



Dar Bantlı Konuşma İzgel Zarfının Yapay Sinir Ağları Kullanılarak Genişletilmesi ve Kalitesinin Artırılması

Bandwidth Expansion of Narrowband Speech using Artificial Neural Networks and Quality Enhancement

Gürhan Bulu^{1*}, **Ahmet Semih Bingöl¹**

¹ Hacettepe Üniversitesi, Mühendislik Fakültesi, Elektrik ve Elektronik Mühendisliği Bölümü, Ankara, TÜRKİYE
Sorumlu Yazar / Corresponding Author *: bulu@ee.hacettepe.edu.tr

Geliş Tarihi / Received: 14.02.2019

Kabul Tarihi / Accepted: 12.09.2019

Araştırma Makalesi/Research Article

DOI: 10.21205/deufmd.2020226406

Atıf şekli/ How to cite BULU, G., BİNGÖL, A.S. (2020). Dar Bantlı Konuşma İzgel Zarfının Yapay Sinir Ağları Kullanılarak Genişletilmesi ve Kalitenin Artırılması. DEUFMD 22(64), 47-58.

Öz

Geleneksel telefon sistemlerinde, 7–8 kHz'e kadar bileşenleri olan insan sesi, 3.4 kHz bant genişliğine sahip bir alçak geçiren süzgeçten geçirilip 8 kHz'te örneklenerek gönderilir. Kodlama ve benzeri işlemlerden kaynaklanan hiçbir kayıp olmasa bile, yüksek frekans bölgesinin süzülerek kaybolmasından kaynaklanan bir kalite kaybı söz konusudur. Bu kayıp, anlaşılabilirliği pek etkilememekle beraber, konuşma kalitesinde hissedilir bir bozulmaya yol açmaktadır. Bu çalışmada, izgel zarfın düşük frekanslı bölgelerinden faydalanılarak, süzülen yüksek frekanslı bölgelerin izgel zarfı elde edilmiştir. İzgel zarfın genişletilmesi olarak da nitelenebilecek bu işlem için yapay sinir ağları yöntemi kullanılmıştır. Daha sonra genişletilen bu izgel zarf kullanılarak yüksek frekanslı bölgeler kaynak-süzgeç modeliyle yeniden oluşturulmuş ve böylece konuşma kalitesinin artırılması hedeflenmiştir. Geliştirilen bu yöntem telefon kalitesindeki konuşmalar için kullanılmakla birlikte, daha düşük bant genişliğine (1.8 kHz) sahip konuşmalar için de kullanılabilir.

Anahtar Kelimeler: İzgel zarf genişletme, İzgel zarf dışdeğerleme, Konuşma kalitesi iyileştirme, Sinir ağları

Abstract

In traditional telephony, human speech which has components up to 7-8 kHz is transmitted after low-pass filtering at 3.4 kHz and sampling at 8 kHz. Even if this filtering operation has little effect on intelligibility, it generally causes a perceptible degradation in speech quality. This work deals with the reconstruction of this lost high frequency component of telephone speech using the low frequency part of the speech. Artificial neural networks are used to obtain the high frequency region, which can also be named as spectral envelope extrapolation. Then, source excitation model is used to regenerate the high frequency speech using this extended high frequency region. This method is developed to enhance the telephone quality speech, however it may also be used for speech signals which have lower bandwidths (1.8 kHz).

Keywords: Spectral envelope extension, Spectral envelope extrapolation, Speech quality enhancement, Neural networks

1. Giriş

Günümüzde, insan tarafından kullanılan en temel iletişim aracı konuşmadır. Bu sebeple, konuşmayı bir yerden diğerine iletmek için pek çok konuşma kodlama ve iletim tekniği geliştirilmiştir. Bunlardan en temel ve basit olan ve tüm telefon sistemlerinde de kullanılan teknikte, konuşma yaklaşık 3.4 kHz'lik bir alçak geçirgen süzgeçten geçirildikten sonra 8 kHz'de örneklenir ve her bir örnek 8 bit ile nicemlenerek 64 kbit/s veri hızında sayısal konuşma elde edilmiş olur. Bu teknikte elde edilen konuşma orta-yüksek kalitede olmakta ve "telefon kalitesi" adı ile anılmaktadır. Günümüzde bant genişliğinden tasarruf etmek için konuşma daha düşük bit hızlarında iletilmektedir. Daha düşük bit hızları sağlamak için son yıllarda basitten karmaşığa pek çok konuşma kodlayıcı geliştirilmiştir. 64 kbit/s veri hızındaki orijinal konuşmaya yakın kalitedeki konuşma günümüzde 2.4 kbit/s civarındaki hızlarla elde edilebilmektedir. Ancak 3.4 kHz bant genişliği ve 8 kHz örnekleme hızı "telefon kalitesindeki" konuşma kodlayıcılarında değişmeden kalmıştır.

Konuşmanın yüksek kalitede sayılabilmesi için 7 kHz'lik bir alçak geçirgen süzgeç ile filtrelenip 16 kHz'le örneklenmesi ve sonrasında da 13-16 bit ile nicemlenmesi gerekir [1]. 3.4 kHz'e bant sınırlanmış, orta-yüksek kalitedeki konuşmada 3.4-7 kHz arasında kalan kısım yok sayılmaktadır. Yok sayılan bu kısım konuşmanın anlaşılabilirliğini pek fazla etkilememekte fakat kalitesinin bir miktar düşmesine sebep olmaktadır. Bazı kaynaklarda bu işlemin konuşmada % 10-20 arasında bir kalite kaybına yol açtığı belirtilmektedir [2]. Konuşma kalitesini artırmaya yönelik literatürde birçok çalışma bulunmakla birlikte bu çalışmaların hemen hemen tamamında gürültü azaltma ve mikrofon karakteristiği düzeltme gibi konular üzerinde durulmuştur. Alçak geçirgen filtre kaynaklı kalite kayıpları üzerine odaklanan pek fazla çalışmayla karşılaşılmamıştır. Karşılaşılan çalışmalar ise başlıca iki grup altında toplanabilir.

Birinci gruptaki çalışmalarda geniş bantlı konuşma ya frekans ekseninde sıkıştırılır ya da yüksek frekans bileşenlerine ait bir takım yan bilgiler dar bantlı konuşma ile beraber gönderilir. Bu gruba örnek çalışmalardan bir tanesi yüksek bant genişliğine sahip olan konuşmanın frekans ekseninde sıkıştırılıp daha

düşük bant genişliğinde (3.4 kHz) konuşmanın elde edilmesi ve elde edilen sıkıştırılmış konuşmanın klasik ses iletim ve kodlama teknikleriyle kodlanmasıdır [3]. Bir diğer çalışmada ise yüksek frekansa ait çok kısıtlı bilginin (150 b/s) dar bantlı konuşmayla beraber gönderilerek konuşma kalitesinin artırılmasıdır [4]. Bu gruba örnek verilebilecek diğer çalışmalarda ise yüksek frekans bileşenlerine ait yan bilgiler sıkıştırılıp gizli damgalama teknikleriyle dar bantlı konuşma içine gömülerek karşı tarafa iletilir ve ilgili alıcılarda bu bilgiler kullanılarak kaliteli konuşma elde edilir [5,6].

İkinci gruptaki çalışmalarda ise dar bantlı konuşma, iletim kanalında herhangi bir değişikliğe uğramadan olduğu gibi iletilir. Bu başlık altındaki çalışmalarda konuşma kalitesi artırma işlemleri alıcı tarafında ve konuşmanın özellikleri kullanılarak yapılır. Esas olarak bu çalışmalar geniş bantlı konuşmanın izgel zarfının elde edilmesi ve elde edilen bu zarf kullanılarak konuşmanın kayıp bölgelerinin yeniden oluşturulması veya sentezlenmesi üzerine kurulmuştur. Bu tür çalışmalara ilk örnekler izgel zarf genişletmede tablo yöntemini kullanan ve genişletilen izgel zarfın altının, kaynak-süzgeç yöntemiyle doldurulduğu çalışmalardır [7,8]. Daha sonraki çalışmaların bazılarında ise yüksek frekanslı bölgenin izgel zarfının altı çok bantlı uyarım [9] ve sinüs yöntemiyle doldurulmuştur [10]. Tablo yöntemiyle izgel zarf genişletmenin bir sonraki aşamasındaki çalışmalarda ise sesli ve sessiz konuşma bölgeleri için iki farklı tablonun kullanıldığı görülmüştür [11]. Daha sonra ardarda gelen konuşma parçalarının birbirine bağlı olduğundan yola çıkılarak Saklı Markov Modeli tablo yöntemiyle birlikte kullanılmaya başlanmıştır [12]. Bu yöntemlerin dışında ayrıca konuşma izgel zarfının konuşmanın özellikleri (perde, formant vb.) kullanılarak genişletildiği [13]; geniş bantlı konuşma parametrelerinin dar bantlı parametrelerden bir dönüşüm matrisi kullanılarak elde edildiği [14] ve uyarım sinyalinin harmonikliği parametrik olarak modelleyip konuşmanın bant genişliğini artıran çalışmalar da literatürde mevcuttur [15]. Son yıllarda yapay sinir ağlarıyla (YSA) konuşma bant genişliğini artıran çalışmalar literatürde yer almaya başlamıştır. Bunlara dar bantlı filtre bankalarını YSA kullanarak genişleten [16],[17] ile derin sinir ağları (DSA) kullanarak izgel bant genişleten [18] çalışmalar örnek verilebilir.

Konuşma izgel zarfını genişletme üzerine literatürde yer alan çalışmalar [19].’da özetlenmiştir.

Bu çalışmada, konuşmanın örtüşme önler süzgeç tarafından atılan yüksek frekans bölgesinin yine aynı süzgeç tarafından geçirilen düşük frekans bölgesinden yararlanılarak yeniden oluşturulması hedeflenmektedir. Konuşmanın izgel zarfını oluşturan insan ses üretim yolunun 250000-500000 kadar farklı olasılığa sahip olabileceği literatürde belirtilmektedir [20]. Konuşmanın filtreden geçirilerek kaybolan 3.4 kHz üstü bölgesi de bu sonlu sayıdaki ses yolu konfigürasyonu ile üretilmiştir ve dolayısıyla, bu bölgeye ait izgel zarf düşük frekans bölgesinin izgel zarfından elde edilebilir. Hatta yüksek frekans bölgesinin anlaşılabilirliğe katkısı çok daha az olduğundan bu bölge için daha az sayıda farklı konfigürasyon kullanılabilir. Yüksek frekans bölgesine ait izgel zarfın elde edilmesi işlemine bazı kaynaklarda izgel zarf dışdeğerleme, diğer bazı kaynaklarda ise izgel zarf genişletme adı verildiği görülmüştür. Bu yöntem kullanılarak kalite artırma işlemi sırasıyla iki ana başlık altında toplanabilir. Birinci başlıkta yüksek frekanslı izgel zarf düşük frekanslı izgel zarftan elde edilir; ikinci başlıkta ise elde edilen yüksek frekans bölgesine ait izgel zarfın altı doldurulur; bir başka deyişle yüksek frekans bileşenlerine sahip konuşma sentezlenir. Bu çalışmada konuşmanın izgel zarfının genişletilmesi için YSA; genişletilen izgel zarfın altında kalan bölgenin doldurulması için ise kaynak-süzgeç yöntemi kullanılmıştır.

Bu çalışma altı bölümden oluşmaktadır. Bölüm 2’de konuşmanın oluşumu ve modellenmesi ile ilgili kısa bilgi verilmiştir. Ayrıca izgel zarfın genişletilmesinde kullanılacak konuşma parametrelerinin elde edilmesi ve bu parametrelerin işlem yapmaya uygun gösterim biçimleri tartışılmıştır. Bölüm 3’te YSA kullanılarak izgel zarf genişletilmiştir. Burada kullanılan YSA’lar çok karmaşık bir yapıya sahip olmayan çok katmanlı sinir ağlarıdır ve bu ağlar geri yayılım algoritması kullanılarak eğitilmişlerdir. Bölüm 4’te ise YSA kullanılarak genişletilen zarfların altı doldurularak konuşma kalitesi artırılmıştır. Bu bölümde ayrıca genişletilen izgel zarfa ait kazanç parametresinin hesaplanma yöntemi ve kaynak süzgeç modelinde kullanılan uyarım sinyalinin genişletilmesi üzerinde de durulmuştur. Bölüm

5’te yapılan öznel testler anlatılmış ve son olarak da Bölüm 6’da sonuçlar değerlendirilmiştir.

2. Konuşmanın ve İzgel Zarfın Modellenmesi

İzgel zarf genişleterek konuşma kalitesini artırmak için öncelikle konuşma sinyalinin üretimi ve modellenmesi hakkında bilgiye sahip olmamız sonrasında da dar bantlı ve geniş bantlı konuşma parçalarının modellenmesi gerekmektedir.

İnsanlar konuşmak için akciğerlerinde biriken havayı kullanırlar. Konuşmanın olmadığı durumda, akciğerlerde biriken bu hava hiçbir engelle karşılaşmadan dışarı verilir. İnsanlar konuşacakları zaman bu havayı ses sistemlerinde çeşitli rezonans, türbülans, ya da patlamalar oluşturarak dışarı verirler [21]. Akciğerlerden sonra hava, herhangi bir değişikliğe uğramadan hançerenin sonundaki ses tellerine kadar gelir. Bundan sonra iki farklı durum söz konusudur. Birinci durumda, ses tellerinin titreşimi ile gelen hava akımına belli bir şekil kazandırılıp az çok periyodik dalga şekline sahip olan sesli (sedalı) seslerin üretilmesi sağlanır. İkinci durumda ise gelen hava akımı ses telleri titreşmediği için periyodik bir uyarıma dönüşmez ve böylece sessiz seslerin (sedasız) oluşumu sağlanır. Ses tellerinin saniyede yapmış olduğu titreşim miktarı perde frekansı, f_0 olarak adlandırılır. Konuşmanın, akciğerlerden gelen havanın ses üretim yolunu modelleyen bir süzgeci uyarmasıyla oluştuğu kabul edilen bu modelde, sesli ve sessiz sesler için sırasıyla dürtü dizisi ve beyaz gürültü olmak üzere iki farklı uyarım dizisi kullanılır [21]. Burada ses tellerinin etkisi $G(z)$, ağız ve geniz boşluğunun etkisi $V(z)$ ve dudakların etkisi $R(z)$ olarak tanımlanırsa ses üretim yolu sesli sesler için $H_v(z) = G(z)V(z)R(z)$ ve sessiz sesler için ise $H_u(z) = V(z)R(z)$ biçiminde modellenebilir. Konuşma sinyalinde sesli bölgeler sessizlere oranla çoğunluktadır ve sesli konuşma parçaları tüm kutup süzgeçle modellenir, bu yüzden ses üretim yolunu modelleyen süzgeç,

$$H(z) = \frac{g}{A(z)} = \frac{g}{1 + a_1z^{-1} + \dots + a_pz^{-p}} \quad (1)$$

biçiminde TK (Tüm Kutup) filtre olarak seçilir. Bu filtre katsayılarını hesaplamak için çeşitli hızlı yöntemler mevcuttur [21]-[23]. Burada g , filtre kazancı ve p ise TK filtrenin derecesidir. Filtre derecesi genelde $p=2B+(2$ ya da $4)$ biçiminde seçilir B ise konuşmanın kHz

cinsinden bant genişliğidir. Örnek olarak 3.4 kHz'e bant sınırlı 8 kHz ile örneklenmiş sesli bir konuşma parçasında 3-4 tane formant frekansı vardır. Her bir formantın 2. dereceden tüm kutup bir filtreyle modellenebileceği düşünülürse formantlar için 6-8 tane parametre gerekmektedir. Ayrıca 2-4 tane parametrenin de genel izgel dengeyi modelleyeceği düşünülürse, 8 kHz ile örneklenmiş konuşma için 10-12. dereceden bir TK filtrenin, 16 kHz ile örneklenmiş bir konuşma için ise 16-20. dereceden bir TK filtrenin yeterli olacağı görülür. Bunun yanında sessiz konuşma parçaları için formantlar tanımlı değildir. Ancak, sessiz konuşma parçaları konuşmada az oldukları ve 10-12. dereceden TK modellendiklerinde pek fazla hata oluşmadığı için, sessiz konuşma parçaları da TK modellenirler [21]. Fakat TK katsayıları nicemlenmeye ve kodlamaya elverişli değildir; örneğin bu katsayılar nicemlenirken meydana gelebilecek en küçük nicemleme hatası bile izgel zarfı büyük ölçüde etkileyebilir ve tüm kutup süzgeç kararsız olabilir. Dolayısıyla TK katsayılarının nicemlemeye ve kodlamaya daha elverişli logaritmik alan oranları [24], arcsinüs yansıma katsayıları [25] veya çizgi izgel frekanslar [26] cinsinden ifade edilmeleri daha uygun olur. Bu çalışmada TK katsayıları kodlamaya ve nicemlenmeye en uygun gösterim biçimi olan ÇİF'ler (çizgi izgel frekanslar) cinsinden ifade edilmiştir.

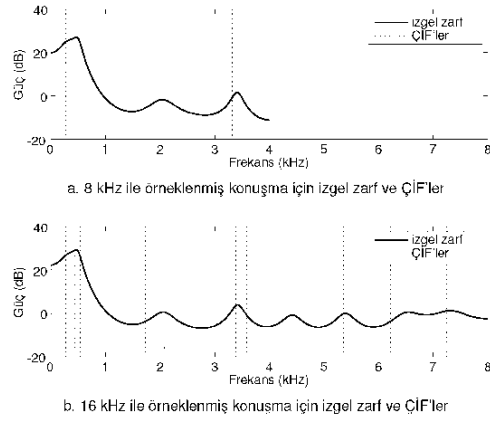
Konuşma yolu izgel zarfı, TK süzgeç katsayıları cinsinden ifade edildiğinde ve kazanç katsayısı bir olarak alındığında Denklem 1'deki $A(z)$, ters filtre olarak adlandırılır. Bu ters filtre kullanılarak ÇİF'leri bulmamızı sağlayacak $P(z)$ ve $Q(z)$ polinomları,

$$\begin{aligned} P(z) &= A(z) + z^{-(p+1)}A(z^{-1}) \\ Q(z) &= A(z) - z^{-(p+1)}A(z^{-1}) \end{aligned} \quad (2)$$

biçiminde ifade edilebilir [26]. ÇİF'ler işte bu $P(z)$ ve $Q(z)$ polinomlarının kökleri olarak tanımlanırlar ve hepsi birim çember üzerindedir. Ayrıca bu kökler birbirlerinin içine geçmiş ve artan bir sırada bulunurlar ve duyarlılıkları bölgeseldir. Son olarak ÇİF'lerden TK katsayılarına kayıpsız bir şekilde geri dönülebilir.

Yukarıdaki özelliklere uygun olarak 8 kHz ve 16 kHz'te örneklenmiş aynı konuşma parçalarına ait izgel güç yoğunlukları ve ÇİF'ler Şekil 1'de

görülebilir. Şekil dikkatli incelendiğinde iki ya da üç ÇİF'in bir formantın özelliklerini belirlediği ve düşük frekanslı bölgelerdeki ÇİF'lerin hemen hemen aynı değerlere sahip oldukları görülebilir. Bir ÇİF değerindeki değişiklik o bölgedeki izgel güç yoğunluğunu etkiler, ancak daha uzak bölgeler ise neredeyse bu değişimden hiç etkilenmemektedir. ÇİF'lerin tüm bu kullanışlı özelliklerine rağmen bu çalışma sırasında olumsuz bir özelliği ile karşılaşmıştır. Örneğin Şekil 1'de 16 kHz ile örneklenmiş konuşmaya ait ÇİF'lerin 4 kHz'in altında kalanlarının, 8 kHz ile örneklenmiş olan aynı konuşmaya ait olanlarla



Şekil 1. ÇİF'lerin izgel zarfta dağılımları

tamamen aynı olmadıkları görülebilir. Hatta başka konuşma örneklerinde 16 kHz ile örneklenmiş konuşmaya ait ÇİF'lerden 4 kHz'in altında kalanların sayısının 10'dan fazla ya da az olduğu durumlarla karşılaşmıştır. Bu durum bu bölümden sonra anlatılacak olan izgel zarf dışdeğerlemede birtakım problemler çıkarmıştır. Bu problemlerden en önemlisi yüksek frekanslı bölgeye ait ÇİF'lerin olmaları gereken bölgenin $[Fs/2-Fs]$ (Fs :örnekleme frekansı) dışına taşmalarıdır. Bu durumda sınır bölgelerindeki ÇİF'ler birbirleri içine geçebilmekte ve geri dönüşümde elde edilecek doğrusal öngörümlü kodlama (DÖK) katsayıları tamamen değişebilmekte ve hatta bazen DÖK katsayılarını elde etmek mümkün olmamaktadır.

3. İzgel Zarfın Genişletilmesi

Çalışmanın bu bölümünde geniş bantlı konuşma parametreleri yapay sinir ağları (YSA) kullanılarak elde edilmeye çalışılmış ve bu işlem için çok katmanlı yapıya sahip olan bir sinir ağı seçilmiştir. Sinir ağının girişi dar bantlı konuşma

parametreleri, çıkışı ise genişletilen zarfa ait parametrelerdir. Bu yapı basit ve anlaşılabilir olmakla birlikte aynı zamanda benzer işlemi yapan yöntemlerle karşılaştırılabilir sonuçlar verdiği için seçilmiştir. Burada önce YSA'lar ve kullanılan yöntem hakkında kısa bilgi verilecek daha sonra yöntemin uygulanış biçimleri ve sonuçlara yer verilecektir.

Yapay sinir ağları, insan beyninin özelliklerinden biri olan öğrenme yolu ile yeni bilgiler türetebilme, yeni ilişkiler kurabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacı ile geliştirilen bilgisayar sistemleridir. YSA, olayların örneklerine bakmakta, onlardan ilgili olay hakkında genellemeler yapmakta, bilgiler toplamakta ve daha sonra hiç görmediği örnekler ile karşılaştıkça öğrendiği bilgileri kullanarak o örnekler hakkında karar verebilmektedir. YSA'lar sahip oldukları özelliklerden dolayı, tanımlanan bir işi herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacı ile geliştirilen bilgisayar sistemleridir. Başka bir bakış açısıyla YSA teknolojisi insanlığın doğayı araştırma ve taklit etme çabalarının en son ürünlerinden bir tanesi olarak da yorumlanabilir ve son dönemde yapay zeka olarak da adlandırılmaktadır.

YSA'lar, basit biyolojik sinir sisteminin çalışma şeklinin benzetimini yapmak için tasarlanan programlardır. Benzetimi yapılan sinir sistemi birçok sinir hücresi (nöronlar) içerir ve bu sinir hücreleri çeşitli şekillerde birbirlerine bağlanarak sinir ağını oluştururlar. YSA'lar, ağırlıklandırılmış şekilde birbirlerine bağlanmış birçok işlem elemanlarından oluşan matematiksel sistemlerdir. Bir işlem elemanı, aslında aktarım işlevi (transfer fonksiyonu) olarak da anılan bir denklemdir. Bu işlem elemanı, diğer nöronlardan sinyalleri alır; bunları birleştirir, dönüştürür ve sayısal bir sonuç ortaya çıkartır.

YSA'lar, geleneksel işlemcilerden farklı şekilde işlem yapmaktadırlar. Geleneksel işlemcilerde, tek bir merkezi işlem elemanı her hareketi sırasıyla gerçekleştirir. YSA'lar ise her biri büyük bir problemin bir parçası ile ilgilenen çok sayıda basit işlem elemanlarından oluşmaktadır. En basit şekilde, bir işlem elemanı, bir girdiyi bir ağırlık kümesi ile ağırlıklandırır, doğrusal olmayan bir şekilde dönüşümünü sağlar ve bir

çıkı değerini oluşturur. İlk bakışta, işlem elemanlarının çalışma şekli oldukça basittir. Ancak, hesaplamaların gücü, toplam işlem yükünü paylaşan işlem elemanlarının birbirleri arasındaki yoğun bağlantı yapısından gelmektedir.

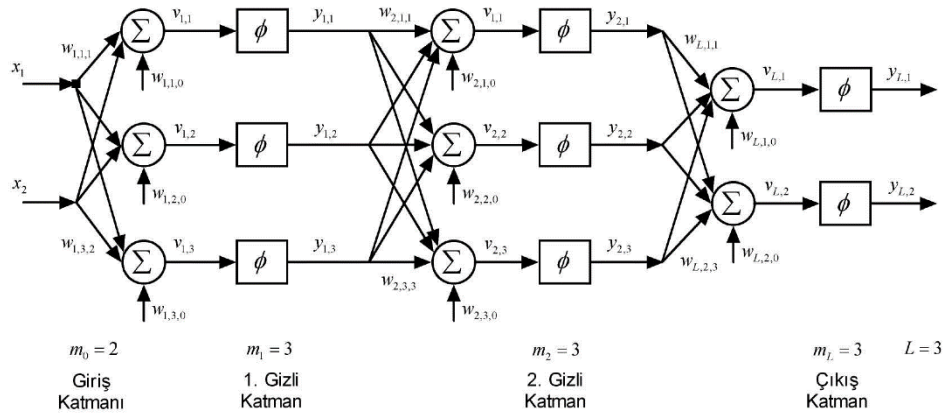
Çoğu YSA'da, benzer karakteristiğe sahip nöronlar tabakalar halinde yapılandırılırlar ve transfer fonksiyonları eş zamanlı olarak çalıştırılırlar. Hemen hemen tüm ağlar, giriş ve çıkış nöronlarına sahiptirler. YSA'nın ana ögesi olan matematiksel fonksiyon, ağın mimarisi tarafından şekillendirilir. YSA'ların davranışları, yani giriş ve çıkış arasında nasıl ilişkilendirdikleri, öncelikli olarak nöronların transfer fonksiyonlarına, ne şekilde birbirlerine bağlandıklarına ve bu bağlantıların ağırlıklarına bağlıdır.

Bu çalışmada YSA, düşük bant genişliğine sahip konuşmaya ait parametrelerden yüksek bant genişliğine sahip konuşma parametrelerini elde etmek için kullanılacaktır. Önceki çalışmalardan yola çıkılarak nasıl bir YSA seçilmesi gerektiği tam olarak bilinmemekle birlikte, bu çalışmada Şekil 2'deki gibi çok katmanlı bir YSA kullanılmıştır [23]. Kullanılan bu çok katmanlı yapı, geriye yayılım algoritmasıyla eğitilmiş ve dar bantlı konuşma parametrelerinden geniş bantlı konuşma parametrelerinin elde edilmesi için kullanılmıştır.

Çok katmanlı sinir ağları yapıları itibariyle çok basit fakat oldukça etkili ve kullanışlıdır. Bu ağlar birbirine ağırlıklandırılarak bağlanmış birçok nörondan oluşurlar. Çok katmanlı YSA'ların genellikle; giriş, gizli ve çıkış katmanları olmak üzere üç ana katmanı vardır. Bu çalışmada kullanılan sinir ağları öğreticiyle öğrenme yöntemi kullanılarak eğitilmiştir. Bu yöntemde geriye yayılım algoritması kullanılarak hesaplanan çıkış ile istenen değer arasındaki hata mümkün olduğunca küçültülmeye çalışılır.

N tane giriş ve çıkıştan oluşan bir eğitim setinde giriş verileri, $x(n)$, $n = 1, 2, \dots, N$ ve istenen çıkış verileri de $d(n)$, $n = 1, 2, \dots, N$ olsun. Bu durumda sinir ağının o anda hesaplanan çıkış verilerinin $y(n)$, $n = 1, 2, \dots, N$ olduğu durumda çıkıştaki karesel hata,

$$\epsilon(n) = \frac{1}{2} \|d(n) - y(n)\|^2 \quad (3)$$



Şekil 2. Kullanılan çok katmanlı yapay sinir ağı mimarisi.

şeklinde ve bu hatanın ortalaması da

$$\varepsilon_{avg} = \frac{1}{N} \sum_{n=1}^N \varepsilon(n) \quad (4)$$

kullanılarak hesaplanır. YSA'lar bu ortalama karesel hatayı en küçültecek biçimde nöronlar arasındaki bağlantıları değiştirerek eğitilirler. Eğitilen bu sinir ağı daha sonra tasarlanan sistem için kullanılabilirler. Bu çalışmada sinir ağının eğitilmesi için geri yayılım (back propagation) algoritması kullanılmıştır. Geri yayılım algoritması bayır inişi (steepest descent) temelli bir eğitim algoritmasıdır. Bu algoritmayı anlatmak için öncelikle Şekil 2'deki simge ve gösteriliş biçimlerinin açıklanması gerekmektedir [23].

- n , index numarası giriş ve çıkış verilerinin yerini göstermektedir. $x(n)$ ve $y(n)$ sırasıyla n 'inci giriş ve çıkışı göstermektedir.
- k , yineleme sayısını göstermektedir.
- l , katman sayısını göstermektedir, $l=0$ giriş katmanını ve $l=L$ ise çıkış katmanını belirtmektedir.
- i , girişi belirtmek için, j , ise nöron numarasını belirtmek için kullanılmaktadır.
- m_l l 'inci katmandaki nöron sayısını belirtmek için kullanılır. Burada m_L çıkış katmanındaki m_0 ise giriş katmanındaki nöron sayısını belirtir.
- $w_{l,j,i}$, l 'inci katmandaki j 'inci nörona, $(l-1)$ 'inci katmandaki i 'inci nörondan gelen ağırlığı belirtir. Eğitim sırasında n 'inci giriş göre ağırlıklardaki değişim $\Delta w_{l,j,i}(n)$ biçiminde gösterilebilir ve bu durumda k 'inci

yinelemedeki ağırlık $w_{l,j,i}^{[k]}$ olur. l 'inci katmandaki ağırlıklar $m_l \times (1 + m_{l-1})$ boyutunda bir matris ile ifade edilebilir. Bu matristeki kolon sayısının bir fazla olması her bir nörona sabit değer olarak 1'in giriş olarak kabul edilmesindedir.

Eğitim algoritmasındaki temel amaç ortalama hatayı en küçültmektir. Ortalama karesel hata ise ağdaki tüm ağırlıkların bir işlevidir ve dolayısıyla bu hatanın küçültülmesi için tüm ağırlıkların her bir yinelemede (iteration) güncellenmesi gerekmektedir. Bu güncelleme işlemi ise n 'inci giriş için bayır inişi yöntemini kullanarak

$$w_{l,j,i}^{[k+1]} = w_{l,j,i}^{[k]} - \mu \frac{\partial \varepsilon(n)}{\partial w_{l,j,i}^{[k]}} \quad (5)$$

biçiminde yapılır ve burada μ öğrenme parametresi olarak tanımlanır. Bu güncelleme işlemi her bir $n = 1, 2, \dots, N$ giriş kümesi için ağırlıklar bir sonuca yakınsayınca kadar ve muhtemelen birden fazla kere devam eder. Bu algorithmada öğrenme parametresinin uygun şekilde seçilmesi kritik noktalardan bir tanesidir. Eğer μ olması gerekenden büyük bir değer seçilirse algorithmadaki adımlar çok büyük olur ve ağırlıklar bir yere yakınsayamaz; tam tersi durumda ise adımlar çok küçük olur ve bölgesel en uç noktaları geçemezler ve dolayısıyla ağırlıklar yine olması gereken yere yakınsayamazlar. Diğer bir kritik nokta ise kısmi türevlerin $\partial \varepsilon(n) / \partial w_{l,j,i}^{[k]}$ alınmasıdır. Çıkış katmanı ile saklı katmanlardaki türevlerin alınmasında birtakım farklılıklar vardır, çünkü saklı katmanlarda olmayan istenen değer, $d_j(n)$

çıkış katmanında mevcuttur. Dolayısıyla son katmandan başlayarak geriye doğru kısmi türevler hesaplanabilir.

Bu çalışmada YSA'lar dar bantlı konuşma izgel zarfının genişletilmesi için kullanılacaktır. Şekil 2'de görülen ve sinir ağlarının giriş ve çıkışlarını gösteren yapının girişleri TK modellenmiş konuşma parametrelerinden elde edilmiş ÇİF'ler, çıkışları yine TK modellenmiş geniş bantlı konuşma parametrelerinden elde edilen ÇİF'in yüksek frekans bölgelerine ait olanlarıdır. Telefon kalitesindeki konuşma parçaları 10. dereceden ve yüksek kaliteli konuşma parçaları ise 20. dereceden TK modellenir. Dolayısıyla Şekil 2'de görülen ve sinir ağlarının giriş ve çıkışlarını gösteren yapının girişleri 10. dereceden TK parametrelerinden elde edilmiş ÇİF'ler, çıkışları ise 20. dereceden TK parametrelerinden elde edilen 20 tane ÇİF'in son 10 tanesidir. Burada kullanılan çok katmanlı sinir ağında 10 nöronluk giriş ve yine 10 nörondan oluşan çıkış katmanının yanında 30 nörondan oluşan bir gizli katman da bulunmaktadır. Bu YSA, önce eğitilmiş daha sonra geniş bantlı izgel zarf parametrelerinin elde edilmesi için kullanılmıştır. Bu ağın eğitimi için toplam kırk dakikalık bir konuşma veri tabanı yeterli olmuştur. Bu veri tabanındaki konuşma parçalarından yaklaşık 100,000 tane eğitim vektörü elde edilmiştir. Bu eğitimin sonunda kullanılan eğitim parametrelerinden bağımsız olarak ortalama karesel hata 0.004'ten küçük bulunmuştur. Bu yöntemle genişletilen ve eğitim setinde olmayan konuşma parçaları için [27]'de anlatılan ortalama izgel bozulum yaklaşık olarak 3.8 dB olarak bulunmuştur. Bilgi teorisi bakış açısıyla 3.4 kHz'e bant sınırlı olan telefon kalitesindeki konuşma parametreleri ile 7 kHz'e bant sınırlı olan konuşma parametreleri karşılıklı olarak 2.3-2.4 bit bilgi içerirler [28]. Bu bilgiden yola çıkılarak izgel zarflar arasında oluşan bozulumun yaklaşık 3 dB'den küçük olamayacağı sonucuna varılır [28]. Bu koşullar altında konuşmalar için bulunan 3.8 dB'lik izgel bozulum teorik sınırdan büyük olmakla birlikte olabilirlik sınırları içindedir. Eğitim setindeki konuşma parçaları için ortalama izgel bozulum yaklaşık olarak 3 dB bulunmuştur. Bu da teorik üst sınıra karşılık gelmektedir.

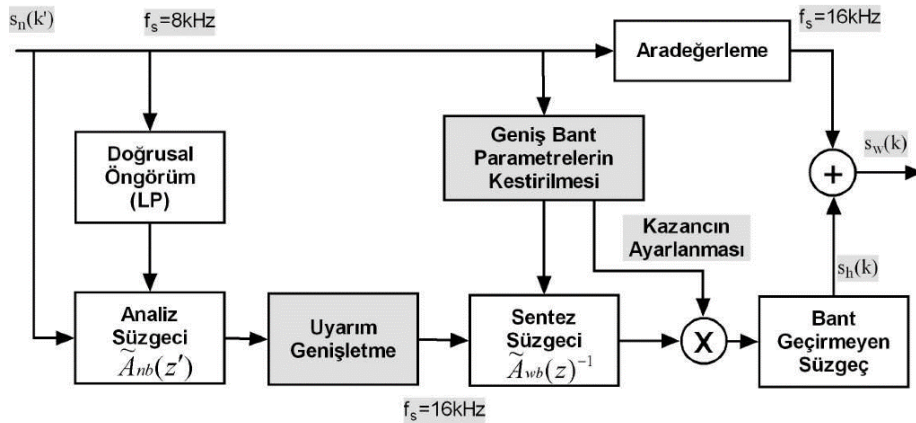
Bu yapı ile genellikle, 8 kHz ile örneklenmiş 3.4 kHz ile bant sınırlı konuşma parçalarının bant genişliklerini 7 kHz'e çıkararak kalitelerinin artırılması hedeflenmekle birlikte daha düşük

bant genişliğine sahip konuşma parçaları için de kullanılabilir. Burada, önceki çalışmalara ek olarak benzer bir yapı 4 kHz ile örneklenmiş 1.8 kHz'e bant limitli olan konuşmaları 3.5 kHz bant genişliğine çıkarmak için gerekli izgel zarfları genişletmekte de kullanılmıştır. Bunun için öncelikle 1.8 kHz'e bant sınırlı konuşma 6. dereceden TK modellenip parametreleri bulunmuş ve bunlar Şekil 2'de görülen sinir ağının girişleri olarak alınmışlardır. Sinir ağının çıkışı ise telefon kalitesine sahip konuşmaya ait 10. dereceden parametrelerin son 4 tanesidir. Bu durumda giriş katmanı 6 nöron, çıkış katmanı 4 nöron ve gizli katman ise 12-16 nörondan oluşacak biçimde seçilmiştir. Bu sinir ağı yapısı da bir öncesine benzer bir şekilde ve aynı eğitim setiyle eğitilmiş ve ortalama bozulum yaklaşık olarak 3.9 dB bulunmuştur.

4. Konuşma Kalitesinin Artırılması

Çalışmanın bundan önceki bölümünde sinir ağları kullanılarak, dar bantlı konuşmanın izgel zarfından geniş bantlı konuşmaya ait izgel zarf elde edilmiştir. Bu bölümde genişletilen izgel zarflar Kaynak-Süzgeç Modelinde kullanılacak ve yüksek frekans bileşenlerine sahip daha kaliteli konuşmalar elde edilmeye çalışılacaktır. Ayrıca Kaynak-Süzgeç modeli kullanılarak konuşma kalitesinin artırılması sırasında gerekli olan geniş bantlı konuşma için uyarım sinyalinin ve kazancın bulunma yöntemleri hakkında bilgi verilecektir.

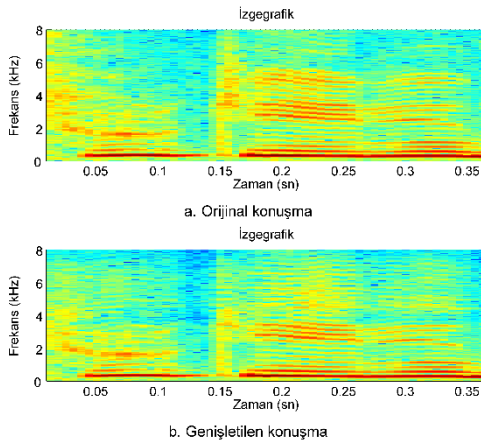
Kaynak-süzgeç modeliyle yüksek frekanslı konuşma bölgelerinin elde edilme işleminin genel işleyiş şeması Şekil 3'te verilmiştir. Bu yöntemde önce düşük bant genişliğine sahip konuşmaya ait TK parametreleri bulunmuş ve konuşma ters süzgeçten geçirilip uyarım sinyali de ayrıca elde edilmiştir. Daha sonra zarf ve uyarım sinyali genişletilmiştir. Ardından uyarım sinyali ve genişletilen TK parametreleri kullanılarak geniş bantlı konuşma sinyali elde edilmiştir. Ancak düşük bant genişliğine sahip orijinal konuşma zaten elde olduğu için ihtiyaç olan yüksek frekanslı bölge, elde edilen geniş bantlı konuşmanın bir yüksek geçiren bir süzgeçle süzülmesiyle elde edilmiştir. Bundan sonra sadece yüksek frekans bileşenleri dar bantlı konuşmaya eklenip yüksek kaliteli konuşma elde edilmiştir. Şekil 4'te bu yöntemle genişletilmiş ve orijinal konuşmalara ait izgegrafikler (spectrogram) görülmektedir.



Şekil 3. Dar bantlı konuşmanın kaynak-süzgeç modeliyle genişletilmesinin genel işleyiş şeması

4.1. Uyarım sinyalinin genişletilmesi

Konuşma kalitesini artırmak için eklenecek yüksek frekans bileşenlerinin elde edilmesi için TK parametrelerinin yanında uyarım sinyalinin de genişletilmesi gerekmektedir. Bunun için öncelikle uyarım sinyali hakkında ön bilgiye sahip olmak gerekmektedir. Şekil 4a'da 16 kHz ile örneklenmiş konuşma parçasının izgeografından de anlaşılacağı gibi sesli konuşmanın 5.5 kHz'ten sonra hemen hemen hiç harmonik bileşeni yoktur. Ayrıca uyarım sinyalinin izgel zarfının da tüm frekans bandı boyunca sabit olduğu düşünülebilir.



Şekil 4. 16 kHz ile örneklenmiş orijinal ve genişletilmiş konuşma parçaları

Yüksek bant genişliğine sahip izgel zarfı elde etmek için, orijinal konuşmadan elde edilen uyarım sinyaline yüksek frekanslı bileşenlerin eklenmesi gerekmektedir. Bunun için öncelikle

dar bantlı uyarım sinyali yukarı örneklenip aradeğerlenerek örnek sayısı artırılmıştır. Daha sonra, bu sinyale yüksek frekans bileşenlerine sahip sinyal eklenerek geniş bantlı uyarım sinyali elde edilmiştir. Bu çalışmada dar bantlı uyarım sinyali için bulunan sinüsler yüksek frekans bandına kaydırılarak dar bantlı uyarım sinyaline eklenmiş ve böylece geniş bantlı uyarım sinyali elde edilmiştir. Konuşmanın sinüslerin toplamı biçiminde ifade edilebilmesi aslında uyarım sinyalinin sinüslerin toplamı biçiminde ifade edilebilmesine dayanmaktadır. Burada bu iki temelden yola çıkarak dar bantlı uyarım önce sinüslerin toplamı olacak

$$E_L[n] = \sum_{l=1}^{L^k} A_l^k \cos(n \omega_l^k + \theta_l^k) \quad (6)$$

biçiminde modellenmiş ve tüm sinüs parametreleri bulunmuştur [26]. Daha sonra bu sinüsler yüksek frekans bandına kaydırılarak toplanmış ve böylece uyarım sinyalinin yüksek frekanslı bölgeleri aşağıdaki denklem kullanılarak elde edilmiştir.

$$E_H[n] = \sum_{l=1}^{L^k} A_l^k \cos(n (\omega_l^k + \omega_s) + \theta_l^k) \quad (7)$$

Burada ω_s , frekans eksenindeki kaymayı gösterir ve genelde dar bantlı konuşmanın bant genişliği kadar seçilir. Örneğin 8 kHz ile örneklenmiş konuşma için ω_s , 3.5 kHz civarında seçilebilir. Bu çalışmada frekansların devamlılığını sağlamak için, ω_s dar bantlı konuşmanın kestirilen en yüksek frekansı olarak seçilmiştir. Ardından bu yüksek frekanslı sinyale dar bantlı uyarım sinyali

eklenerek, k çerçevesi için geniş bantlı uyarım sinyali

$$E_{EX}^k[n] = E_L^k[n] + E_H^k[n] \quad (8)$$

biçiminde bulunur. Sessiz konuşma bölgeleri için elde edilen bu genişletilmiş uyarım sinyaline (uyarım sinyalinin gücüyle orantılı olacak biçimde) gürültü de eklenebilir. Eğer konuşma 3.4 kHz'ten 7 kHz'e genişletilecekse eklenecek gürültü [5.5 – 7] kHz bandında olmalıdır.

4.2. Geniş bantlı konuşma için kazancın bulunması

Konuşma Eş.1'deki gibi TK modellendiğinde g , kazanç parametresi olarak tanımlanır. Konuşmanın sessiz olduğu bölgede kazanç yüksek, sesli olduğu bölgelerde ise kazanç düşük değerler alır. Geniş bantlı konuşmayı elde etmek için izgel zarf ve uyarım sinyaline ek olarak işte bu kazanç parametresinin de geniş bantlı konuşma için bulunması gerekmektedir. Burada öncelikle m'inci çerçeve için dar bantlı konuşmaya ait ortalama izgel güç yoğunluğu $\bar{\Phi}_n(\omega; m) \approx g_n^2 / |\bar{A}_n(\omega; m)|$ biçiminde ifade edilebilir. Ardından da yine aynı çerçeve için geniş bantlı (genişletilmiş) konuşmaya ait ortalama izgel güç yoğunluğu $\bar{\Phi}_e(\omega; m) \approx g_e^2 / |\bar{A}_e(\omega; m)|$ şeklinde ifade edilir. Daha sonra orijinal dar bantlı konuşmaya ait ortalama izgel güç yoğunluğu ile genişletilmiş konuşmaya ait izgel güç yoğunluğunun düşük bantlı kısmı birbirine eşitlenip g_e parametresi

$$g_e = \sqrt{\frac{|\bar{\Phi}_n(\omega; m)|^2}{|\bar{\Phi}_e(\omega; m)|^2}} g_n \quad (9)$$

biçiminde bulunabilir. Burada $\bar{\Phi}_n(\omega; m)$ ve $\bar{\Phi}_e(\omega; m)$ darbantlı ve genişletilmiş zarfın dar bant kısmını gösteren düzleştirilmiş ortalama izgel güç yoğunluklarıdır.

5. Özel Testler

Bundan önceki bölümlerde önce konuşmanın izgel zarfı genişletilmiş daha sonra ise bu zarf kullanılarak konuşma kalitesi artırılmıştı. Ancak kalitesi artırılan konuşmaların başarımının ölçülmesi için nesnel bir başarıım ölçütü bulunmamaktadır. Nesnel başarıım ölçütü olarak sadece izgel zarfların birbirine yakınlıklarını belirten izgel bozulum ölçütü kullanılmaktadır. Bu ölçütün konuşma kalitesiyle ilişkisi olmakla beraber aralarında tam olarak birebir bir ilişki yoktur. Bunun sebebi ise konuşma kalitesinin

kullanılan sentez yöntemiyle ve dinleyiciden dinleyiciye değişmesidir. Günümüzde, konuşma kalitesini ölçmek için Ortalama Kanı Puanı (Mean Opinion Score, MOS) testi kullanılmaktadır [27]. Bu çalışmada kullanılan OKP (Ortalama Kanı Puanı) testinde konuşma 1 ile 5 (5 en iyi olmak üzere) arasında derecelendirilmiştir. Sonuçları değerlendirmek için kullanılan bir diğer testte ise dinleyicilere, dinledikleri konuşma parçaları arasındaki tercihleri sorulmuştur.

5.1. Ortalama kanı puan testleri

Bu test sırasında dinleyicilere ikisi kadın ikisi erkek olmak üzere toplam dört farklı konuşma parçasının her biri için beş tane olmak üzere toplam 20 tane test cümlesi dinletilmiştir. Bu konuşma parçalarından üçü televizyon haberlerinden biri ise elde bulunan ses kayıt sistemiyle elde edilmiştir. Televizyondan kaydedilen sesler bir miktar gürültülü, diğer konuşma ise çok kalitelidir. Dinletilen her bir konuşma parçasının örnekleme frekansı ve izgel zarfının genişletilmiş olup olmadığı Tablo 1'de belirtilmiştir. Bu tabloda görülen 4 kHz ve 8 kHz ile örneklenen işlenmemiş konuşmalar, 16 kHz ile örneklenen geniş bantlı konuşmadan örtüşme önler süzgeçle süzülüp örnekleme frekansları düşürülerek elde edilmişlerdir. Diğer test konuşmaları ise yarı örnekleme hızındaki konuşmalardan Şekil 3'te ayrıntılı biçimde gösterilen yöntem kullanılarak kalitesi artırılan konuşmalardır. Testlerde kullanılan tüm izgel zarflar sinir ağları yöntemi kullanılarak genişletilmişlerdir.

Dinleyicilere, gürültülü olmayan ofis ortamında sessiz çalışan bir dizüstü bilgisayardan Philips SBC HP200 kulaklığı kullanılarak konuşma parçaları dinletilmiş ve konuşmalara 1 ile 5 arasında, 0.25 hassasiyette puanlar vermeleri istenmiştir. Bir kişi için ortalama kanı puanı (OKP) testinin yapılması ortalama 20 dakika sürmüştür. Daha önceden herhangi bir işitme testi yapılmamış 20 ile 40 yaşları arasında 15 dinleyiciyle yapılan OKP testinin sonucu Tablo 1'de görülmektedir.

Tablo 1'in son sütunu tüm konuşma parçaları için ortalama puanı göstermektedir. Bu sütun dikkatle incelendiğinde 1.8 kHz'e bant sınırlı 4 kHz ile örneklenen konuşmanın (S_4) izgel zarfı genişletilip kalitesi artırıldığında elde edilen konuşmanın ($S_{8,ka}$), 1.8 kHz'e bant sınırlı konuşmaya göre daha tercih edilir olduğu

Tablo 1. 15 dinleyici için OKP sonuçları.

Konuşma ve Örneklem Frekansı	Yöntem	1.Cümle	2. Cümle	3. Cümle	4. Cümle	Ortalama
S_4	Yok	2,38	2,47	2,25	2,53	2,41
$S_{8,ka}$	Kaynak-Süzgeç	2,63	2,25	2,50	3,12	2,63
S_8	Yok	3,68	3,52	3,65	3,90	3,69
$S_{16,ka}$	Kaynak-Süzgeç	3,65	3,58	3,40	4,32	3,74
S_{16}	Yok	4,01	3,95	3,92	4,60	4,12

söylenbilir. Aynı işlemler uygulanarak kalitesi artırılmaya çalışılan telefon kalitesindeki 3.4 kHz'e bant sınırlı konuşma için ise aynı şeyi söylemek her zaman mümkün değildir. Çünkü; birinci ve üçüncü konuşma parçaları için 3.4 kHz'e bant sınırlı işlenmemiş konuşma parçalarının OKP'si zarfı genişletilerek kalitesi artırılmış konuşmalarinkinden daha yüksektir. Ancak ikinci ve dördüncü konuşma parçaları için kalitesi artırılan konuşmaların kalitesi, 3.4 kHz'e bant sınırlı konuşmanın kalitesinden çok daha yüksek bulunmuştur. Bu durumda kaynak süzgeç modeli ile kalitesi artırılan konuşmaların ($S_{16,ka}$), orijinal 3.4 kHz'e bant sınırlı konuşmalara (S_8) göre dinleyiciler tarafından daha tercih edilir oldukları sonucuna varılır. Dinleyicilerle testler bittikten sonra yapılan değerlendirmelerde, televizyondan kaydedilen konuşmalardaki gürültünün kendilerini rahatsız ettiğine ve ayrıca bu gürültünün, kalitesi artırılan konuşmalara olumsuz etki ettiği sonucuna varılmıştır. Ancak tüm bunlara rağmen elde edilen OKP sonucu [13]'teki sonuçlarla paralel çıkmıştır.

5.2. Karşılaştırmalı testler

Karşılaştırmalı testler, OKP testlerinde de kullanılan 4 test konuşma parçası kullanılarak gerçekleştirilmiştir. Karşılaştırma testleri, 1.8 kHz ile bant sınırlı konuşmanın izgel zarfı genişletilerek 3.4 kHz'e çıkarılması ve 3.4 kHz'e bant sınırlı işlenmemiş konuşmanın bant genişliğinin 7 kHz'e çıkarılması için ayrı ayrı yapılmıştır. Her iki testte de dört test cümlesi için dar bantlı konuşma, genişletilmiş konuşma ve işlenmemiş geniş bantlı konuşma, 16 dinleyiciye dinletilmiştir. 1.8 kHz'e bant sınırlı ve bu

konuşmadan izgel zarf genişletilerek kalitesi artırılan konuşma için dinletilen toplam 64 test cümlesinden 50 (%78) tanesinde kalitesi artırılan konuşma, dar bantlı konuşmaya göre dinleyiciler tarafından tercih edilmiştir. Aynı test 3.4 kHz'e bant sınırlı konuşma için yapıldığında ise kalitesi artırılan konuşmanın tercih edilme sayısı ise 21'de (%32) kalmıştır. Ancak tercih edilen bu 21 cümlelerin 10 tanesi sadece gürültüsüz konuşma için gerçekleşmiştir. Bu durumda gürültüsüz konuşma için frekans bandı genişletilen konuşmanın dar bantlı konuşmaya göre tercih edilme oranı %62 olmuş ve bu sonuç [17]'de elde edilen %82'lik tercih elde edilme oranından düşük olmakla birlikte yine de kabul edilebilir bir sonuç olarak değerlendirilebilir.

6. Tartışma ve Sonuç

Bu çalışmada, telefon ve daha düşük bant genişliğine sahip konuşmaya ait izgel zarfı genişletilerek konuşma kalitesi artırılmıştır. İzgel zarfı genişletme işlemi için sinir ağları yöntemi kullanılmıştır. Elde edilen sonuçlar literatürde dar bantlı konuşmanın izgel zarfını genişletme üzerine yapılan az sayıdaki diğer çalışmayla benzer sonuçlar vermiştir [11],[28]. Çalışmada, yapay sinir ağlarıyla telefon kalitesindeki konuşma genişletilmiş olmakla birlikte önerilen yöntemin daha dar bantlı konuşmalar için de (örneğin 1.8-2 kHz) kullanılabileceği gösterilmiştir.

İzgel zarfı genişletilirken performans ölçütü olarak ortalama karesel izgel bozulum kullanılmıştır ve telefon kalitesindeki konuşma için ortalama bozulum 3.8 - 3.9 dB aralığında bulunmuştur. Bulunan bu bozulum 3 dB [28]

olan teorik sınırdan büyüktür ancak kabul edilebilir sınırlar içerisinde ve benzer çalışmalardakilerle paralellik içindedir [11],[28]. Ayrıca, konuşmanın kalitesi elde edilen izgel zarfın orijinal zarfa yakınlığıyla orantılı olmakla birlikte kalite artırmada kullanılan sentezleme yöntemine de bağlıdır.

Konuşma kodlama ve kalite artırma için yapılan çalışmalarda konuşmanın dar bantlı kısımları için izgel bozulunun 1 dB ve yüksek bantlı kısımları için 2.5 – 3 dB olmasının konuşma kalitesinde izgel zarftan kaynaklanan büyük bir kalite kaybına yol açmayacağı belirtilmiştir [1].

İzgel zarf genişletmedeki amaç dar bantlı konuşmanın kalitesinin iyileştirilmesidir. Elde edilen bu yüksek frekans bileşenlerine sahip izgel zarf, kaynak-süzgeç modelinde kullanılarak konuşma kalitesi artırılmıştır. Yapılan öznel testlerde elde edilen yüksek frekans bileşenlerine sahip konuşma parçaları için konuşma kalitesi özellikle 1.8 kHz'ten bant genişliği genişletilen konuşmalar için hissedilir derecede artmıştır. Telefon kalitesindeki konuşmalar için gürültülü konuşma bölgelerinde konuşma kalitesi bir miktar düşmüş, gürültüsüz bölgelerde ise kalitede önemli bir miktar artış görülmüştür. Ayrıca dinleme testlerinde kulaklık kullanılmıştır. Kulaklıkla dinlenen konuşma

parçalarındaki çok az gürültü bile dinleyicileri rahatsız etmiş ve test cümlelerine verdikleri puanlar düşmüştür. Aynı konuşma parçaları kaliteli bir hoparlör kullanılarak dinletildiğinde dinleyicilerin gürültüden daha az rahatsız oldukları görülmüş ve konuşmaların dinleyicilerden aldıkları puanlar artmıştır.

Bu çalışmada önerilen konuşma kalitesi artırma yöntemlerinde herhangi bir ek bilgi iletimi gerekmediğinden iletim bant genişliğini artırıcı etkileri yoktur. Konuşmanın kodlanmış olup olmadığı, kodlandığı da hangi yöntemle kodlandığı önem taşımamaktadır. Günümüzde telefon kalitesindeki konuşmaları kodlamak için geliştirilen kodlama tekniklerinde genelde 10. dereceden DÖK parametreleri kodlanarak alıcı tarafa iletilmektedir. 1.8 kHz'e bant sınırlı konuşmaların bant genişliğinin artırılması sırasında kullanılan izgel zarf genişletme yönteminin bu kodlama teknikleriyle beraber kullanılarak 10 yerine 6 tane DÖK parametresinin karşı tarafa gönderilebileceği ve dolayısıyla daha düşük bit hızlarında konuşma kodlamanın mümkün olacağı düşünülmektedir. Yöntemin, tüm bu avantajları sayesinde dar bantlı konuşma iletiminde yaygın kullanım alanları bulabileceği değerlendirilmektedir.

Kaynakça

- [1] European Telecommunications Standards Institute (ETSI), "Speech Processing Transmission & Quality Aspects (STQ) QoS parameter definitions and measurements," ETSI, ETSI EG 201 769-1 [online], http://www.etsi.org/deliver/etsi_eg/201700_201799/20176901/01.01.01_60/eg_20176901v010101p.pdf, v1.1.1, 2000.
- [2] O'Shaughnessy, D. 1987. *Speech Communications - Human and Machine*. Addison-Wesley, Reading, Massachusetts.
- [3] Patrick, P. J., Steele, R. ve Xydeas, C. S., 1993. Frequency compression of 7.6 kHz speech into 3.3 khz bandwidth, *IEEE Trans. on Comm.*, Cilt. 31(5), s. 692-701. DOI: 10.1109/TCOM.1983.1095876.
- [4] Epps, J. ve Holmes, W. H. 2001. A new very low bit rate wideband speech coder with a sinusoidal highband model, *IEEE International Symposium on Circuits and Systems ISCAS*, Sydney.
- [5] Geiser, B. ve Vary, P. 2007. Backwards compatible wideband telephony in mobile networks: Celp watermarking and bandwidth extension, *Proc. of ICASSP*, Honolulu.
- [6] Prasad, N ve Kumar, T. K. 2016. Speech bandwidth extension using magnitude spectrum data hiding, *Int. Conf. on Computing, Analytics and Security Trends (CAST)*, Pune.
- [7] Carl, H., Heute U. 1994. Bandwidth enhancement of narrowband speech signals, *Proceedings of EUSIPCO, VII European Signal Processing Conference*, Edinburg, 1178-1181.
- [8] Yoshida, Y., and Abe, M 1994. An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping, *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, Yokohama, 1591-1594.
- [9] Chan, C-F., and Hui, W-K. 1996. Wideband enhancement of narrowband coded speech using MBE re-synthesis, *Proc. Of 3rd Int. Conf. on Signal Processing (ICSP)*, Beijing.
- [10] Epps, J., Holmes, W. H. 1998. Speech enhancement using STC-based bandwidth extension, *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, Sidney.
- [11] Epps, J., Holmes, W. H. 1999. A new technique for wideband enhancement of coded narrowband speech, *Proc. of IEEE Workshop on Speech Coding*, Porvoo. DOI: 10.1109/SCFT.1999.781522.
- [12] Jax, P., Vary, P., 2000. Wideband extension of telephone speech using a hidden markov model, *Proc. of IEEE Workshop on Speech Coding*, Delevan. DOI: 10.1109/SCFT.2000.878427.
- [13] Gustafsson, H., Lindgren, U.A., Claesson, I. 2006. Low-complexity featuremapped speech bandwidth extension, *IEEE Trans. on Audio, Speech and Language Processing*, Cilt. 14(2), pp. 577-588. DOI: 10.1109/TSA.2005.855837.

- [14] Chennoukh, S., Gerrits, A., Miet, G., Sluijter, R. 2001. Speech enhancement via frequency bandwidth extension using line spectral frequencies, Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP-2001, Salt Lake City, DOI: 10.1109/ICASSP.2001.940919.
- [15] Vaseghi, S., Zavarehei, E., Qin, Y. 2006. Speech bandwidth extension: Extrapolations of spectral envelop and harmonicity quality of excitation, Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing ICASSP-2006, Toulouse, DOI: 10.1109/ICASSP.2006.1660786.
- [16] Pulakka, H., and Alku, P. 2011. Bandwidth extension of telephone speech using a neural network and a filter bank implementation for highband mel spectrum, IEEE Trans. on Audio, Speech, and Language Processing, Cilt. 19(7), pp. 2170 - 2183. DOI: 10.1109/TASL.2011.2118206.
- [17] Kontio, J., Laaksonen, L., Alku, P. 2007. Neural network-based artificial bandwidth expansion of speech, IEEE Trans. on Audio, Speech, and Language Processing, Cilt. 15(3), pp. 873 - 881.
- [18] Wang, Y., Zhao, S., Liu, W., Li, M., Kuang, J., 2015., Speech bandwidth expansion based on deep neural networks, Proc. of INTERSPEECH, Dresden.
- [19] Prasad, N., Kumar, T. K. 2016. Bandwidth extension of speech signals: A comprehensive review, IJ. Intelligent Systems and Applications, Cilt. 2016(2), pp. 45-52, DOI: 10.5815/ijisa.2016.02.06.
- [20] Lahouti, F., Fazel, A. R., Safavi-Naeini, A. H., Khandani, A. K. 2006. Single and double frame coding of speech LPC parameters using a lattice-based quantization scheme, IEEE Trans. on Audio, Speech and Language Processing, Cilt. 14(5), s. 1624-1701. DOI: 10.1109/TSA.2005.858560.
- [21] Rabiner, L. R., Schafer, R. W. 1978. Digital Processing of Speech Signals, Prentice-Hall, New Jersey.
- [22] Kay, S. M. 1988. Modern Spectral Estimation: Theory and Application, Prentice-Hall, New Jersey.
- [23] Moon, T. K., Stirling, W. C. 2000. Mathematical Methods and Algorithms for Signal Processing, Prentice-Hall, New Jersey.
- [24] Viswanathan, R., Makhoul, J. 1975. Quantization properties of transmission parameters in linear predictive systems, IEEE Trans. on Acoustics, Speech and Signal Processing, Cilt. 23(3), s. 309-321. DOI: 10.1109/TASSP.1975.1162675.
- [25] Gray, Jr. A. H., Markel, J. D. 1976. Quantization and bit allocation in speech processing, IEEE Trans. on Acoustics, Speech and Signal Processing, Cilt. 24(6), s. 459-473. DOI: 10.1109/TASSP.1976.1162857.
- [26] F. Itakura, 1975 Line spectrum representation of linear predictive coefficients of speech signals, Journal of the Acoustical Society of America, Cilt. 57(1), p. 35. DOI: 10.1121/1.1995189
- [27] Gray, R., Buzo, A., Gray, A., Matsuyama, Y. 1980. Distortion measures for speech processing, IEEE Trans. on Acoustics, Speech, and Signal Processing, Cilt. 28(4), s. 367-376. DOI: 10.1109/TASSP.1980.1163421.
- [28] P. Jax. 2004 Bandwidth Extension for Speech, s. 171-235. Larsen, E., Aarts, R.M. 2004. Audio Bandwidth Extension: Application of Psychoacoustics, Signal Processing and Loudspeaker Design, John Wiley & Sons.