Communications Commission, Captioning of Internet Video Programming, <http://www.fcc.gov/guides/captioning-internet-video-programming>, 05.05.2014.
- [3] Internet: CISCO, Cisco Visual Networking Index: Forecast and Methodology, 2013–2018, [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white\\_paper\\_c11-481360.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html), 18.05.2014.
- [4] Internet: K. Harrenstien, Finally Caption Playback, <http://googlevideo.blogspot.com.tr/2006/09/finally-caption-playback.html>, 10.05.2014.
- [5] Internet: K. Harrenstien, Automatic Captions in YouTube, <http://googleblog.blogspot.com.tr/2009/11/automatic-captions-in-youtube.html>, 10.05.2014.
- [6] Internet: Google, Automatic Captions, <https://support.google.com/youtube/answer/3038280>, 02.06.2014.
- [7] Internet: J. Siegel, YouTube is going LIVE, <http://youtube-global.blogspot.com.tr/2011/04/youtube-is-going-live.html>, 15.05.2014.
- [8] Internet: Google, Youtube Streaming Guide: Live Caption Requirements, <https://support.google.com/youtube/answer/3068031>, 02.06.2014.
- [9] C. Munteanu, G. Penn, R. Baecker, T. Zhang, "Automatic Speech Recognition for Webcasts: How Good is Good Enough and What to Do When it Isn't", **ICMI '06 8th International Conference on Multimodal Interfaces**, New York, USA, 39-42, 2006.
- [10] C. Aksoylar, S. O. Mutluergil, H. Erdogan, "Bir Türkçe Konuşma Tanıma Sisteminin Anatomisi", **Signal Processing and Communications Applications Conference**, Antalya, 512-515, 2009.
- [11] B. Shiver, "Utilizing automatic speech recognition to improve Deaf accessibility on the Web", **DePaul CDM School of Computing Research Symposium (SOCRS 2011)**, Chicago, 4-8, 2011.
- [12] A. Iglesias, L. Moreno, B. Ruiz, J. L. Pajares, J. Jiménez, J. F. López, P. Revuelta, J. Hernandez, "Web educational services for all: the APEINTA project", **8th International Cross-Disciplinary Conference on Web Accessibility**, New York, USA, 1-3, 2011.
- [13] M. Walde, "'SpeechText': Enhancing Learning and Teaching by Using Automatic Speech Recognition to Create Accessible Synchronised Multimedia", **ED-MEDIA 2005 World Conference on Educational Multimedia, Hypermedia & Telecommunications**, Montreal, 4765-4769, 2005.
- [14] S. Pfeiffer, **The Definitive Guide to HTML5 Video**, Apress, A.B.D., 2010.
- [15] B. Gold, N. Morgan, **Speech and Audio Signal Processing: Processing and Perception of Speech and Music**, John Wiley & Sons Inc., New York, 1999.
- [16] J. C. Junqua, **Robust speech recognition in embedded system and PC applications**, Kluwer Academic Publishers, Norwell, MA, A.B.D., 2000.
- [17] J. Holmes, W. Holmes, **Speech Synthesis and Recognition**, Taylor & Francis Inc., Bristol, PA, 2000.
- [18] M. A. Anusuya, S. K. Katti, "Speech Recognition by Machine: A Review", *International Journal of Computer Science and Information Security*, 6(3), 181-205, 2009.
- [19] M. Aniss, **How Does Voice Recognition Work?**, Gareth Stevens Publishing, New York, 2014.
- [20] D. C. Gibbon, Z. Liu, **Introduction to video search engines**. Springer, Berlin, 2008.
- [21] A. Choudhary, R. Kshirsagar, "Process Speech Recognition System using Artificial Intelligence Technique", *International Journal of Soft Computing and Engineering*, 2(5), 239-242, 2012.
- [22] A. Burton, "Do you understand me now? - An introduction to speech recognition", **10th Interactive Multimedia Conference**, Southampton, 2012.
- [23] W. Lasecki, C. Miller, A. Sadilek, A. Abumoussa, D. Borrello, R. Kushalnagar, J. Bigham, "Real-time captioning by groups of non-experts", **25th annual ACM symposium on User interface software and technology (UIST '12)**, New York, 23-34, 2012.
- [24] Internet: J. Adorf, Web Speech API, <http://home.in.tum.de/~adorf/>, 01.06.2014
- [25] Internet: W3C, Web Speech API Specification, <https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html>, 27.05.2014