

Medical Report Generation in a Structured and Interactive Way Using Speech Driven Approach

Kaya KURU^a, Kemal ARDA^b, Nazife BAYKAL^c

^a Bilgi İşlem Merkezi, Gülhane Askeri Tıp Akademisi, Ankara

^b Radyoloji AD, Yüksek İhtisas Hastanesi, Ankara

^c Enformatik Enstitüsü, Orta Doğu Teknik Üniversitesi, Ankara

Abstract

The primary objective of the present study is to develop a method, for capturing medical data in the field of medical reporting as an alternative interface additional to keyboard, mouse and transcriptionists. In this view, we built a pilot speech-interfaced application for medical reporting. In this study, we blend the advantages of isolated speech recognition (ISR) as being speaker-independent and continuous speech recognition (CSR) as being speaker-dependent together by restricting the sequence of words suitable to the fields presented by interactive style in a very large vocabulary lexicon by a language model. We tried to point out how semi-structured (the understanding of open-ended and close-ended conditions) and interactive medical reporting benefited from SR technology by using CSR and ISR respectively. Complete interactivity is provided with an interface to remove the look-away problem in radiology during the examination as the user hears questions (text-to-speech) and answers these questions with his/her voice (speech-to-text) by using a headphone in great concentration without the need to look at the monitor and return computer to enter data by using mouse and keyboard without losing hygienic environment.

Preliminary results have shown the success of SR technology in entering structured medical data with interactivity. The recognition process is done very fast and more accurately with the proposed and applied methodology when compared to the usage of user-dependent CSR in free text medical report generation. The same colon reports belonging to the same patients were generated by four different methods with the use of the same user and same-trained speech files and the recognition success of SR is increased with the method that we proposed. The medical reports can be prepared more accurately with the proposed method by preventing possible lacks by guiding radiologists through examination of the system; hygienic working environments can be preserved. Over and above, the recognition success of SR can be increased to one hundred percent in case of transforming the kinds of medical reports to exact structured formats, in other words, the answers to the asked questions are close-ended (interpretation is not permitted). Furthermore, medical data are collected in a structured format to be evaluated for further studies.

This study in radiology is significant as pointing out that a voice-enabled, structured, interactive reporting system can be used by giving the experts a means to produce medical reports quickly, efficiently, and much more cheaply than those, which are currently in widespread use as transcriptionists.

Key Words

Speech Recognition, interactive medical reporting, structured medical reporting

Konuşma Tanıma Sistemi Kullanılarak Bilgisayar ile Etkileşimli, Yapısal Tıbbi Rapor Hazırlanması

Özet

Bu çalışmanın asıl hedefi, tıbbi raporlama alanında, tıbbi veri elde etmek maksadı ile klavye, fare ve tıbbi sekreterlere ilave alternatif bir veri toplama metodu geliştirmektir. Bu kapsamda tıbbi raporlama için ses arayüzlü bir pilot uygulama geliştirdik. Bu çalışmada bir dil modeli ile kelime bilgisi çok büyük olan bir sözlük içerisinde etkileşimli olarak sistem tarafından sunulan konuya uygun daraltılmış kelime bilgisi kullanarak, kullanıcı bağımsız ayrık konuşma tanımının ve konuşmacı bağımlı sürekli konuşma tanımının

avantajlarını bir araya getirdik. Biz bu çalışmada, tıbbi rapor oluşturmanın, yarı yapılandırılmış (açık uçluluk ve kapalı uçluluk) ve etkileşimli bir anlayışla, ayırık ve sürekli konuşma tanımadan ne ölçüde faydalanabileceğini ortaya koymaya çalıştık. Radyolojide muayene sırasında, ihtiyaçtan, yapılan işten başka bir yere yönelme problemi, oluşturulmuş olan bir arayüz ile tamamıyla etkileşimli olarak ortadan kaldırılmıştır; kullanıcı yaptığı işe iyi bir şekilde konsantre olmuş bir durumda, fare ve klavye kullanarak veri girmek için bilgisayara dönmeden, monitörü görmeden ve çalıştığı hijyenik ortamdan ayrılmadan, cevaplama gereken soruları bir kulaklık vasıtası ile duymakta (yazıdan konuşmaya) ve bu soruları kendi sesini kullanıp (konuşmadan yazıya) cevaplayarak raporunu oluşturabilmektedir.

İlk sonuçlar, etkileşimli olarak, yapısal tıbbi veri girişlerinde konuşma tanıma teknolojilerinin çok başarılı bir şekilde kullanılabilirliğini göstermektedir. Önerilen ve tatbik edilen yöntemde, serbest metinle tıbbi rapor oluşturmada devamlı konuşma tanımının kullanılmasına nazaran, tanıma süreci çok hızlı gerçekleşmekte ve doğruluk oranı önemli ölçüde artırılabilir. Aynı kullanıcı ve aynı ses eğitim verilerinin kullanılması ile aynı hastalara ait aynı kolon raporları dört farklı yöntemle oluşturulmuş ve bizim önerdiğimiz metot ile konuşma tanımının başarısı yükseltilmiştir. Önerdiğimiz metodoloji ile radyoluğun, sistem tarafından doğru yönlendirilmesi ve girilmesi gereken rapor detaylarını unutmamasının önüne geçilmesi ile tıbbi raporlar çok daha doğru hazırlanabilmektedir; hijyenik çalışma ortamları korunabilmektedir. Bunun yanında, rapor türlerinin tam olarak yapısal bir hale dönüştürülebilmesinin sağlanması durumunda, bir diğer değişle, cevaplanması gereken sorulara verilecek cevaplar kapalı uçlu, çoktan seçmeli olması durumunda, kullanım öncesinde hiçbir ses eğitimi yapılmadan konuşma tanımının başarısı yüzde yüz olabilmektedir. Ayrıca tıbbi veriler ileride işlenebilirlik açısından yapısal bir formatta toplanabilmektedir.

Bu çalışma, radyolojide, ses kullanımlı, yapısal, etkileşimli bir tıbbi raporlama sisteminin, mevcut yaygın olan tıbbi sekreter kullanılmasına nazaran, uzmanların tıbbi raporları hızlıca, etkili, doğru ve çok daha maliyet etkin oluşturmalarını sağladığını göstermesi açısından önemlidir.

Anahtar Kelimeler

Konuşma tanıma, etkileşimli tıbbi raporlama, yapısal tıbbi raporlama

1. Giriş

Karşılıklı konuşmalı bilgisayar geliştirme çabası, son 30 yılın henüz tam olarak ulaşılamamış bir hedefi olarak görünmektedir. Bilgisayar teknolojisinde önemli gelişmelere rağmen, klavye ve fare halen vazgeçilemez bir veri giriş yöntemidir. Pek çok gelişmeler yaşanırken, konuşma tanımayı etkin bir şekilde kullanılmaktan alıkoyan bir takım kısıtlamalar bulunmaktadır. Konuşma tanıma teknolojisinin, ancak, zorlayıcı bazı sebepler olması durumunda kullanılması halinde işe yarayabileceği çok yaygın bir düşünce olarak göze çarpmaktadır. Muayene sırasında veya görüntü okuma sırasında ellerin ve gözlerin meşgul olduğu ve hareketli durumlar bu zorlayıcı sebeplere örnek olarak verilebilir. Konuşma tanımının yaygın olarak kullanılmasının önünde bulunan engellerin aşılması konusunda da çok fazla bir çaba gösterilmediği aşikar bir gerçek olarak önümüzde durmaktadır. Daha önce yapılan bir çalışma, geleneksel veri giriş tekniklerinden ziyade ses arayüzünün veri girişinde kullanılmasının, sağlık hizmeti verenlerin, tıbbi uygulamaları daha sıklıkla kullanacaklarını göstermektedir [1]. Ellerin ve gözlerin meşgul olduğu durumlarda, elde edilen bulguları, muayene sırasında veya muayene sonrasında bilgisayarlara girmek, herhalde, sağlık çalışanlarının maruz bırakıldıkları bir tür işkence olsa gerek.

Mevcut sürekli konuşma tanıma sistemleri, doğal konuşma hızında (dakikada 150 kelime) yüzde 95' in üzerinde olan yüksek bir doğruluk başarısı yakaladıklarını iddia etmektedirler [2]. Ama konuşma tanıma, konuşmacı bağımlılık, doğal konuşma hızı sürekliliği ve sahip olunan kelime hazinesi gibi bazı kısıtlamalara sahiptir. Konuşmacı bağımlı sistemlerde, her kullanıcının sesinin eğitilmesi zorunluluğu bulunmaktadır ve bu şekliyle her kullanıcının sesini tanımaya yönelik olan konuşmacı bağımsız sistemlerden daha yüksek bir başarı oranına ulaşılabilir. Sürekli konuşma tanıma sistemleri kelimeleri doğal konuşma ritminde tanıırken, ayırık konuşma tanıma sistemlerinde, kelimeler arasında belli duraklamalar yapılması zorunlu kılınmaktadır. Sürekli konuşma daha tercih edilir olmasına rağmen, konuşulan cümle içerisindeki kelime sınırlarını tespit etmedeki zorluklar sebebiyle işlem yapmak çok daha zordur. Kelime hazinesi 20 kelime ile 100.000' in üzerinde olabilen değişimler göstermektedir. Çok geniş kelime hazinesi tanıma başarısını düşürebilirken, dar kelime hazinesi ise

konuşma esnasında istenmeyen bir takım kısıtlamalar yapmayı zorunlu hale getirmektedir. Konuşma tanıma sistemleri yapılan en son çalışmalar ile, geniş kelime hazineleri, konuşmacı bağımsız, sürekli tanıma sistemlerine yönelme eğilimlerindedir. Tıp alanında bu tür değişiklikler ilk olarak radyoloji, patoloji ve endoskopi gibi alanlarda görülmektedir. Konuşma tanıma arayüzü ile bir doktor kendi raporunu direkt olarak bilgisayara girebilmekte ve bu şekilde tıbbi sekreterlerinin kullanılmasına gerek kalmamakta, bu durum da raporun kullanıma hazır hale gelmesini son derece hızlandırmaktadır. Konuşma tanıma kullanılarak serbest metin ile hazırlanan tıbbi raporlardaki kelimelerin değişik anlamlar içermesi ve doğal dil işleme (natural language processing) metotlarının henüz tam gelişmemiş olması sebebiyle bu raporlardaki bilgileri ölçmek ve analiz etmek pek mümkün olamamaktadır. Yapısal sesli veri girişi, bu problemlerin ortadan kaldırılmasına, bilginin kodlanmış ve ölçülebilir bir parçasını bir dizi içerisinde tanıyarak çözüm olmaktadır. Etkileşimli metot sesin kullanımındaki etkinliğini arttırmakta ve uygulamalarda standart ekranlar sağlamaktadır.

Ellerin meşgul olduğu durumlarda, serbest metinde, etkileşimli olmadan, konuşma tanımanın kullanılarak test edildiği bazı çalışmalar yapılmıştır. Bir grup, diş muayenesi sırasında, klinik verileri kaydetmek için konuşma tanımanın uygulanabilirliği üzerinde çalışmıştır [3]. Bu şekildeki sistemler, sonuçları kaydetmek için ihtiyaç duyulan diş asistanına olan ihtiyacı ortadan kaldırmaktadır. Bu çalışmalarda konuşma tanımanın daha yavaş olduğu görülmüştür. Ama asistan tarafından kaydedilen verilerin bilgisayar ortamına aktarılması da söz konusu olduğunda konuşma tanımanın daha hızlı olduğu görülmüştür. Farkın istatistiksel olarak önemli olmamasına rağmen, kullanılan konuşma tanıma sisteminde pek çok hata oluştuğu görülmüştür. Çalışma özetle konuşma tanıma sisteminin geleneksel yöntemlere bir alternatif olarak kullanılabileceğini tavsiye etmiştir. Bir başka grup, anestezi uzmanlarının ellerinin dolu olduğu durumlarda, anestezi kayıtlarının tutulduğu bir sistem için bir konuşma tanıma arayüzü geliştirmiştir [4]. Sistem 300 kelime kullanmış ve doğruluk oranı çalışma ortamının gürültülü olmasına rağmen yüzde 96 olarak görülmüştür. Konuşma tanımanın kullanıldığı bir diğer uygulama ise kemik sintigrafisi verilerinin analizi sırasında doktorlar terminale bakmak yerine görüntüleri incelerken elde etmiş oldukları verileri kendi seslerini kullanarak kaydetmişler ve 580 ses kayıt raporunda, rapor yazma süresinin, elle yazma veya tıbbi sekreter kullanılmasına nazaran daha kısa sürede tamamlanabildiği sonucuna varılmıştır.

Bu çalışmada tıbbi raporlama için ses arayüzlü bir pilot uygulama geliştirdik. Bizim geliştirdiğimiz uygulama, ilk aşamada radyoloji alanında, konuşma tanıma teknolojisinin bilgisayar ile tam etkileşimli kullanılması hedeflenmiş, konuşma tanıma becerisini zaman içerisinde kendiliğinden geliştirebilen bir tür öğrenen sistem olarak tanımlanabilir.

2. Gereç ve Yöntem

Tıbbi raporlama için geliştirdiğimiz ses arayüzlü uygulama nesne tabanlı modelleme ve üç katmanlı mimari (presentation layer, business layer, and data layer) ile VB.NET yazılım teknolojisi kullanılarak geliştirilmiş, veri tabanı Microsoft SQL Server'dır. Uygulama kullanıcıya klavye, fare veya konuşma tanıma arayüzünü kullanabilme seçenekleri sunmaktadır. Kullanıcılar bir modaliteden diğerine çok kolay bir şekilde geçebilmektedir. Uygulamada oluşturulmuş olan altyapı ile kullanıcılar kendi rapor formatlarını kendileri oluşturabilmekte ve bir bilgisayar uzmanına ihtiyaç kalmamaktadır.

Biz bu çalışmada, tıbbi rapor oluşturmanın, yarı yapılandırılmış (açık uçluluk ve kapalı uçluluk) ve etkileşimli bir anlayışla, ayırık ve sürekli konuşma tanımadan ne ölçüde faydalanabileceğini ortaya koymaya çalıştık. Öncelikle, daha önce pek çok farklı raporlama türü için oluşturulmuş olan 500.000 rapor sistem tarafından otomatik olarak okunup, farklı kelimeler tespit edilerek 100.000 kelimelik bir tıbbi sözlük oluşturuldu. Aynı zamanda sözlük içerisinde üçlü gramer yapıları da oluşturularak bir kelimenden sonra en sıklıkla gelebilecek kelimelerin olasılıkları da hesaplandı. Kullanıcı bir saat yirmi dakikalık bir ses eğitimine tabi tutularak, konuşma tanıma sırasında kullanılacak ses eğitim dosyaları oluşturuldu. Kullanıcı ses eğitimi ve raporlar oluşturma gürültünün olmadığı bir ortamda gerçekleştirildi.

Daha sonra, aynı konuşma tanıma algoritmaları ile, aynı kullanıcı, aynı eğitim verileri ve aynı sözlük kullanılarak, konuşma tanımanın başarısını dört farklı metodolojide değerlendirilmeye çalışıldı. Birinci ve ikinci metodolojide etkileşimli olmayan, kullanıcıdan bilgisayara tek taraflı (speech_to_text) ve serbest metin yazılan raporlar için konuşma tanımanın başarısı ölçülmeye çalışıldı. Ancak, farklı olarak birinci metodolojide sözlüğün tamamı tanıma esnasında kullanılırken, ikinci metodolojide rapor çeşidine göre uygun daraltılmış kelime hazinesi sisteme yüklenerek, tanıma bu kelime hazinesi kullanılarak yapıldı. Örnek hasta raporları hazırlanırken, tanınması gereken tüm kelimeler sözlük içerisine ve daraltılmış kelime hazinesine tanıtılmıştır. Üçüncü metodolojide yapısal ve bilgisayar ile etkileşimli bir tıbbi raporlama sistemi, konuşmadan yazıya (speech to text) ve yazıdan konuşmaya (text-to-speech) özelliklerinin beraber kullanıldığı bir konuşma tanıma arayüzü ile birleştirilmiştir ve konuşma tanımanın doğruluk oranı ölçülmüştür. Rapor formatı yapılandırılmıştır: ideal bir kolon raporunun sahip olması gereken bilgiler, 120 konuya ayrıldı ve bu konular cevaplanması gereken 120 soruya dönüştürüldü. Bu soruların 64 tanesi çoktan seçmeli (kapalı uçlu), 56 tanesi de açık uçlu (yorumaya açık) şeklinde oluşturulmuştur. Spektrum yanlılığı (spectrum bias) ortadan kaldırmak maksadı ile belli yaş grup aralıklarına ve cinsiyete göre normal değerler tanımlanmış ve soruların sorunun cevap kısmında bu normal değer getirilmiştir. Rapor etkileşimli olarak oluşturulmakta; sistem sesli olarak bir soru sormakta (text_to_speech) ve kullanıcıda bu soruyu cevaplamaktadır (speech_to_text). Sorunun cevaplanmasının ardından bir sonraki soruya kullanıcının sesli işaretiyle geçilmekte ve tüm rapor bu şekilde doldurulmaktadır. Yine konuşma kullanılarak daha önce cevaplanmış olan bir soruya değişiklik yapılabilmesi için geçilebilmektedir. Rapor soruları iki parçanın bir araya getirilmesinden oluşmaktadır. Birinci parça sorulmak istenen konu, ikinci parça ise soru kalıbından oluşmaktadır. Örneğin sorulmak istenen konu “Rektum segmentinin çapı”, soru kalıbı ise “kaç cm dir” bir araya getirilerek “Rektum segmentinin çapı kaç cm dir” sorusu elde edilir. Kullanıcı cevabını verdiğinde ise konu ile cevap bir araya getirilerek, oluşturulan rapor yapısal formatta yazdırılabildiği gibi klasik düz metin olarak da yazdırılabilmektedir. Ayrıca kullanıcı soruları cevaplarırken daha önce oluşturduğu kısımları da bir bütün olarak ekranın alt tarafından klasik metin olarak takip edebilmektedir. Radyoloji raporlaması üzerine yapılan bazı çalışmalar, kavramsal psikolojik ve sosyolojik olarak iletişimin klasik metin ile yapılmasını en etkili yollardan biri olduğuna işaret etmektedirler [5, 6, 7]. Çok aşırı derecede yapılandırılmış raporların veri girişi ve incelenmesi esnasında kavramsal odaklanmada kayıplara sebep olacağı da bir başka çalışmada ifade edilmiştir [8].

Dördüncü metodolojide is 120 konuluk soru 40 ana soru ve 80 alt soruya ayrıldı ve soru sorma dinamik bir hale dönüştürüldü. Ana soruya ait bir anormal değer olduğu takdirde, şayet o soruya ait bir alt soru var ise o soru getirildi, yok ise bir sonraki ana soruya geçildi. Şayet muayene sırasında herhangi bir problemle karşılaşılma ise 40 ana soru cevaplanarak rapor oluşturma tamamlandı. Üçüncü ve dördüncü metodolojide normal değer dışında bir değer oluşmadı ise kullanıcı cevap kısmına bir değer girmeden sesiyle bir sonraki soruya geçmekte ve raporu tamamlamaktadır. Bu durumda sistem dinamik olarak yaş grubuna ve cinsiyete göre cevap kısmını kendisi doldurmaktadır. Tüm metodolojide raporlar oluşturulurken sorulara cevaplanırken iki yöntem izlenmiştir: birinci yöntemde hep olumsuz, yani normal değer dışında olan bir cevap verilmiş; ikinci yöntemde is hep olumlu cevaplar verilmiştir.

Kullanıcının konuşmasının enerjisi belli bir seviyenin (dB) üzerindeyse algılanabilmekte ve buna göre bir eşik değeri sisteme tanımlanabilmektedir. Bu eşik değeri kullanıcının ses tonunu alabilecek kadar düşük ve çevredeki gürültüleri alamayacak kadar büyük olmalıdır.

3. Bulgular

Şüphesiz ki, kullanıcı karşılaştığı konuşma tanıma hatalarını düzelttikçe ve sisteme doğru hallerini tanıttıkça, sistemin bu hataları bir daha yapma olasılığının düşecek olmasından dolayı konuşma tanıma başarısı daha yükselecektir. Tanımda meydana gelen hatalar düzeltilmeden dört farklı yöntemdeki başarı hesaplanmıştır. Birinci metot kullanılarak, üzerinde çalıştığımız kolon raporunda 10 hasta için düz metinde yapılan rapor oluşturma ile başarı, yüzde 81 olmuştur. İkinci metot kullanılarak, daraltılmış

kolon raporu kelime bilgisi tanım kümesinin sisteme yüklenerek kullanılması ile başarı yüzde 90'a arttırılmıştır. Bu esnada sürekli konuşma tanıma kullanılmış ve iki taraflı bir etkileşim kullanılmamış, sadece konuşma tanımının konuşmayı metine dönüştürme fonksiyonu kullanılmıştır. Ayrıca kullanıcı bilgisayara bilgi girebilmek için yaptığı işten (görüntü okuma veya hasta muayenesi) ayrılarak monitörün karşısına geçmek durumunda kalmıştır.

Üçüncü metot kullanılarak rapor formatının yapısal bir hale dönüştürülmesi ile elde edilen sonuçlar tanıma başarısını yüzde 98 olarak önemli ölçüde yükseltmiştir. Kapalı uçlu, çoktan seçmeli sorularda hata hiç oluşmazken, hatalar açık uçlu sorularda meydana gelmiştir. Radyolojide muayene sırasında, ihtiyaçtan, yapılan işten başka bir yere yönelme problemi, oluşturulmuş olan bir arayüz ile tamamıyla etkileşimli olarak ortadan kaldırılmıştır; kullanıcı yaptığı işe iyi bir şekilde konsantre olmuş bir durumda, fare ve klavye kullanarak veri girmek için bilgisayara dönmeden, monitörü görmeden ve çalıştığı hijyenik ortamdan ayrılmadan, cevaplaması gereken soruları bir kulaklık vasıtası ile duymakta (yazıdan konuşmaya) ve bu soruları kendi sesini kullanıp (konuşmadan yazıya) cevaplayarak raporunu oluşturabilmektedir.

Dördüncü metot da ise, dinamik bir anlayışla, tüm sorulara normal değerlerin dışında bir cevap verildiğinde yine üçüncü metot da olduğu gibi 120 soru cevaplanmak zorunda kalınmıştır. Rapor oluşturma zamanı açısından önemli bir fark oluşmamıştır. Üçüncü ve dördüncü metotlarda cevapların normal değerlerde olduğu durumlarda raporlar %100 doğrulukla oluşturulmuşlar. Fakat tüm sorulara olumlu cevaplar verildiğinde rapor süresi 6 da bir oranda düşerek 5 dakika olmuştur. Muayene sırasında cevaplanacak soruların tamamının olumlu (normal değerler içinde) veya tamamının olumsuz (normal değerler dışında) olmasının pek sıklıkla karşılaşılan bir durum olmaması göz önüne alındığı taktirde, dördüncü metodun kullanılması üçüncü metodun kullanılmasına nazaran maksimum 6 kata kadar bir zaman tasarrufu sağlayabileceği söylenebilir. Bizim belirlediğimiz bilgileri içeren bir kolon raporunun klasik yöntemlerle yazılmasının ortalama 20-30 dakika arasında bir zaman alabileceği göz önüne alındığında, dördüncü metodun kullanıldığı sistemin, klasik yöntemlere göre zaman açısından sağlayabileceği olumlu katkı ortaya çıkmaktadır.

Bir konuşma tanıma arayüzüne sahip bir sistemin kullanılmasında karşılaşılan en önemli problemlerden ve en sıkıcı aşamalarından biri kullanıcı sesinin sisteme tanıtılması maksadı ile eğitilmesi zorunluluğudur. Yoğun çalışma ortamlarında bulunan doktorların eğitim için zaman ayırmaları kimi zaman zor olabilmektedir. Rapor türlerinin tam olarak yapısal bir hale dönüştürülebilmesinin sağlanması durumunda, bir diğer değişle, cevaplanması gereken sorulara verilecek cevaplar kapalı uçlu, çoktan seçmeli olması durumunda, kullanım öncesinde hiçbir ses eğitimi yapılmadan konuşma tanımının başarıları yüzde yüz olabilmektedir. Ayrıca tıbbi veriler ileride bilimsel işlenebilirlik açısından tam yapısal bir formatta toplanabilmektedir. Bu durumda kullanıcı sadece ayırık konuşma tanımayı kullanarak açılan menüden uygun olan cevabı seçebilmektedir. Yaptığımız çalışmalar, mevcut ayırık konuşma tanıma algoritmalarını kullanarak menü içerisinde bulunan 1000 farklı değerden yapılan seçimlerin yüzde yüz doğrulukla olabileceğini göstermektedir. 1000-1500 arasında başarı yüzdesi ortalama yüzde 99 civarında, 1500-2000 arası yüzde 98 civarında olmaktadır. Bir soruya verilecek cevapların 1000 farklı değeri geçemeyeceği düşüncesinden hareketle, ayırık konuşma tanımının tam olarak yapısal hale dönüştürülmüş rapor türlerinde yüzde yüz doğrulukla çalışabileceğini söyleyebilmekteyiz. Ayırık konuşma tanımının kullanıldığı durumda uygulama konuşmaları bir dilde konuşmalar nasıl seslendiriliyor ise anlamakta ve konuşulunun karşılığını menüde bulunan seçeneklerle karşılaştırmakta ve uygun olanı ile eşleştirmektedir.

4. Tartışma

İnsanlar arasındaki iletişimde bile, zaman zaman yanlış anlamalar olabilmekteyken, mevcut çalışmadaki konuşma tanımının yüzde 98 başarıları, rapor yazmada konuşma tanıma teknolojinin bize sunduğu önemli bir fırsat olarak görülmelidir.

Konuşma tanıma arayüzünün tıbbi raporlama esnasında çok yaygın olarak kullanılmamasının bir sebebi de henüz standart hale getirilememiş rapor formatları olarak gösterilebilir. Halihazırdaki teknoloji bize mevcut durumdan çok daha fazla hizmet etmeye hazır durumdadır. Bu çalışma, rapor formatlarının tam olarak standart bir hale getirilmesi ile konuşma tanıma arayüzünün çok etkile bir şekilde kullanılabilmesini göstermiştir. Bu çalışmada, etraftaki gürültüyü konuşma sinyalinden ayırt edebilecek bir algoritma kullanılmamıştır. Konuşma tanımadaki tanıma başarısının yüksek olabilmesi için ortamın çok sessiz olması gerekmektedir. Bu türlü uygulamaların hastane ortamlarında yaygın olarak kullanılabilmesi için çok gürültülü hastane ortamlarında konuşmacının sesini, çevreden gelen gürültüden çok iyi bir şekilde ayırt edilebilmesi bir zorunluluk olarak karşımızda çıkmaktadır. Son zamanlarda bu işi yapabilen çift mikrofona konuşma tanıma cihazlarının kullanılması konuşmacı sesini, çevre gürültüsünden çok iyi bir şekilde ayırt edebilecek durumda değerlendirilmektedir. Bu cihazlar, sesin aldığı mesafede, karesi oranında etkisini yitireceği prensibinden hareketle çalışmaktadır. Biz de geliştirmiş olduğumuz uygulamayı bu teknoloji ile entegre etmeye çalışmaktayız.

Konuşma tanıma sisteminin yaygın bir şekilde kullanılabilme başarısı ancak bu teknolojiye ihtiyaç duyan bir pazarın mevcut olması ile doğru orantılıdır. Nitekim böyle bir pazar da mevcuttur: bazı insanlar tekrarlı germe hasarları (repetitive strain injuries), kas yetersizliği (muscular dystrophy) ve benzeri rahatsızlıklar sebebiyle yazı yazmakta bir takım fiziksel sıkıntılar içerisindeyler [9]. Konuşma tanımalı arayüzler pek çok teknoloji rahatsızlığının önüne de geçmektedir.

Halen rapor formatları konusunda ulusal ve uluslar arası bir takım standartların geliştirilmesine ihtiyaç duyulmaktadır. Bu şekilde bir standartlaşma araştırma bulgularının karşılaştırılmasına, sağlık birimleri arasında verilen hizmet kalitelerinin karşılaştırılmasına, daha sağlıklı kişisel ve toplumsal sağlık verilerinin oluşturulmasını sağlayabilecektir.

Geliştirilmiş olan uygulamanın standart hale getirilmiş ekranlarında kullanıcı arayüzünün işlevselliğinin artırılması için halen bazı çalışmaların yapılmasına ihtiyaç bulunmaktadır.

5. Sonuç

Bu çalışma, ses kullanımlı, yapısal, etkileşimli bir tıbbi raporlama sisteminin, mevcut yaygın olan tıbbi sekreter kullanımına nazaran, uzmanların tıbbi raporları hızlıca, etkili, doğru ve çok daha maliyet etkin oluşturmalarını sağladığını göstermesi açısından önemlidir. Ayrıca, doktorlar bir tıbbi sekretere ihtiyaç duymadan kendi raporlarını oluşturabilmeleri sebebiyle doktorlar ve hastalar arasındaki kişisel gizlilik ilişkisi tam olarak korunabilmektedir.

Konuşma tanıma arayüzünün kullanılması ile metinler klavye kullanımına nazaran daha hızlı bir şekilde oluşturulabilmektedir. En iyi daktilo kullanan kişi, dakikada 50 ile 90 kelime arasında üretebilirken, konuşma tanıma arayüzü ile dakikada 100 ile 180 kelime arasında, metin iki katı bir hız ile ve üstelik çok az veya hiçbir fiziksel çaba olmadan yapılabilmektedir [10]. İki taraflı etkileşimli konuşma tanıma arayüzünün kullanılması ile (1) hastalar ve bilgisayarlar arasındaki gelgitler ortadan kaldırılarak rapor üretim hızı artırılmıştır; (2) Tıbbi sekreterlere veya klavye kullanımına olan ihtiyaç ortadan kaldırılmıştır; (3) bilgisayar kullanımı ile ortaya çıkabilen bazı hastalıkları (repetitive strain injuries vb.) azaltılmıştır; (4) Rapor türünün tam olarak yapısal bir hale getirilmesi ile ve ayrık konuşma tanımının kullanılması ile kullanıcı kullanımı öncesinde yapılması gereken eğitim ihtiyacı ortadan kaldırılabilmesi görülmüştür. (5) Büyük bir oranda yapısal bir hale dönüştürülen raporlarda ayrık konuşma tanıma ve sürekli konuşma tanımının kullanılması tanıma başarısını önemli ölçüde arttırmıştır.

Geliştirilmiş olan mevcut sistemi, hastane bilgi sisteminin (HBS) veya PACS sisteminin bir parçası haline getirmek ve bu şekilde kullanımını yaygınlaştırmak sonraki hedef olarak gösterilebilir. Konuşma tanıma teknolojilerinin tıbbi uygulamalar ile entegre edilmesi rapor oluşturmaya önemli katkılar sağlayabilmekte ve genel çalışma verimliliğinin de artırılmasını sağlamaktadır. Doktorlar, monitöre bakmak yerine, görüntüleri incelerken veya hastaları muayene ederken elde edilen bulguların sonuçlarını buldukları çalışma ortamından ayrılmadan bilgisayara girebilmektedirler. Konuşma tanıma ile sağlanan bilgisayar ile iki taraflı etkileşim verilerin toplanması esnasında ellerin ve gözlerin yapılan işten ayrılmasına mani olmakta ve hijyenik çalışma ortamları da korunabilmektedir.

Bu çalışma, yapılan literatür arařtırmalarında, yapısal ve bilgisayar ile etkileşimli bir tıbbi raporlama sisteminin, konuşmadan yazıya (speech to text) ve yazıdan konuşmaya (text-to-speech) özelliklerinin beraber kullanıldığı bir konuşma tanıma arayüzü ile birleştirildiđi, Türkiye’de ilk ve tek, dünyada ise nadir çalışmalarından biri olarak değeriendirilebilir.

6. Kaynakça

- [1] Feldman CA, Stevens D. Pilot study on the feasibility of a computerized speech recognition charting system. *Community Dent Oral Epidemiol.* 1990; 18:213-5.
- [2] Zafar A, Overhage JM, Clement J, McDonald CJ. Continuous Speech Recognition for Clinicians. *J Am Med Inform Assoc.* 1999; 6(3): 195–204.
- [3] Feldman, C.A., Stevens, D. Pilot Study on the Feasibility of a Computerized Speech Recognition Charting System. *Community Dentistry and Oral Epidemiology*, 1990; 18:213-215.
- [4] Smith NT, Brian RA, Pettus DC, Jones BR, Quinn ML, Sarnat L. Recognition Accuracy with a Voice-Recognition System Designed for Anesthesia Record Keeping. *J Clin Monit* 1990; 6(4):299-306.
- [5] Siström CL. Conceptual Approach for the Design of Radiology Reporting Interfaces: The Talking Template. *Journet of Digital Imaging* 2005 18: 176-187.
- [6] Garrod S. How groups co-ordinate their concepts and terminology: implications for medical informatics. *Methods Inf Med* 1998; 37: 471-476.
- [7] Cimino JJ, Patel VL, Kushniruk AW. Studying the human–computer-terminology interface. *J Am Med Inform Assoc* 2001; 8: 163-173.
- [8] Patel VL, Kaufman DR. Medical informatics and the science of cognition. *J Am Med Inform Assoc* 1998; 5: 493-502.
- [9] Connors K. Speech Recognition. www.hditech.com/files/Fusion_Speech.pdf 30.09.2007
- [10] Colton D. Automatic Speech Recognition Tutorial. <http://colton.byuh.edu/courses/cs441/> 30.09.2007
- [11] Gul, H., Kuru, K. and Gursel, G. (2006). A Versatile, User Driven, Flexible And Scalable Decision Making Tool In Toxicology. In: 3rd National Conference on Medical Informatics, November, 16-19, 2006, Belek-ANTALYA.
- [12] Gul, H., Kuru, K., Gursel, G., Yildiz, O (2005). The Advantages of Electronic Prescription, the Problems and Tackling Them. 2nd National Conference on Medical Informatics, November, 17-20, Belek-ANTALYA.
- [13] Kuru, K., Gul, H., Gursel, G., Arda, K. and Mumcuoglu, E. (2005). Analysis of Resources in Healthcare by Computer Simulation Studies in Healthcare: An Outpatient Clinic Study. In: 2nd National Conference on Medical Informatics, November,17-20, 2005, Belek-ANTALYA.
- [14] Kuru, K., Celasun, B. and Mumcuoglu, E. (2005). The Use of Lab Color Model for Sharpening and Optimization of Digital Images of Hematoxylin & Eosin Stained Microscopical Materials. In: International Symposium on Health Informatics and Bioinformatics / HIBIT, 2005, Belek-ANTALYA.

7. Sorumlu yazarın adresi

Kaya Kuru, MEBS Şube Müdürlüğü, Gülhane Askeri Tıp Akademisi, Ankara, Posta Kodu:06010. Tel:03123042771.

E-posta: kkuru2002@yahoo.com