NOMA YÖNTEMLERİYLE TEK-KANALDA KONUŞMA-MÜZİK AYRIŞTIRMANIN KONUŞMA TANIMA PERFORMANSINA ETKİSİNİN ANALİZİ
ANALYSIS OF EFFECT OF SINGLE-CHANNEL SPEECH-MUSIC SEPARATION USING NMF TO AUTOMATIC SPEECH RECOGNITION

Cemil Demir[1,3], A. Taylan Cemgil[2], Murat Saraçlar[3]

[1]TÜBİTAK-BİLGEM, Kocaeli, Türkiye
[2]Bilgisayar Mühendisliği, Boğaziçi Üniversitesi, İstanbul,Türkiye
[3]Elektrik-Elektronik Mühendisliği, Boğaziçi Üniversitesi, İstanbul,Türkiye
cemil.demir@tubitak.gov.tr, (taylan.cemgil|murat.saraclar)@boun.edu.tr

ÖZETÇE

Bu çalışmada özellikle televizyonda konuşma tanıma uygulamalarında tanıma başarımını önemli oranda düşüren arka plan müziğinin konuşmadan ayrıştırılması için çalışmalar yapılmıştır. Ayrıştırma tek-kanalda yapılacak olduğundan, konuşma ve müzik sinyallerinin eğitim verileri kullanılarak modellenmesi gerekmektedir. Konuşma ve müzik sinyalleri Negatif Olmayan Matris Ayrıştırma (NOMA) yöntemiyle modellenmiştir. Bu çalışmada bir önceki çalışmamızda Kullback-Leibler (KL) ıraksayı kullanılarak yapılan analizler Itakura-Saito (IS) ıraksayı kullanılarak da yapılmıştır. Iraksayların konuşma-müzik ayrıştırma performansına etkisi karşılaştırılmıştır. Aynı zamanda bir önceki çalışmada denenmeyen; konuşma için herhangi bir eğitim kümesi olmadığı durum test edilmiştir. Bunun yanında müzik sinyali için müziğe ait çerçevelerin müziğe ait şablon vektörleri olarak kullanılması önerilmiş ve en yüksek başarım bu şekilde elde edilmiştir.

ABSTRACT

In this study, single-channel speech source separation is carried out to separate the speech from the background music, which degrades the speech recognition performance especially in broadcast news transcription systems. Since the separation is done using single observation of the source signals, the sources have to be previously modeled using training data. Non-negative Matrix Factorization (NMF) methods are used to model the sources. In order to model the source signals, different training data sets, which contain different music and speech data, are created and the effect of the training data sets are analyzed in this study. The performances of the methods are measured not only using separation performance measure but also with speech recognition performance measures.

1. GİRİŞ

Son zamanlarda haber bültenlerini yazılandırmak için geliştirilen Konuşma Tanıma (KT) uygulamaları popüler hale gelmiştir. Televizyon ve radyodaki haber bültenlerini yazılandırmak için geliştirilen bu uygulamalardaki başlıca problemlerden bir tanesi konuşmanın arkaplanında müzik olduğunda geliştirilen KT sistemlerinin performansının ciddi oranda düşmesidir. Bundan dolayı arkaplan müziğini temizlemek, gürbüz KT sistemleri geliştirmek için çok önemlidir. Gerçek hayatta kullanılacak bir KT sistemi, gelecek olan ses sinyalinde önce konuşma-müzik bölütlemesi yapabilecek; daha sonra bu bölütleme sonucunda konuşma-müzik karışımı olarak etiketlenen kısımlarda konuşma-müzik ayrıştırma yapabilecek yeteneğe sahip bir ön modüle sahip olmalıdır. Daha önce yapılan çalışmada [1] KT sistemleri için geliştirilen konuşma-müzik bölütleme yöntemi anlatılmıştır. Tek-kanalda birden fazla konuşmacıya ait konuşmaların birbirinden ayrıştırılması üzerine yapılan bir çok çalışma [2] olmasına rağmen tek kanalda konuşma-müzik ayrıştırma üzerine pek çalışılmamıştır [3, 4]. Tek-kanalda kaynak ayrıştırmada genel olarak Model-temelli ayrıştırma yöntemleri kullanılmakla beraber şimdiye kadar model-temelli yaklaşımlar, aynı sınıftan kaynakların, örneğin farklı konuşmacılara ait konuşmaların [5] ve müzikteki farklı enstrümanların [6], birbirinden ayrılması için kullanılmıştır.

Bu çalışmada daha önceki benzer çalışmamızdan [7, 8] farklı olarak sadece Kullback-Leibler (KL) ıraksayı temelli NOMA kullanmakla yerine Itakura-Saito (IS) ıraksayı temelli NOMA kullanarak da konuşma-müzik ayrıştırma deneyleri yapılmıştır ve iki ıraksayın ayrıştırma performansına etkileri karşılaştırımıştır. Aynı zamanda konuşma sinyali için herhangi bir eğitim kümesi kullanılmadığında konuşma şablon vektörlerinin uyarım matrisleri ile birlikte nasıl kestirileceği ve ayrıştırmanın nasıl yapılacağı anlatıldı. Bu durumda ortaya çıkan konuşma tanıma başarımları incelendi. Test kümesi daha öncekinden farklı olarak temiz konuşmaların 10 farklı cıngıl ile karıştırılması ile elde edildi.

2. YÖNTEM

Tek-kanalda konuşma-müzik ayrıştırma yapmak için konuşma ve müzik kaynaklarının eğitim verileri kullanılarak modellen-

mesi gerekmektedir. Bu modelleme sırasında kullanılacak özniteliklerin ve modelleme yönteminin seçimi önemli olmaktadır. Birden fazla kaynağın toplamı olan karışım sinyalinin öznitelikleri kaynaklara ait negatif olmayan özniteliklerin toplamına eşit olduğu durumlarda NOMA yöntemlerinin kullanılması uygun olmaktadır. Güç Spektrogramı (GS) bu tür özniteliklerdendir. NOMA yöntemi Lee ve Seung [9] tarafından veri incelemede kullanılması amacıyla k-means ve PCA yöntemlerine alternatif olarak önerilmiştir. NOMA yönteminde verilen negatif olmayan veri matrisi, X, için negatif olmayan bileşen matrisleri bulunmaya çalışılmaktadır. Bu bileşen bulma işlemini matematiksel olarak aşağıdaki gibi gösterebiliriz.

$$\mathbf{X} \approx \mathbf{UV} \tag{1}$$

Bu gösterimde U şablon vektörlerini V ise bu şablon vektörlerine ait uyarım değerlerini temsil etmektedir. GS veri matrisi olarak kullanıldığında şablon vektörleri konuşma yada müziğin karakteristik özelliklerini barındıran vektörleri, uyarım matrisi de her bir zaman için bu karakteristik vektörlerine ait uyarımları içermektedir. Konuşma sinyali için yapılan çalışmalarda şablon vektörlerinin konuşmayı oluşturan fonları temsil ettiği gösterilmiştir.

2.1. IS-NOMA

IS-NOMA yönteminde veriye ait olan GS, X, ile şablon ve uyarım matrislerinin çarpımı arasındaki IS uzaklık ölçütü

$$D_{IS}(X\|U,V) = \sum_{f,t}\left\{\frac{S_{ft}}{[U*V]_{ft}} - \log(S_{ft}) + \log([U*V]_{ft}) - 1\right\}$$

en azaltılmaya çalışılmaktadır. Bu gösterimde f ve t sırasıyla frekans ve zaman indekslerini göstermektedirler. Bu uzaklık ölçütünün en azaltılmasını sağlayan çarpımsal güncelleme denklemleri [10] aşağıdaki gibidir:

$$\mathbf{D} = \mathbf{D}. * \frac{(\frac{\mathbf{S}}{(\mathbf{D}*\mathbf{E})^2}) * \mathbf{E}^T}{\frac{\mathbf{1}}{\mathbf{D}*\mathbf{E}} * \mathbf{E}^T} \tag{2}$$

$$\mathbf{E} = \mathbf{E}. * \frac{\mathbf{D}^T * (\frac{\mathbf{S}}{(\mathbf{D}*\mathbf{E})^2})}{\mathbf{D}^T * \frac{\mathbf{1}}{\mathbf{D}*\mathbf{E}}}. \tag{3}$$

Bu gösterimde $\mathbf{1}$, birlerden oluşan uygun boyutlu matrisi göstermektedir.

2.2. NOMA ile Konuşma-Müzik Ayrıştırma

NOMA ile konuşma-müzik ayrıştırmada, eğitim sırasında konuşma ve müzik sinyallerine ait olan GS matrisleri kullanılarak her bir sinyale ait şablon matrisleri öğrenilmektedir. Bu eğitimi

$$S = U_s V_s \quad \text{and} \quad M = U_m V_m. \tag{4}$$

şeklinde gösterebiliriz. Bu gösterimde U_s veU_m sırasıyla konuşma ve müzik sinyalleri için öğrenilen şablon matrislerini temsil etmektedir. Şablon ve uyarım matrisleri çarpımsal güncelleme denklemleri kullanılarak hesaplanmaktadır. Ayrıştırma sırasında, konuşma ve müzik sinyalleri için eğitilmiş olan şablon matrisleri kullanılarak genel şablon matrisi oluşturulur. Genel şablon matrisi sabitlenerek karışım sinyalinin GS matrisine karşılık gelen genel uyarım matrisi çarpımsal güncelleme denklemleri yardımıyla hesaplanır. Bu ayrıştırmayı

$$X = [U_s^* U_m^*][(V_s^*)^T (V_m^*)^T] \tag{5}$$

şeklinde gösterebiliriz. Konuşma ve müzik sinyaline karşılık gelen uyarım matrisleri ve eğitilmiş olan şablon matrisi yardımıyla karışım içindeki konuşma ve müzik sinyalleri geri çatılır. Geri çatma işlemi elde edilen şablon ve uyarım matrisleri kullanılarak her bir kaynağın sonsal olasılıklarını en büyütecek şekilde yapılmaktadır. Bu sonsal olabilirliği en büyütecek kaynak geri çatımları

$$\widehat{S} = X. * \frac{U_s^* V_s^*}{(U_s^* V_s^* + U_m^* V_m^*)}. \tag{6}$$

$$\widehat{M} = X. * \frac{U_m^* V_m^*}{(U_s^* V_s^* + U_m^* V_m^*)}. \tag{7}$$

şeklinde hesaplanmaktadır.

3. DENEYSEL SONUÇLAR

3.1. Başarım Ölçütleri:

Yaptığımız çalışmada konuşma-müzik ayrıştırma ile amaçlanan KT başarımını arttırmak olduğu için ayrıştırma yöntemlerinin performansları KT başarım ölçütü olan Kelime Doğruluk Oranıyla (KDO) ile ölçülmüştür. Aynı zamanda KT başarımı ile ayrıştırma başarımı arasındaki ilişkiyi incelemek amacıyla yöntemlerin ayrıştırma başarımları da ölçülmüştür. Ayrıştırma başarımlarını ölçmek amacıyla ayrıştırılan konuşma içindeki kalan müzik miktarını ölçmek amacıyla Konuşma-Müzik Oranı (KMO) ve konuşmada meydana gelen bozulmayı ölçmek amacıyla Konuşma-Bozulma Oranı (KBO) kullanılmıştır.

3.2. Deney Düzeneği:

Bu çalışmada konuşma-müzik ayrıştırmada kullanılan eğitim verilerinin ayrıştırma başarımına etkisini ölçme amacına uygun olarak deney düzenekleri hazırlanmıştır. Deney kümesi; 8 konuşmacıya ait yaklaşık 2 saat uzunluğundaki konuşmaların ortalama 7 saniye uzunluğundaki 10 farklı cıngıl ile 0, 5, 10, 15 ve 20 dB seviyelerinde yapay olarak karıştırılmalarıyla oluşturulmuştur. Kullanılan cıngıllar televizyon haberlerinde kullanılan cıngıllardan seçilmiştir. NOMA için kullanılan BS ve GS matrisleri 1024 boyutlu pencereleri 512 birim kaydırarak elde edilen çerçevelerin Fourier dönüşümleri alınarak hesaplanmıştır. Eğitim verisi olarak her bir konuşmacı için; kendisine ait başka konuşmalarından oluşan "Kendisi", kendisi dışındaki aynı cinsten olan insanların konuşmalarından oluşan "Diğerleri" ve kendisi ile birlikte kendi cinsinden olan diğer konuşmacılara ait konuşmaların bulunduğu "Tümü" adlı konuşma veritabanları oluşturulmuş ve bu veriler kullanılarak her konuşmacı için NOMA modelleri oluşturulmuştur. Aynı zamanda konuşma sinyali için herhangi bir eğitilmiş model kullanılmadığı durum 'Hiçbiri' olarak adlandırılmıştır. Müzik modellerini eğitmek için de benzer bir yaklaşım kullanılmıştır. Ancak müzik modellerinde "Orjinal" adında veritabanındaki müziğin çerçevelerinin şablon vektörleri olarak kullanıldığı durum da test edilmiştir. Konuşma ve müzik için kullanılan 4 farklı

modelin çaprazlanması sonucu konuşma-müzik ayrıştırmada kullanılacak 16 farklı model çeşidi ortaya çıkmıştır. Bu modellere ait sonuçlar incelenerek konuşma müzik ayrıştırmada konuşma ve müziğe ait eğitim verilerinin ayrıştırma performansına olan etkileri tespit edilmeye çalışılmıştır. Aşağıdaki Tablo 1'de konuşma ve müzik NOMA modellerini eğitmek için kullanılan verilerin özellikleri gösterilmiştir.

Tablo 1: Eğitim Verisi Özellikleri

Özellikler	Konuşma				Müzik			
	Kendisi	Diğerleri	Tümü	Hiçbiri	Orjinal	Kendisi	Tümü	Herkes
Süre(Sn)	120	360	480	0	7	7	63	70
Şablon vektör sayısı	30	30	30	30	224	30	30	30

3.3. Konuşma Tanıma Sistemi

Geliştirilen KT sisteminde kullanılan cinsiyet-bağımsız akustik model yaklaşık 125 saatlik konuşma verileri kullanılarak eğitilmiştir. Akustik model eğitim birimi olarak bağlam-bağımlı üçlüsesler kullanılmıştır. Öznitelik olarak 25 ms uzunluğundaki pencerelerin 10 ms kaydırılması sonucu elde edilen çerçevelerin 13 boyutlu MFKK'ları kullanılmıştır. Bu MFKK vektörlerine fark ve fark-fark vektörleri de eklenerek nihai 39 boyutlu öznitelik vektörleri oluşturulmuştur. KT sisteminde kullanılan dil modeli 200 milyon kelime içeren gazete haber metinlerinden 50 bin kelimelik bir sözlük için üç gram olasılıklarının hesaplanması yoluyla elde edilmiştir.

3.4. Eğitim Verilerinin Performans Analizi:

NOMA modellerini eğitmek için kullanılan eğim verilerinin ayrıştırma performansına etkisini incelemek için oluşturulan 16 modelin kullanılmasıyla elde edilen KMO değerleri Tablo 2 ve 5 de gösterilmiştir. KMO değerleri incelendiğinde 'Orjinal' müzik modelinin diğer modellere göre daha yüksek değerler ürettiği görülmüştür. Konuşma için 'Hiçbiri' modeli dışındaki modellerde, müzik için kullanılan 'Kendisi' ve 'Tümü' modellerinin benzer KMO değerleri ürettiği görülmüştür. Konuşma için 'Hiçbiri' modeli kullanıldığında müzik için 'Tümü' ve 'Diğerleri' modelleri benzer KMO değerleri üretmektedir. Genel olarak KL ıraksayının IS ıraksayına göre daha yüksek KMO değerleri üretmektedir.

KBO değerleri Tablo 3 ve 6 de gösterilmiştir. Bu tablolar incelendiğinde tüm konuşma modelleri için 'Orjinal' ve 'Kendisi' müzik modellerinin 'Tümü' ve 'Diğerleri' modellerine göre daha yüksek KBO değerleri ürettiği tespit edilmiştir. Konuşma için 'Hiçbiri' modeli kullanıldığında 'Tümü' ve 'Diğerleri' müzik modellerinin benzer KBO değerleri ürettiği görülmüştür.

KDO değerleri Tablo 4 ve 7 de gösterilmiştir. Bu tablolar incelendiğinde KL ıraksayı kullanıldığında 'Hiçbiri' konuşma modeli ile 'Tümü' ve 'Diğerleri' müzik modellerinin herhangi bir ayrıştırma yapılmadığı duruma göre daha düşük KDO değerleri ürettiği görülmüştür. Bunun dışındaki tüm durumlarda ayrıştırma yapmanın konuşma tanıma başarımını arttırdığı görülmüştür. Tüm konuşma modelleri için 'Orjinal' müzik modelinin daha yüksek KDO değerleri ürettiği görülmüştür. Aynı zamanda 'Hiçbiri' hariç diğer konuşma modelleri için 'Orjinal' müzik modelinin benzer konuşma tanıma başarımları ortaya çıkardığı görülmüştür.

Tablo 2: KL-NOMA ile elde edilen KMO değerleri (dB)

Çıktı KMO (dB)		Girdi KMO (dB)				
Konuşma	Müzik	0dB	5dB	10dB	15dB	20dB
Hiçbiri	Diğerleri	2.1	13.6	23.9	35.5	45.4
	Tümü	2.9	14.7	26.4	36.9	46.5
	Kendisi	9.9	19.6	32.4	38.0	47.0
	Orjinal	17.9	25.4	38.7	41.4	49.9
Diğerleri	Diğerleri	8.3	17.7	26.2	35.9	44.8
	Tümü	9.8	19.0	27.9	36.7	45.5
	Kendisi	9.9	18.9	30.3	36.5	45.3
	Orjinal	14.6	22.6	34.1	38.6	46.9
Tümü	Diğerleri	8.4	17.9	26.4	36.1	45.0
	Tümü	9.8	19.1	28.1	36.9	45.7
	Kendisi	10.0	19.1	30.5	36.8	45.5
	Orjinal	14.9	22.9	34.5	39.0	47.3
Kendisi	Diğerleri	9.6	18.8	27.2	36.6	45.4
	Tümü	11.2	20.2	28.9	37.5	46.1
	Kendisi	11.0	19.9	31.2	37.2	45.8
	Orjinal	15.3	23.2	34.5	39.0	47.2

Tablo 3: KL-NOMA ile elde edilen KBO değerleri (dB)

Çıktı KBO (dB)		Girdi KMO (dB)				
Konuşma	Müzik	0dB	5dB	10dB	15dB	20dB
Hiçbiri	Diğerleri	8.2	10.0	12.2	14.9	16.7
	Tümü	8.1	9.8	12.2	14.2	15.7
	Kendisi	10.0	12.0	15.1	16.5	18.4
	Orjinal	9.2	11.2	14.5	15.9	17.8
Diğerleri	Diğerleri	10.3	12.6	14.5	16.7	18.3
	Tümü	10.3	12.7	14.5	16.2	17.5
	Kendisi	10.7	13.1	16.0	17.7	19.8
	Orjinal	10.8	13.1	16.2	18.0	20.2
Tümü	Diğerleri	10.2	12.7	14.7	16.9	18.6
	Tümü	10.3	12.9	14.9	16.7	18.2
	Kendisi	10.7	13.2	16.3	18.0	20.2
	Orjinal	10.9	13.3	16.4	18.2	20.6
Kendisi	Diğerleri	9.9	12.2	14.0	16.0	17.5
	Tümü	10.0	12.2	14.0	15.7	17.0
	Kendisi	10.5	12.8	15.6	17.3	19.3
	Orjinal	10.6	12.9	15.9	17.6	19.8

Genel olarak ayrıştırma performansları incelendiğinde IS ıraksayının KL ıraksayına göre daha düşük KMO üretmesine rağmen daha yüksek KBO değerleri ürettiği için konuşma tanıma başarımını daha çok arttırdığı tespit edilmiştir. Konuşma tanıma başarımları incelendiğinde yüksek girdi KMO değerlerinde kullanılan model kombinasyonlarının arasındaki performans farkının azaldığı görülmüştür. Müzik modeli için 'Orjinal' modelinin diğer tüm modellere göre daha iyi sonuç verdiği ve konuşma tanıma açısından kullanılmasının faydalı olduğu tespit edilmiştir.

4. SONUÇ

Bu çalışmada KT performansını arttırmak için kullanılan NOMA yaklaşımlarının performansları değerlendirilmiştir. KL ve IS ıraksaylarının ayrıştırma performansları karşılaştırılmıştır. IS ıraksayının KL ıraksayına göre genel olarak daha iyi ayrıştırma yaptığı tespit edilmiştir. Aynı zamanda farklı eğitim kümeleriyle başarım nasıl değiştiği üzerine analizler yapılmıştır. Müzik modeli olarak 'Orjinal' modelinin diğer bir

Tablo 4: KL-NOMA ile elde edilen KDO değerleri (%)

KDO (%)		Girdi KMO (dB)				
Konuşma	Müzik	0dB	5dB	10dB	15dB	20dB
Hiçbiri	Diğerleri	1.2	7.2	21.3	42.2	54.5
	Tümü	1.4	8.0	25.5	44.1	55.3
	Kendisi	10.7	24.6	49.2	54.5	64.5
	Orjinal	17.2	28.0	51.1	53.3	61.3
Diğerleri	Diğerleri	9.9	25.3	45.1	62.7	70.8
	Tümü	11.5	28.5	50.3	64.3	71.1
	Kendisi	14.3	31.6	58.8	64.4	71.9
	Orjinal	27.5	43.0	67.0	66.5	71.4
Tümü	Diğerleri	9.0	26.8	45.4	63.6	70.4
	Tümü	11.3	29.0	50.4	65.3	71.4
	Kendisi	14.2	31.5	59.9	64.0	71.6
	Orjinal	28.1	43.6	67.8	67.4	72.0
Kendisi	Diğerleri	9.4	25.1	0.0	60.3	68.0
	Tümü	11.1	28.2	48.7	61.5	68.9
	Kendisi	14.5	31.9	57.7	62.2	69.6
	Orjinal	27.5	41.2	63.3	63.9	69.6

Tablo 5: IS-NOMA ile elde edilen KMO değerleri (dB)

Çıktı KMO (dB)		Girdi KMO (dB)				
Konuşma	Müzik	0dB	5dB	10dB	15dB	20dB
Hiçbiri	Diğerleri	1.9	13.0	22.6	33.6	43.4
	Tümü	3.1	14.2	24.7	34.8	44.4
	Kendisi	8.7	18.0	30.2	36.1	45.1
	Orjinal	13.4	21.6	34.4	38.5	47.1
Diğerleri	Diğerleri	7.8	17.0	25.5	35.1	44.1
	Tümü	9.0	18.1	26.9	35.8	44.7
	Kendisi	9.0	17.9	29.0	35.5	44.4
	Orjinal	12.2	20.3	31.6	36.9	45.5
Tümü	Diğerleri	7.7	17.0	25.6	35.3	44.4
	Tümü	9.0	18.2	27.2	36.1	45.0
	Kendisi	9.1	18.1	29.3	35.8	44.7
	Orjinal	12.6	20.7	32.2	37.3	45.9
Kendisi	Diğerleri	8.5	17.5	25.9	35.4	44.4
	Tümü	9.9	18.7	27.5	36.2	45.0
	Kendisi	9.7	18.4	29.5	35.9	44.7
	Orjinal	12.7	20.7	32.0	37.2	45.7

Tablo 6: IS-NOMA ile elde edilen KBO değerleri (dB)

Çıktı KBO (dB)		Girdi KMO (dB)				
Konuşma	Müzik	0dB	5dB	10dB	15dB	20dB
Hiçbiri	Diğerleri	6.8	9.6	12.7	16.6	19.8
	Tümü	6.8	10.0	13.7	17.4	20.7
	Kendisi	9.3	12.1	16.7	18.5	21.7
	Orjinal	9.7	12.4	17.4	18.7	21.7
Diğerleri	Diğerleri	8.5	11.7	14.4	17.6	20.4
	Tümü	8.8	12.0	14.8	17.6	20.1
	Kendisi	9.4	12.3	16.3	18.4	21.6
	Orjinal	10.3	13.0	17.3	19.0	22.0
Tümü	Diğerleri	8.3	11.5	14.3	17.7	20.7
	Tümü	8.6	11.8	14.7	17.7	20.4
	Kendisi	9.3	12.3	16.4	18.6	21.8
	Orjinal	10.2	13.1	17.5	19.1	22.2
Kendisi	Diğerleri	8.6	11.7	14.3	17.5	20.4
	Tümü	8.9	12.1	14.9	17.8	20.6
	Kendisi	9.5	12.4	16.5	18.6	21.7
	Orjinal	10.3	13.1	17.4	19.1	22.1

Tablo 7: KL-NOMA ile elde edilen KDO değerleri (%)

KDO (%)		Girdi KMO (dB)				
Konuşma	Müzik	0dB	5dB	10dB	15dB	20dB
Hiçbiri	Diğerleri	1.4	9.8	26.8	50.6	62.6
	Tümü	2.1	14.0	37.2	55.7	66.1
	Kendisi	14.7	30.7	56.5	59.4	68.1
	Orjinal	31.4	42.6	68.2	62.8	69.8
Diğerleri	Diğerleri	9.9	26.2	44.1	62.9	69.5
	Tümü	12.3	28.8	50.2	64.1	70.7
	Kendisi	17.4	34.1	61.4	64.6	72.0
	Orjinal	39.6	49.2	0	67.2	72.2
Tümü	Diğerleri	9.4	25.2	43.5	62.0	70.0
	Tümü	11.8	28.8	50.9	64.5	70.3
	Kendisi	16.5	33.9	60.4	64.8	71.3
	Orjinal	39.3	49.1	0	67.1	72.2
Kendisi	Diğerleri	11.0	26.3	45.0	62.1	69.0
	Tümü	30.2	30.2	51.5	64.2	70.1
	Kendisi	18.4	35.2	61.8	64.2	71.6
	Orjinal	38.9	49.1	0	66.7	72.6

ifadeyle müziğe ait çerçeveleri şablon vektörleri olarak kullanmanın en iyi başarımı sağladığı görülmüştür.

5. KAYNAKÇA

[1] C. Demir and M. U. Dogan, "Speech-music segmentation for speech recognition," *Proc. of SIU*, 2009.

[2] M.N. Schmidt and R.K. Olsson, "Single-channel speech separation using sparse non-negative matrix factorization," in *Proc. of ICSLP*, 2006, pp. 2614–2617.

[3] S. Kirbiz and B. Gunsel, "Perceptual single-channel audio source separation by non-negative matrix factorization," in *in proc. of SIU*, 2009, pp. 416–419.

[4] S. Yildirim and M. Saraclar, "Single channel music and speech separation using non-negative matrix factorization," in *in proc. of SIU*, 2009, pp. 301–304.

[5] P. Smaragdis, M. Shashanka, M. Inc, and B. Raj, "A Sparse Non-Parametric Approach for Single Channel Separation of Known Sounds," Proc. of NIPS, 2009, pp. 1705–1713.

[6] T. Virtanen, "Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. on ASLP*, vol. 15, no. 3, pp. 1066–1074, 2007.

[7] Dogan M. U. Demir, C., A.T. Cemgil, and M. Saraçlar, "Single-channel speech-music separation using NMF for automatic speech recognition," *Proc. of SIU*, 2009.

[8] C. Demir, A.T. Cemgil, and M. Saraçlar, "Gain Estimation Approaches in Catalog-Based Single-Channel Speech-Music Separation," in *Proc. of ASRU*, 2011, pp. 185–190.

[9] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, 1999.

[10] C. Févotte, N. Bertin, and J.L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.