

Konuşmacı Kimliği Algılanmasında Akustik Özniteliklerin Karşılaştırmalı Analizi

Oytun Türk

Levent M. Arslan

Elektrik ve Elektronik Mühendisliği Bölümü, Boğaziçi Üniversitesi, İstanbul
Sestek A.Ş. AR-GE Bölümü, İstanbul
oytun@sestek.com.tr arslanle@boun.edu.tr

Özetçe

Bu çalışmada farklı akustik özniteliklerin konuşmacı kimliği algılanmasındaki önemleri özlü deneyler yoluyla karşılaştırılmaktadır. İki özlü dinleme deneyi gerçekleştirilmiştir. Birinci deneyde ses borusu, ses perdesi, süre ve enerji özniteliklerinin konuşmacı kimliği algılanmasındaki önemleri ölçülmüştür. Bu dört akustik özneliğin tüm kombinasyonları bir konuşmacıdan diğerine nakledilmiştir. Ses borusu + ses perdesi + süre nakliyle hedeflenen konuşmacıya özlü olarak en benzer çıktı elde edilmiştir. Ses borusu + ses perdesi, ses borusu + süre + enerji ve ses borusu + süre nakilleri de çıktının hedef konuşmacıya benzerliği yönünden başarılı sonuçlar vermiştir. Konuşmacı kimliğinin algılanması açısından en önemli özneliğin ses borusu yapısı olduğu belirlenmiştir. Ses perdesi, süre ve enerji öznitelikleri sırasıyla azalan öneme sahiptir. İkinci deneyde farklı frekans aralıklarındaki işaret içeriklerinin konuşmacı kimliği algılanmasındaki önemleri karşılaştırılmıştır. Bu amaçla, farklı frekans bantlarına ayrıştırılmış ses kayıtları kullanılmıştır. Deneklerin bu frekans aralıklarından konuşmacı kimliğini ne kadar doğrulukla tanıyabildikleri ölçülmüştür. Sonuçlar, en önemli frekans aralığının 1.0-1.8 KHz olduğunu göstermektedir.

1. Giriş

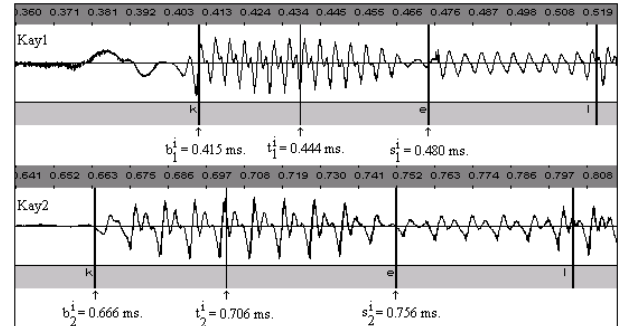
Konuşma işaretlerinden konuşmacı kimliğinin algılanmasında sosyal ve fizyolojik etkenler önem taşımaktadır. Sosyal etkenler, konuşma işaretinin zamanlama, vurgu ve aksan özelliklerini belirlemektedir. Fizyolojik etkenlerse spektral öznitelikler, ses tonu ve ses kalitesini etkilemektedir. Temel frekans (f_0) değeri, formant frekansları, f_0 'ın değişimi ve kaynak sinyalinin spektral eğiminin önemi [1]'de, spektral zarfın önemi [2]'de gösterilmiştir. Konuşmacı kimliğinin algılanması için çeşitli nesnel ve özlü ölçütlerin karşılaştırılmasında medyan f_0 değeri, ses borusu yapısı, kaynak sinyali ve vurgu özniteliklerinin önemli olduğu gözlemlenmiştir [3]. Furu, ortalama spektrum ve ortalama f_0 değerlerinin konuşmacılar arasındaki farkları açıklamada en önemli etkenler olduklarını göstermiştir [4].

Konuşmacı kimliği algılanması konusundaki çalışmalar çeşitli uygulamalar açısından önemlidir. Konuşmacı dönüştürmede kaynak konuşmacının sesi hedef konuşmacının sesine otomatik olarak dönüştürülmektedir [5]. Konuşmacı kimliği açısından önemli özniteliklerin dönüştürülmesi performansı artırmaktadır [6]. Konuşma tanıma uygulamalarında çeşitli uyarlama teknikleriyle akustik öznitelikler değiştirilmekte ve tanıma performansı artırılmaktadır. Konuşmacıya özgü nitelikler konuşmacı tanıma sistemlerinin tasarımında da göz önünde bulundurulmaktadır [7].

Bu çalışma, konuşmacı kimliği algılanmasında önemli etkenlerin araştırıldığı iki özlü dinleme deneyinden oluşmaktadır. 2. bölümde ses borusu, ses perdesi, süre ve enerji özniteliklerinin konuşmacılar arasında nakledilmesi için Perde Senkronize Üstüste Ekleme Yöntemi (PSOLA-Pitch Synchronous Overlap-Add) kullanılmıştır. Bu özniteliklerin tüm kombinasyonları farklı konuşmacılar arasında nakledilmiş ve çıktılar deneklere dinletilerek konuşmacıların tanınabilirliği ölçülmüştür. 3. bölümde farklı frekans aralıklarının konuşmacı kimliğinin algılanmasındaki önemleri özlü bir dinleme deneyiyle ölçülmüştür. Farklı kişilerden alınan ses kayıtlarında vurgu, PSOLA yöntemiyle normalize edilmiş ve bant-geçiren bir süzgeç bankasıyla frekans bantlarına ayrıştırılmıştır. Çıktılar özgün kayıtlarla deneklere dinletilmiş ve konuşmacıların tanınabilirliği ölçülmüştür. 4. bölümde sonuçlar tartışılmaktadır.

2. Deney 1: Akustik Öznitelik Nakilleri

Bu bölümde, bir konuşmacıdan diğerine akustik öznitelik nakilleri için kullanılan PSOLA yöntemleri açıklanmaktadır. Nakiller için kaynak konuşmacı (Kon1) ve hedef konuşmacının (Kon2) aynı metine karşılık gelen ses kayıtları (Kay1 ve Kay2) kullanılmıştır. Deneylerde kullanılmak üzere Kay2'den Kay1'e dört farklı akustik öznitelik nakledilmiştir: ses borusu yapısı, ses perdesi, süre ve enerji. Kay1 ve Kay2 fonetik olarak etiketlenmiştir. Kay1'deki her bir fonemin akustik öznitelikleri Kay2'de o foneme karşılık gelen özniteliklere dönüştürülmüştür. Dönüşüm sırasında Tablo 1'de gösterilen olası bütün öznitelik kombinasyonları ($2^4=16$ kombinasyon) denenmiştir. Bu kombinasyonların ikisi özgün cümlelere karşılık gelmektedir. Tablo 1'de bu ikili "A" ve "a" ile belirtilmiştir. Diğer 14 kombinasyonda, aynı harfle gösterilen nakillerin birinde Kay1'den alınan öznitelikler (örneğin "B") diğerinde ("b") Kay2'den alınmıştır.



Şekil 1: Birbirine karşılık gelen zamanların hesaplanması.

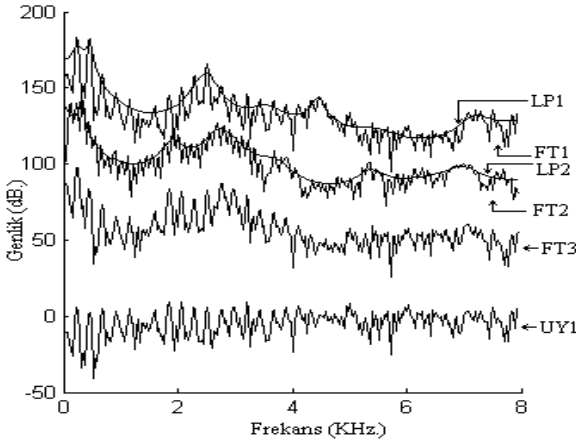
Nakillerde her bir anda birbirine karşılık gelen öznitelikler 1. denklem ile bulunmuştur. Şekil 1'de üstteki grafik Kay1'i, alttaki ise Kay2'yi göstermektedir. /e/ fonemi (i no'lu fonem),

Kay1'de $b_1^i = 0.415$ ms. ile $s_1^i = 0.480$ ms. arasında, Kay2'de $b_2^i = 0.666$ ms. ile $s_2^i = 0.756$ ms. arasında yer almaktadır. PSOLA sırasında Kay1'deki i no'lu fonem içindeki herhangi bir t_1^i anına karşılık gelen Kay2'deki zaman (t_2^i) 1. denklem ile hesaplanır. Şekil 1'de $t_1^i=0.444$ ms. değeri için $t_2^i=0.706$ ms. olarak bulunmuştur.

$$t_2^i = \frac{(t_1^i - b_1^i)(s_2^i - b_2^i)}{s_1^i - b_1^i} + b_2^i \quad (1)$$

Nakil türü (Kay2'den Kay1'e)	Kısa Gösterim	Ses Borusu	Perde	Süre	Enerji
Özgün (Kon1)	A	Kay1	Kay1	Kay1	Kay1
Özgün (Kon2)	a	Kay2	Kay2	Kay2	Kay2
Ses borusu	B	Kay2	Kay1	Kay1	Kay1
Perde + süre + enerji	b	Kay1	Kay2	Kay2	Kay2
Perde	C	Kay1	Kay2	Kay1	Kay1
Ses borusu + süre + enerji	c	Kay2	Kay1	Kay2	Kay2
Süre	D	Kay1	Kay1	Kay2	Kay1
Ses borusu + perde + enerji	d	Kay2	Kay2	Kay1	Kay2
Enerji	E	Kay1	Kay1	Kay1	Kay2
Ses borusu + perde + süre	e	Kay2	Kay2	Kay2	Kay1
Ses borusu + perde	F	Kay2	Kay2	Kay1	Kay1
Süre + enerji	f	Kay1	Kay1	Kay2	Kay2
Perde + Enerji	G	Kay1	Kay2	Kay1	Kay2
Ses borusu + süre	g	Kay2	Kay1	Kay2	Kay1
Ses borusu + enerji	H	Kay2	Kay1	Kay1	Kay2
Perde + süre	h	Kay1	Kay2	Kay2	Kay1

Tablo 1: Tüm akustik öznelik nakil kombinasyonları.



Şekil 2: Kay1 doğrusal öngörü analizi (LP1) ve Fourier (FT1) spektrumları, Kay2 doğrusal öngörü analizi (LP2) ve Fourier (FT2) spektrumları, nakil çıktısı Fourier spektrumu (FT3) ve Kay1 uyarım spektrumu (UY1).

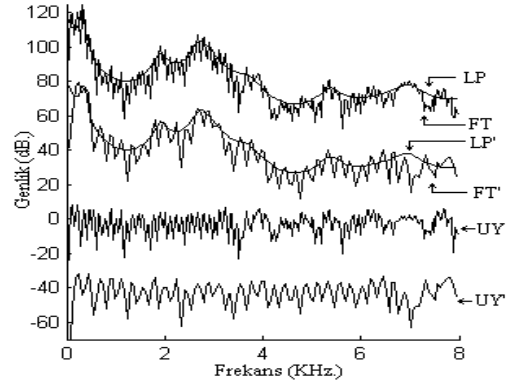
2.1. Ses Borusu Yapısı Nakli

Ses borusu yapısı (Tablo 1'de "B") perde senkronize olarak Kay2'den Kay1'e nakledilmiştir. Kay1 işlenirken 1. denklem kullanılarak Kay2'de, üzerinde işlem yapılan foneme karşılık gelen zaman hesaplanır. O andaki ses borusu parametreleri doğrusal öngörü analiziyle bulunur. Doğrusal öngörü analizine dayalı spektrum Şekil 2'de LP1 ile gösterilmiştir. Fourier dönüşümünden elde edilen spektrum (FT1), LP1 spektrumu ile bölünerek uyarım spektrumu (UY1) bulunur. Bu spektrum

Kay2'den elde edilen LP spektrumu (LP2) ile çarpılınca nakil çıktısının Fourier spektrumu (FT3) elde edilir. FT3'ün zaman uzayındaki karşılığı ters Fourier dönüşümüyle hesaplanır ve PSOLA yönteminde kullanılır. FT2, Kay2'nin Fourier spektrumuna karşılık gelmektedir. (Çizimlerde LP1 ve FT1 80 dB., LP2 ve FT2 40 dB. yukarı kaydırılmıştır.)

2.2. Ses Perdesi Nakli

Kay2'ye uygulanacak anlık ses perdesi dönüşüm oranı ($\beta(t)$) Kay1'deki anlık f_0 değeri ($f_1(t)$) ve buna karşılık gelen Kay2'deki anlık f_0 değerinden ($f_2(t)$) 2. denklem kullanılarak hesaplanır. Kay1'deki doğrusal öngörü analizi spektrumu Bölüm 2.1'deki gibi bulunur. Ses perdesinin artırılması ($\beta(t)>1.0$) Şekil 3'te gösterilmiştir. Öncelikle özgün doğrusal öngörü analizi spektrumu (LP), uygulanacak ses perdesi değiştirme oranında sıkıştırılır. Sıkıştırılmış LP spektrumuyla karşılık gelen uyarım spektrumu parçası (UY') çarpılır ve ses perdesi artırılmış işaretin Fourier spektrumu (FT') elde edilir. $\beta(t)<1.0$ için ses perdesinin düşürülüp daha kalın bir sesin elde edilmesi gerekmektedir. Bu amaçla ses perdesi artırma yöntemine benzer bir yöntem kullanılabilir (LP spektrumunun genişletilmesi ve uyarım spektrumuyla çarpılması). Fakat bu durumda genişletilen LP spektrumunun yüksek frekanslarına karşılık gelen uyarım spektrumunun da interpolate edilmesi gerekmektedir. İnterpolasyon için yaygın olarak kullanılan yöntem, uyarım spektrum değerlerinin yüksek frekanslarda tekrar edilmesidir. Fakat bu yöntem gürültüye yol açabilmektedir. Gürültünün engellenmesi için ses perdesinin azaltılması zaman uzayında PSOLA yöntemiyle gerçekleştirilir. Ses perdesinin iki katı genişliğindeki parçalar üstüste ekleme sırasında birbirinden uzaklaştırılarak ses perdesi istenen oranda azaltılır. PSOLA yöntemlerinde gürültüyü azaltmak için $\beta(t)$ 0.5 ile 2.0 arasında sınırlandırılmıştır.



Şekil 3: Özgün doğrusal öngörü (LP), Fourier (FT) ve uyarım (UY) spektrumları; ses perdesi artırılmış işaretin doğrusal öngörü (LP'), Fourier (FT') ve uyarım (UY') spektrumları.

$$\beta(t) = \frac{f_2(t)}{f_1(t)} \quad (2)$$

Bazı durumlarda Kay1'de sesli olan bir kısım Kay2'de sessiz bir kısma karşılık gelebilir. Bu durumda sessiz kısım komşu sesli f_0 değerleri arasında doğrusal olarak interpolate edilir.

2.3. Süre Nakli

Kay1 ve Kay2'deki karşılıklı fonem sürelerinin oranı genel olarak her fonem için farklı olduğundan Kay1'e zamanla değişen oranlarda süre değişimi uygulanmalıdır. PSOLA

sırasında ses perdesiyle senkronize sinyal parçalarının uygun sayıda yinelenmesi veya bazı parçaların atlanmasıyla süre değişimi gerçekleştirilebilir. Bu yöntem sabit süre değişimi için yüksek kalitede sonuç vermektedir. Fakat süre değişim oranı zamanla değiştiğinde, çok küçük (<0.5) ya da çok büyük (>2.0) olduğunda ses kalitesi düşmektedir. Gürültünün azaltılması için Kay1'deki fonem sürelerinin Kay2'deki fonem sürelerine dönüştürülmesi yerine Kay1'deki ses borusu yapısı, ses perdesi ve enerji öznitelikleri Kay2'ye nakledilmiştir. Kay2'deki uyarım sinyali özelliklerinin ses perdesi ve enerji dışında korunması nedeniyle süre nakli (Tablo 1'de "D"), ses borusu + ses perdesi + enerji naklinin (Tablo 1'de "d") tümleyeni olarak değerlendirilmelidir.

2.4. Enerji Nakli

Enerji nakli (Tablo 1'de "E") için Kay2'deki anlık enerji değerinin, $e_2(t)$, Kay1'deki karşılık gelen anlık enerji değerine, $e_1(t)$, oranı 3. denklem yardımıyla hesaplanır. Belirlenen enerji değiştirme oranı, $\varepsilon(t)$, Kay1'deki anlık enerji değerinin değiştirilmesinde kullanılır.

$$\varepsilon(t) = \frac{e_2(t)}{e_1(t)} \quad (3)$$

Tablo 1'deki diğer nakiller ("F", "G", "H", "f", "g" ve "h"), Bölüm 2.1-2.4'de anlatılan dört temel yöntemlerin birlikte uygulanmasıyla gerçekleştirilir.

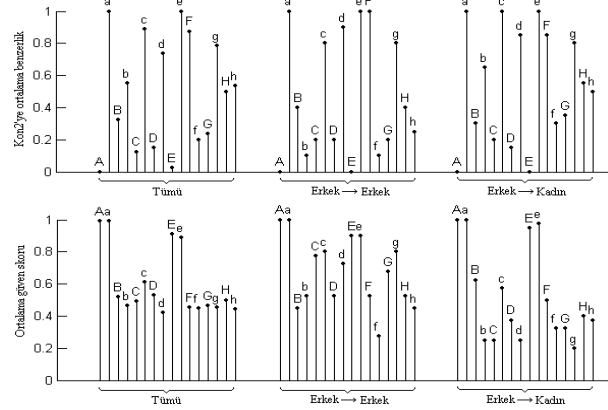
2.5. Dene Yöntemi

Dört akustik özneliğin konuşmacı kimliğinin algılanmasına etkilerinin incelenmesi için bu kısımda öznel bir dinleme deneyi tasarlanmıştır. 4 bayan, 4 bay konuşmacıdan, Türkçe bir veri tabanı kaydedilmiştir. Veri tabanı 50 sözcük ve 29 tümceden oluşmaktadır. Her konuşmacı çifti için 16 sözcük ve 16 tümce rastgele seçilmiştir. Her sözcük ve tümcede Tablo 1'de gösterilen öznel nakillerinden biri uygulanmıştır. Konuşmacı çiftleri tüm cinsiyet kombinasyonlarını içerecek şekilde seçilmiştir. Deneklere 128 kayıt üçlüsü dinletilmiştir. Her üçlü, iki özgün konuşmacı kaydıyla Tablo 1'deki nakillerden birini içermektedir. Deneklerden dinledikleri üçlüdeki nakil çıktısının hangi konuşmacıya ne oranda benzediğini belirtmeleri istenmiştir. Konuşmacı kimliği için üç farklı seçim yapmak mümkündür: "1. konuşmacı", "2. konuşmacı" ve "hiçbiri". Denek, seçtiği konuşmacı kimliği konusunda ne kadar emin olduğunu belirtmek için güven skoru olarak adlandırılan ikinci bir skor daha vermiştir. Güven skoru 1 (düşük) ile 5 (yüksek) arasında seçilen tamsayı değerlerinden oluşmaktadır. Deneklerden, konuşmacının iki konuşmacıya da yakın olması durumunda konuşmacı kimliği için hiçbiri seçimini yapmaları ve güven skoru olarak 1'i (düşük) seçmeleri istenmiştir.

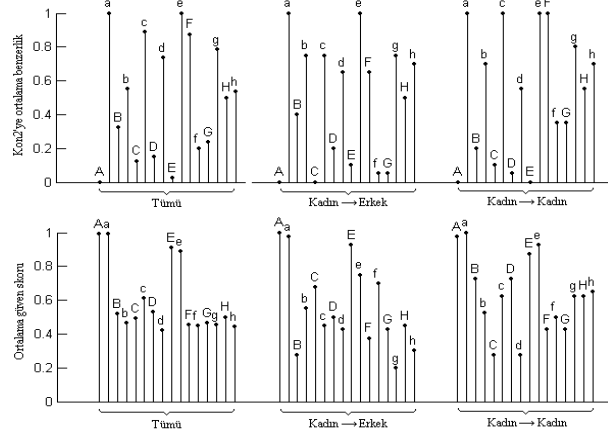
2.6. Sonuçlar

Konuşmacı kimliği seçimleri sayısal değerlere dönüştürülmüştür ("1. konuşmacı" seçimi için 0.0, "hiçbiri" seçimi için 0.5 ve "2. konuşmacı" seçimi için 1.0). Güven skoru en yüksek değer 1.0 olacak şekilde normalize edilmiştir. Farklı cinsiyet kombinasyonları ("Tümü", "Erkek→Erkek", "Erkek→Kadın", "Kadın→Erkek" ve "Kadın→Kadın") için ortalama değerler ve çeyreklerarası aralıkları (CAA) hesaplanmıştır. Standart sapmaya benzer şekilde bir değişkene ait değerlerin ortalama çevresinde ne

kadar dağınık olduklarını gösteren CAA, deneklerin yanıtlarındaki örtüşmeyi ölçmek amacıyla kullanılmıştır. Dene sonuçlarında 0.0'a yakın CAA değerleri deneklerin kararlarının uyum içinde olduğunu göstermektedir. Aşağıdaki değerlendirmelerde CAA değerleri göz önünde bulundurulmuştur.



Şekil 4: "Tümü", "Erkek→Erkek", ve "Erkek→Kadın" nakillerinde Kon2'ye ortalama benzerlik (Üst) ve ortalama güven skorları (Alt).



Şekil 5: "Tümü", "Kadın→Erkek", ve "Kadın→Kadın" nakillerinde Kon2'ye ortalama benzerlik (Üst) ve ortalama güven skorları (Alt).

Şekil 4 ve 5'te ortalama skorlar gösterilmiştir. Her şekilde iki grafik bulunmaktadır. İlk grafikte ortalama konuşmacı kimliği skorları, ikinci grafikte ortalama güven skorları gösterilmiştir. Her grafikte farklı bir cinsiyet kombinasyonuna karşılık gelen üç grup mevcuttur. Örneğin Şekil 4'te üstteki grafikte ilk grup "Tümü" olarak işaretlenmiştir. Bu grupta "B" harfi ile işaretlenmiş üçüncü doğru, Tablo 1'deki ses borusu yapısı nakline ("B") karşılık gelmektedir. Ortalama konuşmacı kimliği skoru 0.32'dir. Altta grafikte ortalama benzerlik oranının 0.52 olduğu görülmektedir. Bu durumda ses borusu yapısı nakliyle elde edilen çıktı 1. konuşmacıya ("Kon1") yakındır. Dolayısıyla sadece ses borusu yapısının dönüştürülmesi hedeflenen konuşmacıya ("Kon2") benzerliği sağlamamaktadır. Her gruptaki ilk iki doğru özgün konuşmacılara (sırasıyla "Kon1" ve "Kon2") karşılık gelmektedir. Teorik olarak "A" için konuşmacı kimliği skoru 0.0, benzerlik skoru 1.0; "a" için konuşmacı kimliği skoru "1.0", benzerlik skoru 1.0 olmalıdır. Sonuçlar bu değerlere

çok yakın olup deneklerin özgün konuşmacıları başarıyla tanıdıklarını göstermektedir.

Şekil 4 ve 5'te Kon2'ye en benzer çıktının ses borusu yapısı + ses perdesi + süre nakliyle (Tablo 1'de "e") elde edildiği görülmektedir. "e" için ortalama benzerlik oranı yüksek, ÇAA değeri düşüktür. Tümleyen nakil olan enerji nakli (Tablo 1'de "E") Kon2'ye en düşük benzerliğe sahiptir. Ses perdesi ("C") ve ses perdesi + enerji ("G") nakilleri de düşük skorlar almıştır. Ses borusu yapısı + ses perdesi ("F"), ses borusu yapısı + süre + enerji ("c") ve ses borusu yapısı + süre ("g") nakilleri Kon2'ye yüksek oranda benzerlik sağlamıştır. Denekler, ses perdesi ve süre nakillerine birbirine yakın konuşmacı kimliği skorları vermiştir. Ses perdesi nakilleri Kon1 ve Kon2'nin cinsiyetleri farklı olduğunda nakil çıktısının Kon2'ye benzerliğini arttırmıştır. Ortalama skorlar tümleyen nakillerin konuşmacı kimliği skorlarının da birbirini tümlediğini göstermektedir. Örneğin enerji nakliyle ("E") ses borusu yapısı + ses perdesi + süre nakline ("e") verilen konuşmacı kimliği skorları toplamının yaklaşık 1.0 olduğu gözlenmektedir. Bu durum belli akustik özneliklerin nakledilmesinin Kon2'ye benzerlik açısından yetersiz olduğu durumlarda diğer akustik öznelik nakillerinin başarılı olacağını göstermektedir.

3. Deneysel İçeriği

Bu bölümde, farklı frekans bantlarının konuşmacı kimliği algılanmasındaki önemleri öznel bir dinleme deneyiyle ölçülmektedir. Deneyde sekiz konuşmacıdan kaydedilen on Türkçe sözcük kullanılmıştır. Kayıtlar 44.1 KHz örnekleme sıklığında gerçekleştirilmiştir. Sözcük veri tabanı ile deneklerin vurgu bilgisini kullanarak konuşmacıları tanıma olasılıkları azaltılmıştır. Veri tabanındaki vurgusal farkların daha da azaltılması için PSOLA tabanlı bir yöntemle tüm sözcükler sabit f0'da (180 Hz.) sentezlenmiştir. PSOLA sonrasında tüm sözcükler 10 bantlı bir bant-geçiren süzgeç bankasıyla alt-bant bileşenlerine ayrılmıştır. Süzgeç bankasının tasarımında kullanılan 50. derece sonlu dürtü yanıtı süzgeçlerin kesim frekansları Tablo 3'te gösterilmiştir.

3.1. Deneysel Yöntemi

10 deneye sabit f0'da sentezlenmiş kayıtlarla bunların alt-bant bileşenlerinden oluşan ikililer dinletilmiş ve ikililerdeki konuşmacıların birbirine benzerliğiyle ilgili skorlar vermeleri istenmiştir (Tablo 2). Toplam 8 konuşmacı ve 10 alt-bant olduğundan her denek 8x10=80 ikili dinlemiştir. Her alt-bant için skorların ortalaması ve standart sapması hesaplanmıştır. Tablo 3'te sonuçlar verilmiştir.

3.2. Sonuçlar

1034-1895 Hz aralığındaki ikinci alt-banda en yüksek skor verilmiştir (63.3). 2. formantı kapsayan bu alt-bant konuşmacı kimliğinin algılanması açısından en önemli frekans aralığıdır. İkinci en önemli alt-bant 1895-2756 Hz aralığında yer almaktadır.

4. Tartışma

Bu çalışmada farklı akustik özneliklerin konuşmacı kimliğinin algılanmasındaki önemlerinin belirlenmesi için iki öznel dinleme deneyi tasarlanmış ve uygulanmıştır. İlk deneyde PSOLA yöntemi kullanılarak ses borusu yapısı, ses

perdesi, süre ve enerji öznelikleri konuşmacılar arasında nakledilmiştir. Ses borusu yapısı + ses perdesi + süre nakli çıktının hedeflenen konuşmacıya benzerliği açısından en başarılı sonucu vermiştir. Diğer önemli öznelik nakli kombinasyonları ses borusu yapısı + ses perdesi, ses borusu yapısı + süre + enerji ve ses borusu yapısı + süre nakillerini içermektedir. Tek özneliğin nakledildiği durumlarda ses borusu yapısının en önemli özellik olduğu belirlenmiştir. Bu durum [2]'de bildirilen sonuçlara benzerdir. Ses perdesi ve enerji özneliklerinin konuşmacı kimliği algılanmasında ses borusu yapısına göre daha az önemli olduğu gözlemlenmiştir. Cinsiyetler arasındaki nakillerde, ses perdesi süreden daha önemlidir. En az öneme sahip olan öznelik enerjidir. İkinci deneyde farklı frekans bantlarının konuşmacı algılanmasındaki önemleri incelenmiştir. Bu amaçla vurgusal farkları olabildiğince giderilmiş sözcükler, bant-geçiren bir süzgeç bankası yardımıyla alt-bant bileşenlerine ayrılmıştır. Sonuçlar 1034-1895 Hz aralığının konuşmacı kimliğinin algılanmasında en önemli frekans bandı olduğunu göstermektedir. İkinci en önemli aralık 1895-2756 Hz aralığını kapsamaktadır.

Skor	
100	Konuşmacılar aynı
75	Konuşmacılar benzer
50	Fikirim yok
25	Konuşmacılar benzemiyor
0	Konuşmacılar farklı

Tablo 2: Benzerlik skorları.

Alt-bant no	Frek. (Hz.)	Ortalama	Std. Sap.
1	0-1034	42.2	30.9
2	1034-1895	63.3	27.7
3	1895-2756	46.9	28.7
4	2756-3618	35.9	30.4
5	3618-4482	32.8	28.0
6	4482-5346	29.7	38.8
7	5346-6210	39.8	37.5
8	6210-7074	39.1	40.1
9	7074-7938	14.1	31.7
10	7938-8802	12.5	30.4

Tablo 3: Farklı frekans aralıklarının konuşmacı kimliği algılanmasındaki önemi.

5. Kaynakça

- [1] Matsumoto, H., Hiki, S., Sone, T. ve Nimura, T., "Multidimensional representation of personal quality of vowels and its acoustical correlates", *IEEE Trans. AU*, AU-21, sf. 428-436, 1973.
- [2] Itoh, K. ve Saito, S., "Effects of acoustical feature parameters of speech on perceptual identification of speaker", *IECE Transactions*, J65-A, sf. 101-108, 1982.
- [3] Necioğlu, B. F., Clements, M. A., Barnwell III, T. P. ve Schmidt-Nielsen, A., "Perceptual relevance of objectively measured descriptors for speaker characterization", in *Proc. of ICASSP*, sf. 869-872, 1998.
- [4] Furui, S., "Research on individuality features in speech waves and automatic speaker recognition techniques", *Speech Commun.*, 5 (2), sf. 183-197, 1986.
- [5] Arslan, L.M., "Speaker transformation algorithm using segmental codebooks", *Speech Commun.*, 28, sf. 211-226, 1999.
- [6] Türk, O. *New methods for voice conversion*. Yüksek Lisans Tezi. Elektrik-Elektronik Müh. Bölümü. Boğaziçi Üniversitesi, İstanbul, 2003.
- [7] Campbell, J. P., "Speaker recognition: A tutorial", in *Proc. of the IEEE*, 85 (9), sf. 1437-1462, 1997.