

# Model Tabanlı Ses Dizisi Hizalanması

## Model Based Audio Sequence Alignment

Doğaç Başaran, Emin Anarım

Ali Taylan Cemgil

Elektrik ve Elektronik Mühendisliği Bölümü  
Boğaziçi Üniversitesi  
dogac.basaran@gmail.com  
anarim@boun.edu.tr

Bilgisayar Mühendisliği Bölümü  
Boğaziçi Üniversitesi  
taylan.cemgil@boun.edu.tr

### ÖZETÇE

*Bu çalışmada, çoklu ses dizilerinin hizalanması sorunu olasılıksal bir model yaklaşımı ile ele alınmıştır. Önerilen modelde, bağımsız ve eşzamansız kaydedilmiş ses kayıtlarından özümlenmiş, zamanla değişen öznitelikler tanımlanmıştır. Bu çalışmada çoklu ses kayıtlarının tümünün alınmadığı kayıtlı, hiçbir hizalamanın gerekli bilgiyi içermediği durum ele alınmıştır. Eşleştirme işlemi, yaklaşık Bayes çıkarım ile gerçekleştirilmiştir. Burada hizalanma kaymasının kesin sonsal dağılımın örneklenmesi için bir benzetimli tavlama yaklaşımının kullanımı sergilenmiştir. Gerçek ve yapay verilerle yapılan benzetim sonuçlarıyla, zor ve muğlak senaryolar veya kısmi eşleştirmelerde önerilen yöntemin kullanılabilirliği gösterilmiştir.*

### ABSTRACT

We formulate alignment of multiple audio sequences in a probabilistic framework. Our approach defines a generative model for time varying features extracted from audio clips that are recorded independently and asynchronously. We are able to handle missing data and multiple clips where no clip is covering the entire material. The matching is achieved via approximate Bayesian inference. Here, we illustrate a simulated tempering approach for sampling from the exact posterior density of the clip offsets. The simulation results on synthetic and real data suggest that the framework is able to handle difficult ambiguous scenarios or partial matchings.

## 1. GİRİŞ

Ses eşleştirme genel olarak, bilinmeyen bir ses bölümünün büyük bir veri tabanında eşleştirildiği bir tanımlama problemi olarak ele alınır. Literatürde, çok gürültülü koşullarda, yüksek eşleştirme başarımı sağlayan dayanıklı ses parmakizi yöntemleri mevcuttur [1, 2]. Bu bildiride, çoklu eşleştirme problemi ele alınmış ve ses eşleştirme sorununa değişik bir açıdan bakılmıştır.

Kayıt ortamında birçok eşzamansız mikrofon olduğu ve bunların kayıtlarının birbirinden bağımsız ve değişik zamanlarda alındıkları varsayılmıştır. Amaç, örtüşen ya da örtüşmeyen ses kayıtlarının başlangıç noktalarına göre genel bir zaman çizgisinde hizalamaktır. Bu sorunun genel geçer ses eşleştirme yaklaşımlarından en önemli farkı, temiz orijinal ses

kayıtlarının olduğu bir veri tabanının değil sadece muhtemelen kaynağın gürültülü kayıtlarının bulunmasıdır. Ayrıca herhangi bir gözlemin bütün zaman çizgisini kaplamak zorunda olmadığını da düşünmek gerekir. Çoklu ses eşleştirme sorununu ele almada itici güç tam hizalanmış kayıtların kaynak ayırıştırma, iyileştirme veya yeniden oluşturmak için kullanılmaktadır.

Böyle bir senaryo bir gösteri esnasında bir konser salonunda oluşabilir. Bazı dinleyicilerin değişik kalitede kayıt düzenekleri ile konserin kendi beğendikleri kısımlarını kaydettikleri durumda, her bir kayıt farklı bir açıdan elde edilmiş olur. Bu ses kayıtlarını sözünü ettiğimiz amaçlar doğrultusunda kullanabilmek için, her bir kaydın net bir şekilde genel zaman çizgisinde hizalanması gerekir. Benzer yaklaşımlar değişik açılardan çekilmiş imgelerin bir araya getirilerek panoramik görüntü oluşturması probleminde de mevcuttur.

İlkesel olarak soruna ilinti ve şablon eşleştirme gibi belirlenimci yöntemlerle yaklaşmak mümkündür. Fakat bunlarda da belirli sınırlar mevcuttur. Öncelikle ses sinyalleriyle ilgilenirken örnek sayısı genelde çok olur ve çok sayıda kayıt olduğunda işlem maliyeti çok fazladır. Birçok ses eşleştirme uygulaması, çoklu ses kayıtlarında bile ikişerli şekilde çalışır.  $K$  tane ses kaydı olduğunda  $O(K^2)$  mertebesinde ikili eşleştirme yapmak gerekmektedir. Buna ek olarak eğer ses kayıtları örtüşmezse veya bir kayıta bazı veri noktaları kaçırılırsa, ilinti ve şablon eşleştirme yöntemlerinin nasıl uygulanacağı çok açık değildir. İşlem karmaşıklığını ve veri sayısını azaltmanın açık bir yolu direk ses sinyalinin üstünde çalışmak yerine öznitelik uzayında çalışmaktır. Kısa süreli izge büyüklüğü, yerel renk doyumluğu enerji dağılımı, eksi olmayan dizey çarpanlarına ayırma ve kısa süreli pencerelerdeki sinyal enerjisi, ses eşleştirme çatısı altında en çok kullanılan özniteliklerdendir [1, 2]. Fakat özniteliklerle çalışırken bile bir referans zaman çizgisi olmadan, birçok kısa kayıtlarla çalışmak zordur.

Burada, model bazlı bir yaklaşım ve ses öznitelikleri için üretimsel bir model tanımlanmaktadır. Modelleme yaklaşımı herhangi bir öznitelik yöneyini kullanmaya müsaade etse de, biz pozitif öznitelikleri bu çalışmada kullanılmaktayız. Bu modelin bir üstünlüğü de ses kayıtlarında ikili eşleştirme yapmak yerine her kaydı gizli bir ses içeriğine hizalıyor olması ve bunun işlemsel karmaşıklığı azaltmasıdır. Uygun bir puan işlevi tanımlanarak eksikli kayıtlar için bile sinyalleri eşleştirmek mümkündür. Sadece 2 tane kayıt olduğu durumda ise olası

her hizalama için puan işlevi, yarı örtüşme ve örtüşmeme durumlarını tanımlayarak yapmak mümkündür. Ancak, çoklu dizilerde net skorlama mümkün değildir. Bu çalışmada, ortak zaman çizgisi üzerinde çoklu ses sinyallerini eşleştirmek için benzetimli tavlama çatısı altında Gibbs örnekleme yaklaşımı önerilmektedir.

Bölüm 2’de önerilen model ayrıntılı bir şekilde tanımlanmakta ve eşleştirme için uygun bir puan işlevi verilmektedir. Gibbs örnekleme için benzetimli tavlama yaklaşımı Bölüm 3’te ve deney düzeneği, benzetim sonuçları ve tartışmalar Bölüm 4’te verilmektedir.

## 2. ÖNERİLEN MODEL

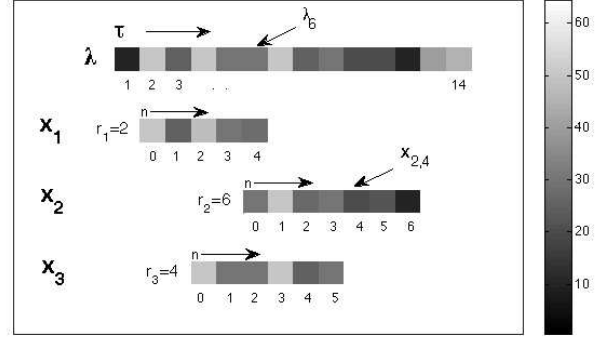
Bu bölümde, eşleştirme problemi için oluşturulan olasılık modeli, basit bir örnekle Şekil 1’de verilmektedir. Model içinde doğrudan ses sinyali yerine, öznelik dizilerini tanımlamıştır. Burada öznelik olarak, altbantlardaki zamanla değişen enerji bilgisi ele alınmaktadır. Enerji bilgisi, 20 ms’lik örtüşmeyen pencerelerde kısa zamanlı Fourier dönüşümü olarak ve dönüşüm katsayılarının karelerini sıklık indeksi üzerinden toplayarak elde edilmiştir. Modelin gösterimini kolaylaştırmak amacıyla, örnekte spektrogramın yalnızca bir altbantı kullanılmaktadır. Ancak çeşitli altbantların eşzamanlı kullanıldığı çok değişkenli durumlar için yaklaşımın genelleştirilmesi mümkündür. Modelin ana fikri, gözlemlenmiş ve doğru şekilde hizalanmış öznelik dizilerinin, aslında alta yatan ve gözlemlenemeyen gerçek ses kaynağının öznelik dizilerinin gürültülü uyarlamaları olarak ele alınmasıdır.

Gizli öznelik yöneyi  $\lambda_\tau$  ile tanımlanmakta ve  $\tau = 1 \dots T$  her yerde geçerli bir zaman çerçevesi indeksi olarak tanımlanmaktadır. Bir tane altbant için öznelik yöneyi bir sayıdır. Şekil 1’de verilen örnekte, üç tane gözlemlenmiş kayıt ve  $x_k$   $k$ ’nci kaydın öznelik vektörünü,  $N_k$  da bu öznelik vektörünün uzunluğunu ifade etmektedir. Ayrıca  $T = 14$ ,  $N_1 = 5$ ,  $N_2 = 7$ , ve  $N_3 = 6$  alınmıştır. Burada  $n$  yerel bir zaman çerçevesi indeksidir ve  $k$ ’nci kaydın hizalama değişkeni  $r_k$  olarak ifade edilmektedir. Örneğin ikinci kayıt her yerde geçerli zamanda  $\tau = 6$  noktasında başladığından,  $r_2 = 6$  olmaktadır. Bu senaryoda, kayıtlar çeşitli noktalarda örtüşmekte ve tam olarak  $x_{1,4}, x_{2,0}$  ve  $x_{3,2}$  her yerde geçerli zamanla  $\tau = 6$ ’da çakışmaktadır. Şekilden de görülebileceği gibi bütün bu katsayı değerleri birbirine çok yakın değerlerdedir. Böyle olmalarının sebebi, bunların ortak gizli kaynak  $\lambda_6$ ’nın gürültülü uyarlamaları olmalarıdır.

Üretimsel model aşağıdaki gibi verilmiştir:

$$\begin{aligned} \lambda_\tau &\sim \mathcal{IG}(\lambda_\tau; \alpha_\lambda, \beta_\lambda) \\ r_k &\sim \prod_{\tau=1}^{T-N_k+1} \pi_{k,\tau}^{[r_k=\tau]} \\ x_{k,n}|r_k, \lambda_\tau &\sim \prod_{\tau=1}^T \mathcal{G}(x_{k,n}; \alpha, \alpha/(\lambda_\tau))^{[n=\tau-r_k]} \end{aligned}$$

Burada  $\mathcal{IG}$  ve  $\mathcal{G}$  sırasıyla ters gamma ve gamma dağılımlarını ifade etmektedir. Bu dağılımlar pozitif rastgele değişkenleri modellemedeki esneklikleri ve eşleniklik özelliklik-



Şekil 1: Model açıklaması.  $\lambda$  gizli,  $x_1, x_2$  ve  $x_3$  gözlemlenmiş

leri nedeni ile seçilmişlerdir.  $k$ ’nci kaydın hizalama değişkeni  $r_k$ ’in  $\tau$  zamanında olma olasılığı  $\pi_{k,\tau}$  ile ifade edilmektedir. Ayrıca  $r_k$  hizalama değişkeninin birbiçimli dağılımlı olduğu varsayılmıştır. Spektrogram büyüklüğü, öznelik yöneyi olarak kullanıldığından, hem  $\lambda_\tau$  hem de  $x_{k,n}$  pozitif dağılımlardır.

Modelin daha iyi anlaşılması için hiper-parametre  $\alpha$ ’yı incelemek gerekmektedir. Üretimsel modele göre  $x_{k,n}$  değişkeninin ortalaması ve değişintisi sırasıyla  $\lambda_\tau$  ve  $\frac{\lambda_\tau^2}{\alpha}$  olarak tanımlanmıştır.

Bu nedenle  $x_{k,n}$ ’in  $\lambda_\tau$ ’dan ne kadar saptığını ayarlayan parametrenin  $\alpha$  olduğu açıktır. Eğer  $\alpha$  büyükse, değişinti küçük demektir, bu da  $x_{k,n}$ ’in  $\lambda_\tau$ ’ya yakın bir değeri olduğu anlamına gelmektedir. Fakat eğer  $\alpha$  küçükse, bu değişintinin yüksek olması yani  $x_{k,n}$ ’in  $\lambda_\tau$ ’dan yüksek miktarda sapabilmesi anlamına gelmektedir.

Burada asıl amaç, gizli öznelik yöneyini kestirmek değil,  $r_{1:K}$  ile ifade edilen en olası dizilimi bulmaktır. Bu birleşik koşullu dağılım  $p(r_{1:K}|x_{1:K,0:N_k-1})$ ’in kipine karşılık gelmektedir. Doğru dizilime dair hiçbir öncül bilginin olmadığını varsayarsak, sonsal dağılım yerine olabilirlik dağılımı  $p(x_{1:K,0:N_k-1}|r_{1:K})$  kullanılabileceği açıktır. Modele göre tam olabilirlik dağılımı aşağıda verilmektedir:

$$\begin{aligned} p(x_{1:K,0:N_k-1}|r_{1:K}) \\ = \int d\lambda_{1:T} \prod_{k=1}^K \prod_{n=0}^{N_k-1} p(x_{k,n}|r_k, \lambda_{1:T}) \prod_{\tau=1}^T p(\lambda_\tau) \end{aligned}$$

Gizli öznelik yöneyinin her bir elemanı  $\lambda_\tau$ , birbirlerinden bağımsızdır ve  $x_{k,n}$ ,  $\lambda_{1:T}$  verildiğinde ise birbirlerinden koşullu olarak bağımsız olacakları açıktır. Gamma ve ters gamma dağılımlarının eşleniklikleri sayesinde tümlevin analitik olarak elde edilmesi mümkündür. Sonuçta oluşan olabilirlik dağılımının logaritması,  $\mathcal{L}(r_{1:K})$ , aşağıdaki gibi ifade edilmektedir.

$$\begin{aligned}
\mathcal{L}(r_{1:K}) &= -T \log \Gamma(\alpha_\lambda) + T\alpha_\lambda \log \beta_\lambda \\
&+ \sum_{\tau=1}^T \sum_{k=1}^K \sum_{n=0}^{N_k-1} [n = \tau - r_k] \left( (\alpha - 1) \log x_{k,n} \right. \\
&\quad \left. - \log \Gamma(\alpha) + \alpha \log \alpha \right) \\
&+ \sum_{\tau=1}^T \log \Gamma(\alpha_\lambda + \sum_{k=1}^K \sum_{n=0}^{N_k-1} [n = \tau - r_k] \alpha) \\
&- \sum_{\tau=1}^T (\alpha_\lambda + \sum_{k=1}^K \sum_{n=0}^{N_k-1} [n = \tau - r_k] \alpha) \\
&\quad \log(\beta_\lambda + \sum_{k=1}^K \sum_{n=0}^{N_k-1} [n = \tau - r_k] \alpha x_{k,n})
\end{aligned}$$

Daha sonra  $\mathcal{L}(r_{1:K})$ 'in  $r_{1:K}$ 'ya göre enbüyüğünü bularak, en iyi dizilim noktalarına aşağıdaki gibi ulaşmak mümkündür,

$$\begin{aligned}
r_{1:K}^* &= \arg \max_{r_{1:K}} \mathcal{L}(r_{1:K}) \\
&= \arg \max_{r_{1:K}} \log p(x_{1:K,0:N_k-1} | r_{1:K})
\end{aligned}$$

Hizalama değişkenleri  $r_k$  ayrık oldukları için arama uzayı sonludur ve her olası hizalama  $r_{1:K}$  için puanlama yaparak en doğru hizalamayı bulmak anlamlıdır. Fakat büyük  $K$  değerleri için arama uzayı çok büyüür, dolayısıyla Gibbs örnekleycisi gibi yaklaşık çıkarım yöntemlerine başvurmak gerekir.

### 3. BENZETİMLİ TAVLAMA İLE GIBBS ÖRNEKLEYİCİSİ

Gibbs örnekleme, en az 2 boyutlu dağılımlardan örnekleme için kullanılan ünlü Markov Chain Monte Carlo (MCMC) yöntemlerinden biridir [4, 5]. Örnekleme yöntemleri çoğunlukla hedef dağılımın analitik hesaplamaları yapmak için çok karışık olduğu durumlarda kullanılır. Önerilen modelde, koşullu birleşik dağılımlardan örnekleme yapmak için Gibbs örnekleycisi kullanılmıştır. Gibbs örnekleycisi, tam koşullu dağılımları kullanarak, değişkenleri tek tek örnekler [3]. Tam koşullu dağılım, diğer tüm rastgele değişken ve verilere göre şekillenmiş değişkenin olasılığı anlamına gelir. Bu tam koşullu dağılımlar çoğunlukla tek boyutlu, izlenebilir ve modele bağlı örnekleme için uygundur.

Örneğin  $\theta$  değişkeninin tam koşullu dağılımı,  $p(\theta|.)$  olarak ifade edilir. Öne sürülen modelden, tam koşullu dağılımlar  $p(\lambda_\tau|.)$  ve  $p(r_k|.)$  aşağıdaki gibi türetilmiştir.

$$\begin{aligned}
p(\lambda_\tau|.) &= \mathcal{IG}(\lambda_\tau; \alpha_\lambda + \sum_{k=1}^K \sum_{n=0}^{N_k-1} [n = \tau - r_k] \alpha, \\
&\quad \beta_\lambda + \sum_{k=1}^K \sum_{n=0}^{N_k-1} [n = \tau - r_k] \alpha x_{k,n}) \\
p(r_k|.) &= \prod_{n=0}^{N_k-1} \prod_{\tau=1}^T \mathcal{G}(x_{k,n}; \alpha, \frac{\alpha}{\lambda_\tau})^{[n=\tau-r_k]}
\end{aligned}$$

Genel olarak MCMC yöntemleri hedef dağılımların yerel kiplerinde takılabilir. Bu yüksek oranda parametre kümesinin ve

rastgele değişkenlerin başlangıç değerleriyle doğrudan ilgilidir. Bütün rastgele değişkenlerin başlangıç değerleri orjinal modelden elde edilse bile, ki bu hiper-parametrelerin bilinmesi veya doğru kestirilmesi anlamına gelmektedir, örnekleycinin yerel bir enbüyükte takılma olasılığı her zaman vardır. Bu durumu engellemek amacıyla kullanılan yöntemlerden birisi aşağıda verilmektedir:

Döngülerin başında örnekler tam koşullu  $p(r_k|.)$  dağılımından değil, bunun bir katı olan  $p(r_k|.)^\beta$  dağılımından çekilir.  $\beta$  parametresi küçük bir değerden başlar ve döngü sayısı arttıkça 1'e doğru yaklaşır.  $\beta$  küçük olduğunda tam koşullu dağılımın doruk noktaları sönmülenir ve daha düz bir hale gelir, böylece daha az olası örneklerin de çekilebilmesi sağlanır.  $\beta$  1'e doğru yaklaştıkça, örnekler gerçek tam koşullu dağılımdan çekilir. Buna Benzetimli Tavlama (BT) denir [3, 6]. Eğer değişkenlerin başlangıç değerleri birleşik dağılımın yerel bir enbüyüğüne götürüyorsa, bu yöntemle dağılımı düzleştirerek örnekleycinin yerel bir enbüyükte takılı kalması sağlanabilir ve dağılımın birincil kipinden örnek çekmeye başlaması beklenebilir. Ayrıca  $\beta$  arttıkça, birincil kip yani global enbüyükte kalır. Fakat bu yöntemle bile örnekleycinin sonunda birincil kipi bulacağına garantisi yoktur.

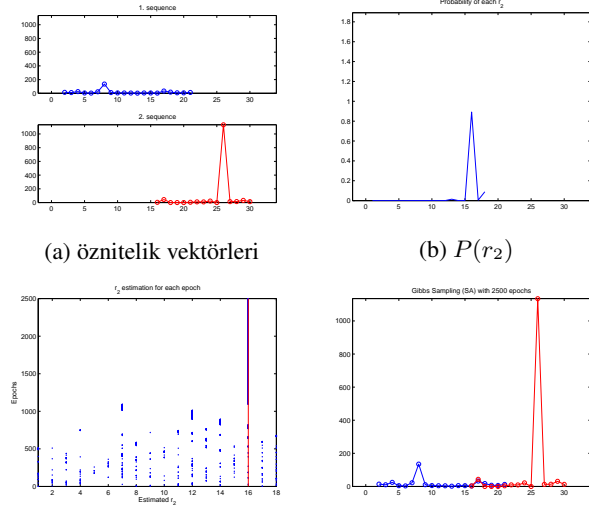
Başka bir BT stratejisi de döngüler sırasında  $\alpha$  parametresini değiştirmek olabilir. Dikkat edilirse, tam koşullu dağılım  $p(r_k|.)$ 'in değışintisi  $\alpha$  parametresiyle ters orantılı olduğu görülebilir. Dolayısıyla küçük bir  $\alpha$  değerinden başlayarak, orjinal değere doğru arttırmak da yine bir tavlama yöntemidir.  $\alpha$  parametresi küçükken, tam koşullu dağılımın değışintisi yüksektir ve bu nedenle daha düz bir dağılıma dönüşür.  $\alpha$  orjinal değerine yaklaştıkça, örnekler gerçek tam koşullu dağılımdan çekilecektir.

### 4. SONUÇLAR

DeneySEL sonuçlar, hem yapay hem gerçek verilerle elde edilmiştir. Yapay veri, hiper-parametre kümesi  $\{\alpha_\lambda, \beta_\lambda, \alpha\}$  ile öne sürülmüş modelden elde edilmiştir. Gibbs örnekleycisinin herhangi bir tavlama olmaksızın sonsal dağılımın yerel eniyeye sıklıkla takıldığı gözlemlenmiştir. Öte yandan benzetimli tavlama yaklaşımı uygulandığında, tavlama stratejisine bağlı olarak daha iyi bir başarıya ulaşılmıştır. Daha önce de belirtildiği gibi, örnekleycinin sonsal dağılımın öncül kipinden örnek alacağını garantilemek mümkün değildir. Bu sıkıntıyı ortadan kaldırmanın tek yolu Gibbs örnekleycisini farklı başlangıç noktalarından çok defalar çalıştırmak ve paysal olabilirlik dağılımını,  $p(r_{1:K} | x_{1:K,0:N_k-1})$ , kullanarak en uygun çıktıyı kabul etmektir.

Önerilen model bazı örnekler kayıp olduğunda bile, sinyalleri başarıyla eşleştirmektedir. Şekil 2'de böyle bir durum sergilenmektedir. Burada kayıt 2'nin kayıt 1 ile örtüşen örneklerinden bazıları silinmiştir. Kayıtlar Şekil 2.a da gösterilmektedir. Şekil 2.b,  $r_1$  verildiğinde her olası hizalama için  $r_2$ 'nin olasılığını göstermektedir. Yukarıda değinildiği gibi benzetimli tavlama, tam koşullu dağılımı düzleştirir ve tavlama dağılımdan alınan örnekler, birbiçimli örneklerden daha iyi sonuç vermektedir. Fakat döngü sayısı arttıkça, örnekler orjinal tam koşullu dağılımdan çekilir ve böylece örnekleyci çoğunlukla dağılımın kipinden örnekler alır. Şekil 2.b'de  $r_2$  kestirimi, her döngü için gösterilmiştir. Şekil 2.d kayıtlar arasındaki son eşleşmeyi göstermektedir. Gerçek veri benze-

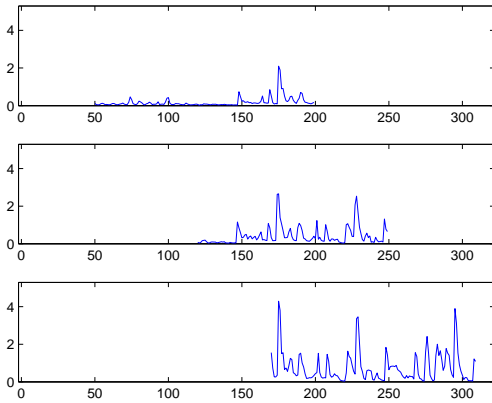
timi için ilk adım öznitelik yöneylerini çıkarmaktır. Veriler, ses kanalları, örtüşmeyen 25ms uzunluğundaki pencelerle bölünerek  $F_s = 8\text{kHz}$  ile örneklenecek ve her pencerenin spektrum büyüklüğü bulunarak elde edilmiştir.



(b) i'nci döngüdeki  $r_2$  kestirimi (c) Gibbs ile eşleşmiş sonuçlar

Şekil 2: sentetik verideki eşleşme

Daha sonra, spektrum büyüklüğü sıklık aralıkları [400Hz - 800Hz],[800Hz - 1600Hz] ve [1600Hz - 3200Hz] olan 3 oktava ayrılmıştır. Katsayılar, tüm oktav için sıklık üzerinden toplanarak öznitelik yöneyi oluşturulmuştur. Bu işlemin sonunda her kaynak 3 spektral öznitelik yöneyi ile gösterilmiştir. Bunun tek band durumundan farkı, bu kez yordam öznitelik yöneyleri yerine öznitelik dizeylerini eşleştirecektir. Şekil 3'de birinci oktav için benzetim sonuçları gösterilmektedir.



Şekil 3: 2000 döngülü Gibbs kullanarak gerçek verideki eşleştirme

Burada çoklu ses eşleştirme sistemleri için olasılıksal bir model önerilmiş ve Gibbs örnekleme için oluşturulmuş tavlama yaklaşımı incelenmiştir. Gerçek ve sentetik veriler

üzerine benzetimlerle belirsiz durumlarla nasıl başedileceği gösterilmiştir. Ayrıca örnekleme başarımı, hiper-parametre kümesi  $\{\alpha_\lambda, \beta_\lambda, \alpha\}$  seçimlerine bağlı olduğu ve bazı durumlarda örnekleme başlangıca bağlı olarak, yerel eniyi kısımlarda takılı kaldığı tespit edilmiştir.

İlerideki çalışmalarda, önerilen modelin, çıkarımını hızlandırmak, modelin hiper-parametrelerini eğitmek, EM gibi değişik yaklaşık çıkarım yöntemlerini kullanmak ve daha gelişmiş tavlama stratejileri üretmek gibi birçok ilginç yönde araştırma yapmak mümkündür.

## 5. Teşekkür

Bu çalışma 2007K120610 numaralı TAM Projesi kapsamında Türkiye Devlet Planlama Teşkilatı tarafından desteklenmiştir.

Ali Taylan Cemgil, bu çalışmada, TÜBİTAK tarafından 110E292 Bayesci Tensor ayrıştırma (BAYTEN) projesi kapsamında desteklenmektedir.

## 6. KAYNAKÇA

- [1] Wang, A.L., "An Industrial-Strength Audio Search Algorithm", 2008.
- [2] M. Muller and F. Kurth and M. Clausen, "Audio Matching via Chroma-based statistical features", 2003
- [3] Cemgil, A. T. and Fevotte, C. and Godsill, S. J. , "Variational and Stochastic Inference for Bayesian Source Separation", Digital Signal Processing, Vol. 17,no. 5 2007, p 891-913.
- [4] Troughton, P.T.; Godsill, S.J.; , "Bayesian model selection for time series using Markov chain Monte Carlo," Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on , vol.5, no., pp.3733-3736 vol.5, 21-24 Apr 1997 doi: 10.1109/ICASSP.1997.604681
- [5] Kashino, K.; Godsill, S.J.; , "Bayesian estimation of simultaneous musical notes based on frequency domain modelling," Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on , vol.4, no., pp. iv-305- iv-308 vol.4, 17-21 May 2004 doi: 10.1109/ICASSP.2004.1326824
- [6] Soo-Young Lee; Kyung Geun Lee; , "Synchronous and asynchronous parallel simulated annealing with multiple Markov chains," Parallel and Distributed Systems, IEEE Transactions on , vol.7, no.10, pp.993-1008, Oct 1996 doi: 10.1109/71.539732