

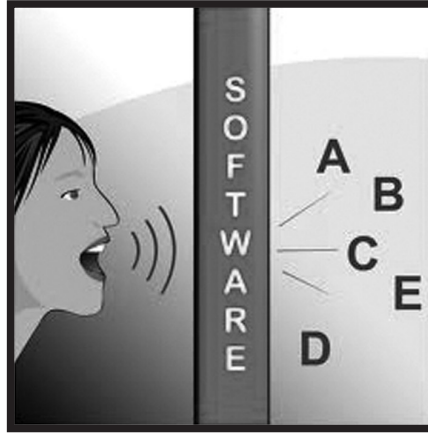
# Konuşma Tanıma Sistemleri

Behnam ASEFİSARAY

Bilgisayar Mühendisi  
bh.asefi@gmail.com

Günümüzde Konuşma Tanıma Sistemleri (KTS) birçok alanda kullanılmaya başlamıştır. Genel olarak bir KTS'nin görevi bir insana ait ses sinyalini alıp ve bu sinyal üzerinde bir takım işlemler yapmaktan ibarettir. Yapılan işlemlerin sonucunda ses sinyalinin hangi kelimelere karşılık geldiği bulunup yazıya dökülmektedir. Demek ki KTS, girdi olarak bir konuşma sinyali alıp çıktı olarak ise metin üretmektedir. Ancak yüzde yüz başarılı ve bir dile ait tüm kelimeleri tanıyan bir KTS'nin yapılması oldukça zordur.

KTS'ler birçok alanda yaygın olarak kullanılmakta olup ve daha hızlı yazmaya olanak sağlamaktadır. Örneğin dikte sistemleri, komut kontrol sistemleri, haber yazılandırma sistemleri vb. gibi teknolojilerde KTS'ler kullanılmaktadır. En çok kullanılmakta olan ve özellikle İngilizce dili üzerinde çok daha başarılı olan bu sistemlerden birisi dikte sistemleridir. Bu sistemler bilgisayar ortamında daha hızlı yazmak için kullanılmaktadır. Dikte sistemlerinde kullanıcı klavye kullanmadan sadece ne yazmak istediğini söyleyerek istediği kelimeleri yazabilmektedir. Şekil 1'de genel olarak bir KTS'nin görevi yer almaktadır. Ayrıca bu sistemler engelli insanlar tarafından çok kolay bir şekilde kullanılıp ve istedikleri kelimeleri sadece söyleyerek yazabilmektedirler.



Şekil 1

KTS'leri tanıdıkları kelime sayısına göre (sözlüklerinde bulunan kelime sayısı) üç farklı kategoriye ayrılmaktadırlar. Az sayıda kelime sözlüğüne sahip olanlar (küçük KTS), orta sayıda kelime sözlüğüne sahip (orta KTS) ve çok sayıda kelime sözlüğüne sahip (büyük KTS) olan KTS'ler.

Küçük KTS'lerin sözlüğünde bulunan ve sistemin tanıyabileceği kelime sayısı 50-100 arası olup ve daha çok komut kontrol sistemlerinde kullanılmaktadır. Bu tür sistemlerin küçük sözlüğe sahip olduklarından dolayı başarı oranları oldukça yüksektir. Bu sistemler için örnek verecek olursak; arabalarda kullanılan komut kontrol sistemleri, robotlarda kullanılan sistemler ve cep telefonlarında kullanılan sistemler bu kategoride yer almaktadır.

Orta KTS'lerin sözlüğünde bulunan kelime sayısı 100-1000 arası olup ve sınırlı sayıda kelime içeren dikte sistemlerinde kullanılmaktadır. Bu tür sistemler için örnek verecek olursak; radyologların rapor üretmek için kullandıkları KTS'ler bu kategoride yer almaktadır.

Büyük KTS'lerin sözlüğünde bulunan kelime sayısı ise 1000'den fazla olup bir dile ait genel dikte sistemlerinde (bir dilde bulunan tüm kelimeleri tanıyabilecek bir sistem) kullanılmaktadır. Ayrıca KTS'ler konuşmacı sayısına bağlı olarak iki farklı kategoriye ayrılmaktadır. Bunlardan birincisi tek bir konuşmacıyı tanıyan ve diğeri ise çok sayıda konuşmacıyı tanıyan sistemlerdir. Tek bir konuşmacıyı tanıyan sistemler sadece bir kişinin sesi için eğitilmiştir. Bu sistemlerin başarı oranı genel olarak oldukça yüksektir. Diğer taraftan konuşmacıdan bağımsız KTS'ler ise çok sayıda konuşmacının sesini tanımaya olanak sağlamaktadır. Konuşmacıdan bağımsız KTS'lerin yapılması oldukça zor olup sınırsız sayıda insan sesini sisteme eğitilmesi gerekmektedir.

İngilizce dili üzerinde yapılan KTS'ler günümüzün en başarılı ve yüksek doğruluk oranına sahip sistemleridir. Bunun nedeni ise İngilizce dili üzerinde daha çok bilimsel araştırmanın yapıpı yatırım yapılmasıdır. Son yıllarda Türkçe KTS'ler üzerinde bilimsel araştırma ve çalışmaların sayısı artış göstermektedir. Türkçe, KTS'lerin yapılmasında bu dilin kendine özgü avantajları ve zorlukları barındırmaktadır. Türkçe'de yazılan her kelimenin yazıldığı gibi okunması, KTS'lerin yapılması için bir avantaj sayılır. Ancak Türkçe'de bulunan kelime sayısının

teorik olarak sonsuz olması ve sondan çekimli bir dil olması ise KTS'ler için zor bir problemdir. KTS'lerin mimarisinde bulunan bileşenler genel olarak aynıdır. Demek ki farklı diller için yapılan KTS'lerin mimarisi ve bunlarda kullanılmakta olan modüller aynıdır. Ancak farklı diller için o dile ait modellerin eğitilmesi gerekmektedir. Başarılı ve konuşmacıdan bağımsız bir KTS'nin yapılabilmesi için sonsuz sayıda insanın sesinin sisteme eğitilmesi gerekmektedir. Dolayısıyla kadın, erkek ve farklı yaş aralıklarında ve çeşitli lehçelerde konuşan insanların sesi kayıt edilip sisteme eğitilmesi gerekmektedir.

Başarılı ve konuşmacıdan bağımsız KTS'ler çok büyük ve zengin ses derlemlerini kullanarak eğitilmektedirler. Eğitilen derlemin boyutu çok önemli olup saat biriminden ölçülmektedir. Örneğin; İngilizce için hazırlanan derlemlerin boyutu 2000 saatten fazla olup binlerce farklı kişinin sesi kayıt edilip sisteme eğitilmektedir. Hâlbuki Türkçe için hazırlanan ve mevcut olan ses derlemlerin boyutu çok daha küçüktür. Şu an Türkçe için mevcut olan ses derlemlerin toplam boyutu 200 saatten azdır. Dolayısıyla Türkçe için başarılı ve konuşmacıdan bağımsız bir KTS'nin yapılabilmesi için oldukça zengin ve büyük bir ses derleminin oluşturulması gerekmektedir.