

Objektif Ses Analizi

OBJECTIVE VOICE ANALYSIS

Fatih ÖGÜT*

*Doç.Dr., Ege Üniversitesi Tıp Fakültesi Kulak Burun Boğaz Hastalıkları AD, İZMİR

Özet

Objektif ses analizi, ses bozukluklarının tanı ve kayıtlanmasında kullanılmaktadır. Bu teknikler linguistik araştırmalarda da kullanılmaktadır. Ses analizinin sağlıklı bir şekilde gerçekleştirilmesi için, ses fiziolojisinin iyi bilinmesi gerekmektedir. Ses bozukluklarının değerlendirilmesi için birçok yöntem ve parametre bulunmaktadır. Sesin değerlendirilmesinde objektif ses analizi yöntemleri kadar, subjektif ses değerlendirme yöntemleri de önem taşımaktadır. Bu çalışmada, öncelikle ses fiziği, subjektif ses değerlendirme yöntemleri anlatıldıktan sonra, objektif ses analizi için günümüzde kullanılan teknikler ve değerlendirme parametreleri sunulmuştur.

Anahtar Kelimeler: Ses fiziolojisi, Ses analizi

T Klin K B B 2002, 2:121-126

Summary

Objective voice analysis is used in recognition and recording of the voice disorders. These techniques are also used in linguistics. The knowledge of the physiology of the voice is very important in voice analysis. The subjective evaluation of the voice is as important as the objective voice analysis methods. Many techniques and parameters are being used to evaluate the voice disorders. In this study, the physics of the voice and the subjective evaluation methods are mentioned and the current techniques and evaluation parameters in the objective voice analysis are presented after the subject.

Key Words: Physiology of voice, Voice analysis

T Klin J E N T 2002, 2:121-126

Foniatri ses ve konuşma bozuklukları tanı ve tedavisini inceleyen bilim dalıdır. Ses ve konuşmayı oluşturan periferik organlar kulak burun boğaz hastalıkları anatomik bölgesinde yer aldığından dolayı, bu patolojiler kulak burun boğaz hekimini yakından ilgilendirmektedir.

Ses ve konuşma bozukluklarının tanı ve tedavisini sağlamak amacı ile kulak burun boğaz klinikleri içinde foniatri merkezleri kurulmaya başlanmıştır. Anabilim dalımızda bu amaçla 1973 yılında foniatri laboratuvarı kurulmuş olup, teknolojinin gelişmesine bağlı olarak modernize edilmiştir.

Ses ve konuşma bozuklukları diğer patolojilere göre bazı özellikler göstermektedir. Tanı koyma amacı ile kullanılan sistemler modern teknoloji gerektirmekte ve yüksek maliyet göstermektedir. Bunun yanında ses kalitesinin incelenmesi ve değerlendirilmesi ayrı bir uzmanlık alanı oluşturmaktadır. Ayrıca ses ve konuşma bozukluklarının tedavi

programları da değişiklik göstermekte ve özellik taşımaktadır. Bunun yanında bu patolojilerin çok değişiklik göstermesi birçok uzmanlık alanını aynı anda ilgilendirmekte ve bu konu ile ilgilenecek kulak burun boğaz uzmanının patolojinin cinsine göre değişik uzmanlık alanları ile ilgili bilgisi olması gerekir. Örneğin, müzik bilgisi olmayan bir kulak burun boğaz uzmanının ses sanatçısındaki patolojiyi oluşturan teknik eksikliği saptaması zorluk göstermektedir. Ses ve konuşma bozukluklarının tanısı iyi bir anamnez ve kulak burun boğaz bakışı ile konabilmektedir. Larengovideostroboskopi ve videokinematografik incelemeler günümüzde tanı için kullanılan en gelişmiş ve en güvenilir yöntemlerdir.

Objektif ses analizleri foniatri içinde giderek önemini arttırmaktadır. Ancak ses analizine geçmeden önce foniatride hastaya yaklaşımın çok iyi planlanması gerekmektedir. Kulak burun boğaz

uzmanı hasta kliniğe girer girmez onunla konuşmaya başlamalıdır. Tam bir değerlendirme konuşmacının sesini tipik olarak nasıl kullandığını ve hasta ses kapasitesini saptamaktan geçmektedir. Böylece daha sonra yapılacak fizik muayene bulguları ile bu bulgular karşılaştırılacak ve en doğru tanıyı bulma mümkün olacaktır (2).

Subjektif ses değerlendirilmesi foniatrı ve logopedi uygulamasında sık olarak kullanılmaktadır. Ses bozukluğunun saptanmasında bir çok semptom bulunmasına karşın deneyimli klinisyenler tarafından en çok kullanılan 3 değişken bulunmaktadır. Bunlardaki sapmaları saptamak suretiyle tanıya gitmede kolaylık elde edilmiş olunur. Bunlar seste hava kaçağı, gerginlik ve kitlesel lezyon etkisi olmaktadır (3-4).

Hava kaçağı boşa hava çıkarmadır ve organik ya da fonksiyonel bir nedene bağlı olabilir (Örneğin tek taraflı kord vokal paralizisi).

Ses gerginliği kendini ses kısıklığı ile belli eder. Bu da çoğunlukla kas gerilimine bağlı olarak ortaya çıkmaktadır. Eğer hasta ses çıkarırken efor sarf ediyorsa ses gergin ve sıklıkla hava kaçaklı olacaktır. Vokal nodüller ses gerginliği sonucu ortaya çıkmaktadır.

Kitlesel lezyonlar sıklıkla kabul edilebilir ses değişiklikleriyle birlikte temel frekansta düşme, şiddet düzensizliği ve ses kısıklığı ile ortaya çıkmaktadır. Reinke ödemi kitlesel lezyon sesinin karakteristik örneğidir.

Bu parametreleri tanımlayarak tecrübeli klinisyenler genellikle bir çok sesi sınıflandırıp larengeal muayeneyi yapmadan larengeal bulgular konusunda tahminde bulunabilirler. Ayrıca temelde bu parametreler dikkate alınmak suretiyle bir takım sesi algısal olarak değerlendiren skalalar ortaya konmuştur (2).

Subjektif Skala Tipleri

Kategoriksel puanlamalar: Bu skalalar konuşma veya ses örneklerinin ayrı belli parametreye bağlı olmadan değerlendirilmesi ile oluşturulmuştur. (örneğin havalı, kaba).

Eşit görünen aralıklı (Equal-appearing interval EAI) skalalarının esası dinleyicilerin ses örnekleri-

ne 1 ve n arasında numara vermesinden geçmektedir. EAI skalası üzerindeki noktalar eşit aralıklı olarak hesaplanması ve istatistikler uygulanması esasına dayanmaktadır.

Görsel analog (Visual Analog "VA") skalaları değişken olmayan çizgilerdir ve genellikle 100 mm uzunluğundadırlar. Dinleyicilerin seslere ait çizgi üzerinde bir nokta yapması suretiyle sesin verilen hangi özelliklere uyduğu bulunur.

Direkt büyüklük yorumunda (Direct Magnitude Estimation, DME), ses örneğine bir numara vermek suretiyle ses kalitesinin nereye ait olduğuna karar verilir. Olası numaralar genel olarak kısıtlanmamıştır. İşaretle DME de diğerlerinden farklı olarak özel bir büyüklüğü olan bir referans ses örneği bulunmaktadır (genellikle 100). İşaretlenmemiş DME de ise dinleyiciler kararlarını kendi kriterlerine göre oluşturdukları referansa göre vermektedirler. Sonuç olarak çiftleştirilmiş karşılaştırmada dinleyiciler 2 uyarıyı karşılaştırırlar. Bazı boyutlarda farklılığın uzanımı konusunda benzerlik/değişkenlik açısından karar verebilirler. Bunlardan Japon foniatrı Derneği tarafından ileri sürülen en sık kullanılan subjektif disfoni değerlendirme skalası, bir EAI skalası olan GRBAS Skalası olmaktadır: Bu skala sesin Grade of severity (disfoni derecesinin önemi), Roughness (kabalık, frekans tutarlılığının bulunup bulunmaması), Breathiness (hava sesi olması), Asthenicity (güçsüzlük, hipokinetik özellikte olma), Strain (gergin, hiperkinetik özellikte olma) gibi özelliklerinin subjektif olarak değerlendirilmesini içermektedir (3,5).

Ses bozuklukları kulakla değerlendirilirken, deneyimli bir KBB uzmanı patolojiyi saptayabilmektedir. Fakat normal seslerde bile geniş bir spektrum olduğundan bir kişiye göre disfonik gelen bir ses diğer kişi tarafından normal olarak değerlendirilebilmektedir.

Eğer laboratuvar olanakları bulunursa disfonik hastanın tanısında klinik değerlendirmeden sonra aşağıdaki aygıtlardan da yararlanılabilmektedir.

1) Glottografi

Fonatuvar vibrasyonlar esnasında glottal bölgede oluşan değişimlerin incelenmesi için glottografi ortaya konmuştur.

2. Elektrolottografi (EGG)
- 3) Fotolottografi (FGG)
- 4) Ultrasound glottografi (UGG)
- 5) Ultra yüksek hızda sinematografi
- 6) Fundamental Frekansmetre
- 8) Elektroaerometre
- 9) Manofon
- 10) Sonograf
- 11) Real Time Oktav Analizatör
- 12) Larengeal EMG
- 13) Akustik analiz

Konumuzu oluşturan akustik analizin daha iyi anlaşılabilmesi için ses fiyolojisinin iyi bilinmesi gerekmektedir.

Ses Fizigi

Ses yer değiştiren dalgalardan meydana gelmiştir. Denebilir ki ses dalgaları birinden diğerine ulaşan materyal ortamların vibrasyonlarıdır. Her materyal komşu partikül üzerine kendinde bulunan hareketi nakletmek için etki yapar. Bir ortam içinde sesin yayılması sesin ve ortamın özelliklerine bağlıdır. Ses dalgası bir düzlem üzerine bir basınç meydana getirir, bu da sesin fizik şiddetini ölçmeye yarar.

Periyodik bir ses üç temel özellik ile karakterizedir.

- Period: Sinyalin birbirini takip eden iki geçişi arasındaki (T) zamanıdır. Saniye ile ölçülür.
- Frekans: Saniyedeki periyod sayısına eşittir. Hertz olarak ölçülür.
- Amplitüd: Sesin şiddeti, insan sesinde dBA olarak ölçülür.

Saf Ses ve Kompoze Sesler:

Saf ses sinüzoidal, vibratuar bir hareket tarafından meydana getirilir. Müzikal ses periyodik, fakat sinüzoidal olmayan seslerdir. Bu seslere kompleks veya kompoze sesler adı da verilebilir. Çünkü kompleks bir vibrasyonun daima belli sayıdaki basit vibrasyonların sonucu olarak ortaya çıktığı kabul edilir. Fourier'nin matematik teorisi

kullanılarak sesin formantik özellikleri elde edilebilmektedir.

Eğer kompleks ses sinüzoidal seslerden oluşuyorsa ve bu sesler temel frekansın tam katlarından oluşursa bunlara harmonik ses adı verilmektedir. İlk oluşan frekansa ise temel frekans denmektedir.

İkinci bir durumda ise f1 ve f2 vs. gibi oluşan frekans bölgeleri temel frekansın tam katları değildirler. Bunlara formant veya parsiyel adı verilmekte ve bu seslere havalı aletlerde rastlanmaktadır.

Bu bilgilere göre sesli harfler kompleks periodik sesler olup, periodik olmayan sessiz harflerden ayrı edilebilmektedir.

Periodik her sinyal sinüzoidal fonksiyonlara ayrıştırılabilir, fakat periodu çok büyük kabul edilecek olursa periodik olmayan herhangi bir sinyal periodikmiş gibi düşünülebilir. Her sinyalin üstüste binmiş sinüzoidal fonksiyonların sonucu olduğu düşünülebilir. Fourier ilkelerine göre sinyalin spektral analizi yapıldığında frekansiyal dağılım ortaya çıkmaktadır. Fourier analizleri ile ilgili daha geniş bilgi ses analizleri bölümünde verilecektir.

Akustik ölçüm olan jitter ve shimmer sık olarak kullanılmaktadır. Bunlar jitter Fo da siklusa da sıklusa olan değişiklikler iken shimmer ise amplitüdeki değişkenlikler olmaktadır. Jitter ve shimmer non-invasif yoldan vokal değişkenliğinin ve farklılığının normal ve patolojik sesde uygun olarak saptanmasına yarar. Bu iki akustik parametrenin veri alımında ve analizinde kullanılan metodların bir çok araştırmacı arasında değişik olması sonuçlardan kuşku duyulmasına neden olmuştur (6).

$$\text{Jitter, Shimmer} = \frac{100}{(N-1) \bar{X}} \sum_{i=2}^N [X_i - X_{i-1}]$$

N: Dalga sayısı

X: Jitterde sinüzoidal ses dalgasının periyodu (süresi)

Shimmerde sinüzoidal ses dalgasının amplitüdü

Gürültüye harmonik yüzdesi, harmoniğin gürültüye oranı olarak rapor edilmiştir. Bu ölçümün disfoni ciddiyet derecesinin değerlendirilmesinde objektif bir metod olduğu bulunmuştur.

$$\text{Gürültüye harmonik yüzde} = 100 \frac{\int_0^T f_A^2(t) dt}{\sum_{i=1}^N \int_0^T [f_i(t) - f_A(t)]^2 dt}$$

$$f_A(t) = \frac{1}{N} \sum_{i=1}^N f_i(t)$$

Bu formülde sinyali oluşturan formantların amplitüdlerinin sinyalde bulunan düzensiz amplitüdlere oranı saptanmaktadır. N : Dalga sayısı, A2: Sinyalin integral alanı, (t): zaman

Bunun yanında bir çok akustik parametre belirtilmiştir. Bunlardan en çok kullanılanları aşağıdaki tabloda gösterilmiştir. (1)

Aerodinamik parametreler	Kısaltması	Kadın ortalama değeri	Erkek ortalama değeri
Maksimum fonasyon süresi (Maximum phonation time)	MPT	16,2 s	21,8 s
Vital kapasite (Vital capacity)	Capacity	3235 cc	4665 cc
Fonasyon katsayısı (Phonation quotient)	PQ	219	235

Akustik parametreler

Ortalama temel frekans (Mean fundamental frequency)	Fo	212 Hz	122 Hz
Temel frekansın standart sapması (Standard deviation of Fo)	STD	2,6 Hz	1,4 Hz
Fonatur temel frekans aralığı (Phonatory fundamental frequency range)	PFR	2,8 semitone	2,5 semitone
Temel frekansın tremor frekansı (Tremor frequency of Fo)	Fftr	2,6 Hz	2,7 Hz
Tremor frekansının amplitüdü (Amplitude of tremor)	Fatr	2,55 Hz	2,77 Hz
Absolüt jitter (Absolute jitter)	Jita	50 ms	69 ms
Jitter yüzdesi (Jitter percent)	Jitt	% 1,04	% 0,81
Relatif pertürbasyon ortalaması (Relative average of perturbation)	RAP	% 0,62	% 0,48
Pertürbasyon katsayısının ton süresi (Period of perturbation quotient)	PPQ	% 0,57	% 0,48
Yumuşatılmış pertürbasyon katsayısının ton süresi (Smoothed period of perturbation quotient)	SPPQ	% 0,72	% 0,72
Temel frekansın varyasyonu (Variation of Fo)	VFo	% 1,25	% 1,17
dB Shimmeri (dB shimmer)	ShdB	0,3 dB	0,31 dB
Shimmer yüzdesi (Shimmer percent)	Shim	% 3,4	% 3,6
Pertürbasyon katsayısı amplitüdü (Amplitude of perturbation quotient)	APQ	% 2,3	% 2,8
Pertürbasyon katsayısı düzleştirilmiş amplitüdü (Smoothed amplitude of perturbation quotient)	SAPQ	% 3,6	% 4,8
Amplitüd maksimum varyasyonu (Variation of maximum amplitude)	VAM	% 11,3	% 13,8

Gürültü harmoni oranı (Noise harmonic ratio)	NHR	0,121	0,133
Ses türbülans indeksi (Voice turbulence index)	VTI	0,044	0,042
Yumuşak fonasyon indeksi (Soft phonation index)	SPI	16,8	18,8

Akustik analizin yapılabilmesi için bazı noktalara dikkat edilmelidir.

1. Örnekleme sayısı

- * Telefon konuşması için 8 kHz.
- * Ses analizi için minimum 16 kHz
- * Kompakt disk odyo standardı için 44.1 kHz.
- * Dijital odyo kasetleri için 48 kHz. örnekleme hızı gerekmektedir.

2. Kayıtlama İşlemleri

Kayıtlamada kullanılacak sistem önem taşımaktadır. Burada özellikle seçilen sistemin standardizasyonu ve güvenilirliğinin test edilmesi gerekmektedir.

3. Temporal ve Bit çözünürlüğü

Distorsiyon analizinde temporal çözünürlük oldukça önemlidir. çözünürlük zamanının ve örnekleme oranının daha iyi olması daha güvenilir sonuçların ortaya çıkmasına neden olmaktadır.

4- Örnekleme süresi

İlk tekniklerde fonasyonun kısa bölümleri analiz ediliyordu. Günümüzde teknolojinin gelişmesiyle daha uzun bölümlerin analizleri yapılabilmektedir.

5. Çıkarma işlemi

Peak-yakalama ya da sıfırı-geçme interpolasyonlu ya da olmadan çıkarma işleminde kullanılabilir. Bu işlemler ses sinyallerinin analizinde kullanılan belirli filtre yöntemleridir.

6. Pre-filtreleme

Low-pass filtreleme kullanıldığı zaman jitter ölçümünün etkilenmemesi için temel frekansın 1 oktav üzerindeki frekansı kesilme özelliğine sahip olması gerekmektedir. Eğer kesilen (cut-off) frekans temel frekansın daha altında olursa hesaplanan jitter beklenenin daha altına düşecektir.

Ses analiz sistemlerinin ortak yapısı incelendiğinde, bir sinyalin analog olarak alınması ve bunun bilgi işlemcinin değerlendirebilmesi için bir konvertörde dijitalize edilerek bilgisayara aktarıldığı görülmektedir. Anabilim dalımızda bu amaçla PCL 812PG laboratuvar kartı kullanılmaktadır. Bu kartla 16 sinyal aynı anda alınabilmektedir. Ancak rezolüsyon düşeceğinden genelde 2 ya da 3 kanal kullanılmaktadır. Kart hızı DMA kullanıldığında tek kanalda 50 kHz e kadar çıkabilmektedir. Genelde ses için 16000 ya da 22000 Hz, EGG için 1600 Hz lik sinyal hızı kullanılmaktadır. Burada kartın rutin programlarından yararlanarak "C" dilinde yazdığımız program kullanılmaktadır. Ancak yeni başlayanlar sound blaster kartını rahatça kullanabileceklerdir. Bu kart ile ilgili birçok hazır yazılım programı bulunmaktadır.

Klasik zaman frekans analiz metodlarında 18. yüzyıl Fransız fizikçisi olan Joseph Fourier'in prensipleri kullanılmaktadır. Fourier Transform formülü aşağıda görülmektedir (6).

$$F T_x (f) = \int_{-\infty}^{\infty} x(t) e^{-2j\pi ft} dt$$

Ancak düşük ve yüksek frekanslı bölgeler içeren bir sinyalde yanıtıcı sonuçlar elde edilebilmektedir. Bu durumda, sabit zaman ve frekans rezolüsyonu için kısa zamanlı Fourier transform formülü kullanılmaktadır. (6)

$$S T F T_x (\tau, f) = \int_{-\infty}^{\infty} x(t) g^*(t - \tau) e^{-2j\pi ft} dt$$

Bu ikinci formülde (t) değeri ile gösterilen zaman aralığına göre analiz yapılmaktadır.

FFT'de 100Hz le 110 Hz, 1000 Hz le 1010 Hz aynı kriterde değerlendirilmektedir. Buna göre hem zaman hem de frekans rezolüsyonu aynı anda artış gösteremez. Buna göre bir zaman penceresi seçildiğinde tüm frekanslar aynı rezolüsyonda değerlendirilmektedir. Ayrıca Fourier transform yöntemi tüm sinyalleri sinüzoidal sinyallerden oluştuğunu kabul etmektedir. Bundan dolayı kompleks sinyallerden oluşan insan sesinin analizini değerlendirmede bu yöntem yetersiz kalmaktadır.

Yeni bir transform yöntemi olan wavelet (dalgaçık) transformu (WT) STFT nin bazı yetersizliklerini giderebilmesi yönünden bazı avantajlar taşımaktadır. Bundan dolayı "Wavelet" analiz yöntemi öne sürülmüştür. Wavelet formülü aşağıda görülmektedir (7,8).

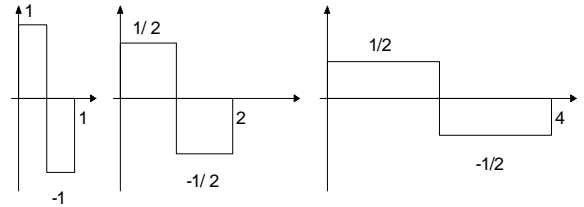
$$W T_x (\tau, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t - \tau}{a} \right) dt$$

Wt de sinyalin analizi h(t) fonksiyonu adı verilen bir ana wavelet fonksiyonu ile a olarak gösterilen skalalarla kontrakte ya da dilate edilerek gerçekleştirilmektedir. A parametre skalası frekans aralığını göstermektedir. Küçük a değerleri yüksek frekanslar, yüksek değerleri ise düşük frekanslar ile ilgilidir. WT de yüksek frekanslar daha yüksek frekans rezolüsyonu ile incelenebilmektedir [3].

Bu analiz yönteminde en basiti Haar dalgaformu olan değişik dalga formlarına göre sinyal analizi yapılabilmektedir.

$$a, b \in \mathbb{R}, a \neq 0$$

$$\psi_{\tau,a}(t) = \frac{1}{\sqrt{|a|}} \psi \left(\frac{t - \tau}{a} \right)$$



$$\psi_{0,2}(t)$$

$$\psi_{0,1}(t) = \psi(t) \quad \psi_{0,4}(t)$$

Logaritmik skalada inceleme yapılabilen bu sistemin en büyük üstünlüğü yüksek frekanslarda yüksek rezolüsyon sağlaması ve çok değişik parametreleri inceleyebilmesidir.

Ancak analiz edilecek sinyal için uygun waveletin seçilmesi önem taşımaktadır. Haar, daubechies, daubechies, coiflet, biorthogonales gibi ana waveletlerin değişik formları seçilebileceği gibi yeni wavelet formülleri de oluşturulup kullanılabilir. Ancak henüz ses analiz standartları belirtilmediği için rutin uygulamalarda güçlük göstermektedir (4,5).

Bilgisayar teknolojisindeki büyük gelişmelere karşın elde edilen sonuçlarla disfoninin tanısı yapılamamaktadır. Tecrübeli bir kulak henüz günümüzde bilgisayarlara üstün gelebilmektedir. Ancak her geçen gün artan teknoloji ve yapılan çalışmalar giderek bu açığı kapatmaktadırlar. Çünkü bu konuda tam ve objektif değerlendirme bilgisayarlar kullanılarak yapılabilecektir. İleride sadece "a" sesi değil bir konuşma bütünlüğü içinde ses incelenebilecek, ve yapılan çalışmalarla disfoni özelliklerine göre tanı ortaya konabilecektir.

KAYNAKLAR

1. Boedts D, Demanez P. Acta Oto-Rhino-Laryngologica Belgica Appendix 1, list of variables and abbreviations 1996; 50: 363-4.
2. Kreiman J, Gerratt RB, Precoda K. Listener experience and perception of voice quality. Journal of Speech and Hearing Research 1990; 33: 103-15.
3. Dejonckere PH. Perceptual evaluation of dysphonia, reliability and relevance. Folia Phoniatr. Basel 1993; 45(2): 76-83.
4. Davis SB. Acoustic characteristics of normal and pathological voices. Asha reports 1981; (11): 97-112.
5. Ögüt F. Disfonili hastanın subjektif ve objektif olarak değerlendirilmesi. KBB de Uluslararası Konferanslar Dizisi Tutanakları. İzmir, 1996: 155-68.
6. Akansu AN, Haddad RA. Multiresolution signal decomposition. Transforms, subbands and wavelets. Academic Press, San Diego, 1992.
7. Rioul O, Vetterli M. Wavelets and signal processing. IEEE SP Magazine 1991 Oct; 14-38.
8. Daubechies I. Orthonormal bases of compactly supported wavelets. Com Pure Appl Math 1988; 41: 909-96.

Geliş Tarihi: 05.09.2000

Yazışma Adresi: Fatih ÖĞÜT

Ege Üniversitesi Tıp Fakültesi
Kulak Burun Boğaz Hastalıkları AD,
İZMİR