

İŞARET HİZALAMA : DIŞBÜKEY OLMAYAN ENİYİLEMAYA DAYALI BİR YÖNTEM

Signal Alignment : A Method Based on Non-Convex Optimization

İlker Bayram

İstanbul Teknik Üniversitesi
ibayram@itu.edu.tr

ÖZETÇE

Bir kaynak işaretinin iki gözleminin hizalanması problemi ele alıyoruz. İşaretler arasındaki gecikmeyi kestirmeye ihtiyaç duymayan bir yöntem öneriyoruz. Yöntem, gözlem işaretlerinden birinin kısa zamanlı Fourier dönüşümü katsayılarına etki edecek bir faz haritası çıkarmaya dayanmaktadır. Faz haritasının çıkarılmasını konveks olmayan basit bir eniyileme problemi olarak düzenliyoruz. Bildiride ayrıca, hizalama yönteminin kaynak işaretin gözlemleri arasındaki gecikmeyi ortadan kaldırırken gürültünün istatistiksel özelliklerini nasıl koruduğunu da tartışıyoruz.

ABSTRACT

We consider the problem of aligning two observations of a source signal. We propose a method that does not require to explicitly estimate the time-delay between the observations. The method relies on extracting a phase map that acts on the short time Fourier coefficients of one of the observations. We formulate the phase map extraction step as a simple non-convex minimization problem. We also discuss how the proposed method preserves the statistical properties of the unwanted noise, while aligning the source of interest in both observations.

1. GİRİŞ

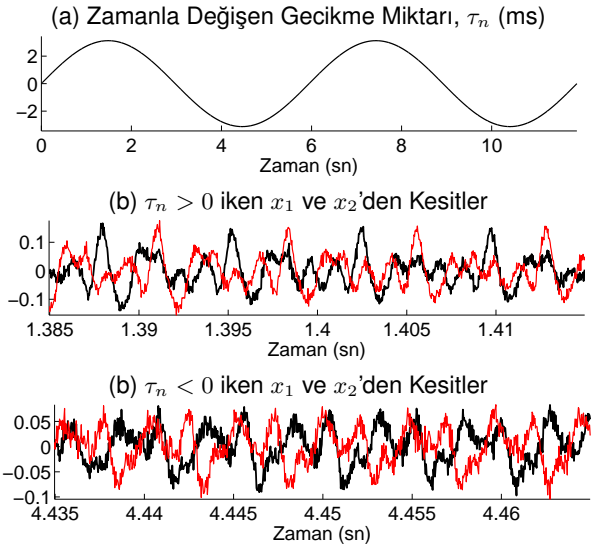
Bu çalışmada, bir ses işaretinin, iki mikrofonla elde edilen gürültülü gözlemlerinin hizalanması problemi ele alıyoruz. Bu problem için, işaretlerin kısa zamanlı Fourier dönüşümü (KZFD) domenindeki görüntülerinden faydalanarak ‘düzenleştirilmiş’ (regularized) bir faz haritası çıkarmayı ve bu faz haritasını kullanarak hizalamayı sağlamayı öneriyoruz. Bu haritanın çıkarılmasını basit bir konveks olmayan eniyileme problemi olarak düzenliyoruz.

Problemi bir örnek üzerinden anlatabiliriz. $s(n)$, gözlemlemek istediğimiz kaynak işareti olsun. Gözlem işaretlerimiz ise, τ_n , zamana bağlı gecikme miktarı, u_i ’ler de gürültü işaretleri olduğu halde,

$$x_1(n) = s(n) + u_1(n), \quad (1a)$$

$$x_2(n) = s(n - \tau_n) + u_2(n) \quad (1b)$$

Bu çalışma TÜBİTAK tarafından desteklenmektedir.



Şekil 1: Denklem 1’de belirtilen modele göre oluşturulmuş işaretler. (a) Zamana bağlı gecikme miktarı, τ_n . (b) x_1 (kalın siyah) ve x_2 sinyallerinin (ince kırmızı) τ_n ’nin pozitif olduğu (ve en büyük değerine yaklaştığı) anlardan birindeki kesitleri. (c) x_1 ve x_2 ’nin τ_n ’nin negatif olduğu (ve en küçük değerine yaklaştığı) anlardan birindeki kesitleri.

şeklinde olsun (bkz Şekil 1). Bu gözlemlerden kaynak işaretini kestirmek için ‘hüzme oluşturma’ yöntemlerinden faydalanabiliriz [2]. Fakat, bunun için, öncelikle, x_2 ’deki τ_n ’den kaynaklanan gecikmeleri ortadan kaldırmamız gerekmektedir. Bir başka deyişle, Γ ile göstereceğimiz bir operatörü, etkisi üzerinden

$$(\Gamma x_2)(n) = x_2(n + \tau_n) \quad (2)$$

şeklinde tanımlarsak, amacımızın bu operatörü gerçekleştirmek olduğunu söyleyebiliriz. τ_n ’i bildiğimiz veya kestirebildiğimiz (farklı yöntemler için bkz. [1, 3]) durumda Γ operatörünü kolayca gerçekleştirebiliriz. Bu çalışmada tartıştığımız yöntem, τ_n ’leri açık bir şekilde kestirmeden kaynak işaret gözlemlerini hizalamayı amaçlamaktadır.

1.1. Notasyon

$g(n)$, N örnek uzunluğunda bir pencere fonksiyonu olduğu halde $g_s(n)$ fonksiyonunu, $s = 0, 1, \dots, N - 1$ için

$$g_s(n) = g(n) \exp\left(j \frac{2\pi}{N} s n\right), \quad (3)$$

olarak tanımlayalım. Bu durumda x 'in KZFD'sini

$$X(k, s) = \langle x(n), g_s(n - kT) \rangle = \sum_n x(n) g_s^*(n - kT) \quad (4)$$

olarak tanımlıyoruz. Burada, k ve s sırasıyla zaman ve frekans değişkenleridir. k değişkeninin, $[1, K]$ aralığında tamsayı değerleri aldığı varsayıyoruz. Ayrıca, N ve T , KZFD'nin parametrelerini oluşturmaktadır. Bildiri boyunca zaman domenindeki işaretler küçük harfle ($x(n)$ gibi), KZFD domenindeki işaretlerse büyük harfle ($X(k, s)$ gibi) gösterilecektir.

2. PROBLEM DÜZENLEMESİ

Denklem (1)'i KZFD katsayıları cinsinden yaklaşık olarak

$$X_1(k, s) = S(k, s) + U_1(k, s), \quad (5a)$$

$$X_2(k, s) = A(k, s) S(k, s) + U_2(k, s) \quad (5b)$$

şeklinde yazabiliriz. Burada $A(k, s)$, τ_n 'ler tarafından yaklaşık olarak

$$A(k, s) = \exp\left(j \frac{2\pi}{N} \bar{\tau}_k\right); \quad \bar{\tau}_k = \frac{1}{N} \sum_{n=k}^{(k+1)N} \tau_n \quad (6)$$

eşitliklerince tanımlanabilir. Ayrıca girişte tartıştığımız Γ operatörünün KZFD domenindeki görüntüsünü $\hat{\Gamma}$ olarak ifade edersek, yine yaklaşık olarak

$$(\hat{\Gamma} X_2)(k, s) = A^*(k, s) X_2(k, s) \quad (7)$$

yazabiliriz.

(6)'daki ' A ' fonksiyonuna dair şu gözlemleri yapabiliriz.

- (i) Tüm (k, s) çiftleri için $|A(k, s)| = 1$.
- (ii) k 'yi sabit olarak düşünürsek, $z = \exp\left(j \frac{2\pi}{N} \bar{\tau}_k\right)$ olduğu halde $A(k, s) = z^s$.
- (iii) Eğer τ_n , ' n 'ye göre yavaş değişen bir fonksiyonsa, ' s ' sabit olarak düşünüldüğünde, $A(k, s)$, ' k 'ye göre yavaş değişen bir fonksiyon olacaktır.

(i) ve (ii), Γ operatörünün zamanla değişen bir geciktirme operatörü olmasından kaynaklanmakta ve ayrıca bu durum için önkoşul oluşturmaktadır¹. (iii) ise, işaret kaynağının veya mikrofonların fiziksel olarak yavaş hareket etmesinin getirdiği bir kısıttır. Bütün gözlemleri (veya kısıtları) göz önünde bulunduran bir düzenleme mümkün olsa da ortaya çıkan problemi sayısal olarak çözmenin zor olacağını düşünüyoruz. Özellikle, (ii)'deki kısıt değişken değiştirerek basit bir şekilde ele alınabilse de, bu durumda karşımıza çok salınımlı ilinti fonksiyonları çıkmaktadır [1]. Bu tip çok salınımlı fonksiyonların varlığı, işaret gürültü oranının (İGO) düşük olduğu durumlarda, en küçükleme problemlerinin çözümünü zorlaştırmaktadır. Bu nedenle, (ii)'deki kısıtı gözardı ederek A 'yı kestirme problemini şöyle düzenliyoruz :

$$\bar{A} = \arg \min_{|\alpha(k, s)|=1} \sum_{s=0}^{N-1} \left[\sum_{k=1}^K |X_1(k, s) - X_2(k, s) \alpha(k, s)|^2 + \lambda \sum_{k=2}^K |\alpha(k, s) - \alpha(k-1, s)|^2 \right]. \quad (8)$$

¹Bu noktada (i)'in, (ii)'nin sonucu olduğunu not edebiliriz. Fakat bunun tersi doğru değildir.

2.1. Eleştiriler

Eniyileme problemini sayısal olarak çözen algoritmaya geçmeden önce düzenlemeye dair birkaç soru ortaya atabiliriz.

- (I) Düzenlemede (ii) gözardı edildiği için, ortaya çıkan faz haritasının (zamanla değişen) bir geciktirme operatörüyle birebir ilişki içinde olmasını bekleyebilir miyiz?
- (II) (8)'deki düzenlemede, $\lambda = 0$ alındığında çözümün,

$$\bar{A} = \frac{X_1}{X_2} \frac{|X_2|}{|X_1|} \quad (9)$$

olduğunu gösterebiliriz. Bu çözüm analitik olarak elde edilebilirken $\lambda > 0$ olduğunda yinelemeli bir algoritmaya ihtiyacımız vardır. Düzenleme teriminin, elde edilen çözümü iyileştirmeye önemli bir katkısı var mıdır?

- (III) Elde edilen faz haritası τ_n 'i kestirmemizi kolaylaştırır mı?

(I) ve (II)'yi yanıtlamadan önce, rastgele süreçlere dair bir sonucu hatırlayalım [6].

Önerme 1. $u(n)$ her n için Gaus dağılımına sahip, durağan bir rastgele süreç, $\tilde{u}(n)$ de u 'dan bağımsız başka bir durağan süreç olsun. Ayrıca, $h(n)$ sabit bir dizi olduğu halde, $v = u * h$ olarak tanımlansın. Eğer $|H(\omega)| = 1$ ise, yani h ile evrişim, birimcil (unitary) bir işlemse,

(a) v , u ile aynı dağılıma sahiptir.

(b) v , \tilde{u} 'dan bağımsızdır. \square

Önerdiğimiz düzenleme, frekans domeninde birim genliğe sahip, zamanla-değişen bir evrişim operatörü çıkarmaya dayanmaktadır. İşaret gürültü oranının (İGO) çok düşük olmadığını varsayarsak, operatörü belirleyen maliyet fonksiyonundaki

- 'veri terimi' (yani (8)'deki ilk terim), kaynak işaretinin baskın olduğu zaman-frekans bölgelerinde \bar{A} 'nın büyük oranda kaynak işaretindeki gecikme tarafından belirlenmesini sağlarken,
- 'düzenleme terimi' (yani (8)'deki ikinci terim), kaynak işaretinin gürültüye göre zayıf olduğu bölgelerde, \bar{A} 'nın gürültü süreçlerinin aldığı değerlerden (realization) bir seviyeye kadar bağımsız olmasını sağlamaktadır.

Bu sayede, \bar{A} ile çarpmayı, yaklaşık olarak, U_1 ve U_2 'den bağımsız birimcil bir operatör olarak düşünebiliriz. Bir başka deyişle,

$$S(k, s) \bar{A}(k, s) \approx S(k, s) A(k, s) \quad (10)$$

olduğu halde, $\bar{U}_2(k, s) = U_2(k, s) \bar{A}(k, s)$ olarak tanımlandığında, \bar{u}_2 'nin u_2 ile istatistiksel olarak benzer özelliklere sahip olacağını ve eğer u_2 , u_1 'den bağımsızsa, \bar{u}_2 'nin de u_1 'den (yaklaşık olarak) bağımsız olacağını düşünüyoruz. Her ne kadar (I)'deki soruya yanıtımız olumsuz ise de, bu tartışmaya dayanarak, düzenlemenin uygulamamız için uygun olduğunu söyleyebiliriz.

Bu noktaları (ve (III)'teki soruyu) ayrıca Bölüm 4'te bir deney etrafında tartışacağız. Fakat önce, (8)'de önerdiğimiz problemi sayısal olarak nasıl çözebileceğimizi ele alalım.

3. DÜZENLEMENİN SAYISAL ÇÖZÜMÜ

(8)'deki problem, $|\alpha(k, s)| = 1$ kısıtı nedeniyle dışbükey olmayan bir problemdir. Fakat maliyet fonksiyonunun ve kısıtın doğası gereği 'coordinate descent' tipi algoritmaların [5] uygulanmasına elverişlidir. Bunu görmek için k ve s koordinatlarını sabitlediğimizi düşünelim ve d_1, d_2, a_0, a_1 sabitlerini

$$\begin{aligned} d_1 &= X_1(k, s), & d_2 &= X_2(k, s), \\ a_0 &= \alpha(k-1, s), & a_1 &= \alpha(k+1, s) \end{aligned} \quad (11)$$

şeklinde tanımlayalım. Bu durumda, problemin (k, s) koordinatları üzerindeki görüntüsü

$$\min_{|\alpha|=1} |d_1 - \alpha d_2|^2 + \lambda (|\alpha - a_0|^2 + |\alpha - a_1|^2) \quad (12)$$

olacaktır. Bu problemi, birkaç ara adımdan sonra,

$$\min_{|\alpha|=1} |\alpha - (d_2^* d_1 + \lambda (a_0 + a_1))|^2 \quad (13)$$

şeklinde yazabiliriz. Bu problem de dışbükey değildir. Fakat $d_2^* d_1 + \lambda (a_0 + a_1) \neq 0$ olduğu durumlarda bu problemin çözümü tektir ve

$$\alpha^* = \frac{d_2^* d_1 + \lambda (a_0 + a_1)}{|d_2^* d_1 + \lambda (a_0 + a_1)|} \quad (14)$$

şeklinde. $d_2^* d_1 + \lambda (a_0 + a_1) = 0$ olduğunda ise, $|\alpha| = 1$ denkleğini sağlayan tüm karmaşık sayılar problemi çözer. Bu gözlemlerden yola çıkarak yakınsak bir 'coordinate descent' algoritması [5] önerebiliriz.

Algoritma 1 (8)'deki Problemi Çözen Bir Algoritma

$R(k, s) \leftarrow X_2^*(k, s) X_1(k, s), \forall s, \forall k$

tekrarla

herbir $i \in \{0, 1\}$ için

$b(k, s) \leftarrow \alpha(k-1, s) + \alpha(k+1, s), \forall s, \forall k$

$\alpha(k, s) \leftarrow R(k, s) + \lambda (b(k, s)), \forall s, \forall k \in (2\mathbb{Z} + i)$

$\alpha(k, s) \leftarrow \alpha(k, s)/|\alpha(k, s)|, \forall s, \forall k \in (2\mathbb{Z} + i)$

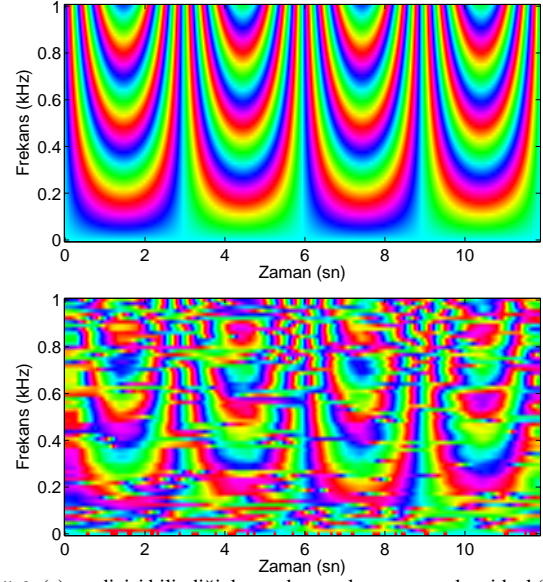
yap

yakınsayana kadar

4. DENEY ve TARTIŞMA

Deney 1. Bu deneyde kaynak işareti (yani $s(t)$) olarak Şekil 3'te spektrogramından bir parça gösterilen işareti kullandık. İşaret, yaklaşık 12 saniyelik, telli bir çalgıyla çalınan bir parçadan oluşmaktadır. Bu işareti kullanarak gözlem işaretlerini (1)'deki modele göre oluşturduk. τ_n dizisi, Şekil 1a'da gösterilmiştir. Gürültü miktarını, İGO = 10 dB olacak şekilde seçtik.

τ_n dizisini (6)'da kullanarak elde ettiğimiz ideal faz haritası (yani (6)'daki A), Şekil 2a'da gösterilmiştir. Şekilde farklı faz değerleri farklı renk tonlarıyla (hue) karşılanmıştır. (8)'i kullanarak sıfırdan büyük bir λ değeri ile elde ettiğimiz faz haritası (yani \bar{A}) ise Şekil 2b'de gösterilmiştir. Bu faz haritasında, dikey eksen boyunca süresizlikler görüyoruz. Bu, elde edilen faz haritasının bir geciktirme operatörüne karşılık gelmediğine işaret etmektedir (bkz. Bölüm 2.1'de soru (I)). Fakat, yatay eksen boyunca değişimin yavaşlığı nedeniyle operatörün yaklaşık



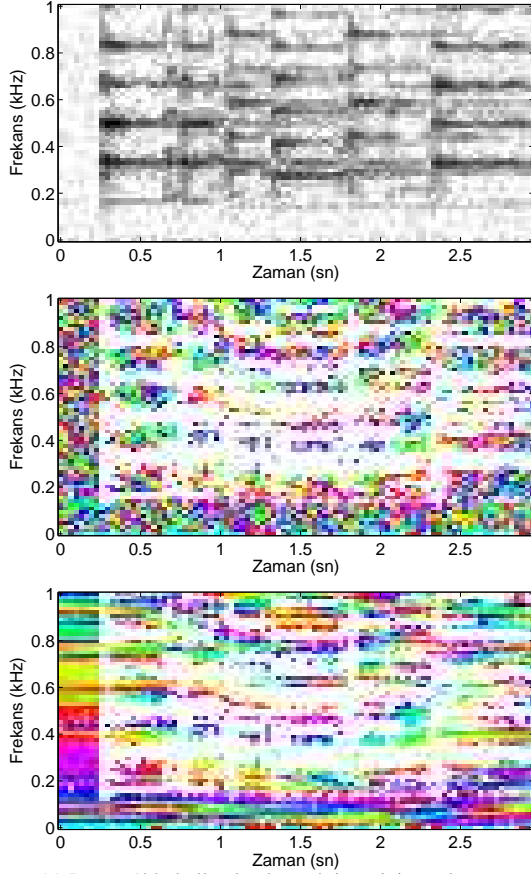
Şekil 2: (a) τ_n dizisi bilindiği durumda uygulanması gereken ideal (bkz. (6)) faz haritası. (b) Önerilen düzenleme sonucu elde edilen faz haritası. (a) ve (b)'deki faz haritaları, özellikle İGO'nun düşük olduğu bölgelerde ayrılmaktadır.

olarak birimcil olduğunu söyleyebiliriz. Bu faz haritasını kullanarak hizalanan işaretlerden kesitler Şekil 4'te gösterilmiştir.

Faz haritasını dikkatli bir şekilde incelediğimizde, dikey eksen boyunca görülen süresizliklerin işaretin gücünün düşük olduğu bölgelerde, yani gürültünün görece olarak baskın olduğu bölgelerde oluştuğunu görüyoruz. Bu noktayı Şekil 3'te göstermeye çalıştık. Şekil 3a'da gürültülü gözlem işaretinin bir parçası gösterilmektedir. Şekil 3b ve c'de ise, sırasıyla $\lambda = 0$ ve $\lambda > 0$ alındığında elde edilen faz haritaları aynı zaman-frekans bölgesi için gösterilmiştir. Her iki faz haritasında da, İGO'nun yüksek (yani işaretin baskın) olduğu bölgelerin 'doyunluk' (saturation) değerini artırarak bu bölgelerdeki katsayıların daha soluk görünmesini sağladık. Böylelikle, gürültünün baskın olduğu bölgelerde faz haritalarının nasıl oluştuğunu daha rahat görebiliyoruz. $\lambda = 0$ aldığımızda, faz haritası rastgele bir görüntü sergilemektedir (bkz. Şekil 3b). Bu, U_2 ve U_1 'in fazlarının birbirinden bağımsız ve yaklaşık olarak düzgün (uniform) dağılıma sahip olmasından kaynaklanmaktadır. Bu faz haritası, sadece gürültülerin aldığı değerler tarafından belirlendiği için ilgili operatörün, gürültü terimlerini bağımlı hale getireceğini söyleyebiliriz. Öte yandan, Şekil 3c'deki haritada, rastgele dağılımın ortadan kalktığını ve belirli bir frekans bandında zamanla değişimin düştüğünü görüyoruz. Dolayısıyla bu faz haritasını kullanarak elde edilen operatörün, gürültü terimlerinden yaklaşık olarak bağımsız olduğunu söyleyebiliriz. Bu sayede gürültü terimleri arasındaki bağımsızlık, faz haritası x_2 'ye uygulandığında da yaklaşık olarak korunacaktır.

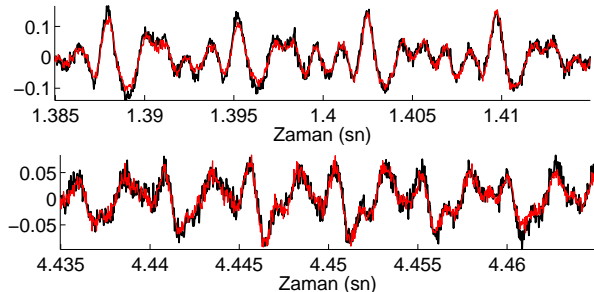
Bu iddiamızı desteklemek için basit bir test yaptık. $\lambda = 0$ ve $\lambda > 0$ durumlarında elde edilen faz haritalarını kullanarak işaretleri senkronize ederek, sırasıyla z_0 ve z_+ işaretlerini elde ettik. Bu işaretlerin İGO'su, sırasıyla 10.46 ve 10.09 dB'dir. Bu noktada, ideal koşullarda, hizalama işleminin İGO'yu etkilememesi gerektiğini hatırlatmak istiyoruz². z_0 ve z_+ 'nın

² z_0 'ın İGO'sundaki artışın kullandığımız KZFD'nin birimcil olmasından kaynaklandığını düşünüyoruz.

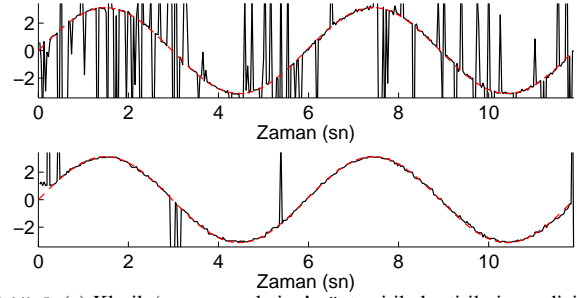


Şekil 3: (a) Deneysel 1’de kullanılan kaynak işaretinin spektrogramından bir parça. ‘ $\lambda = 0^\circ$ ’ (b) ve ‘ $\lambda > 0^\circ$ ’ (c) alındığında elde edilen faz haritaları. (b) ve (c)’de, spektrogramın genliğinin yüksek olduğu bölgeleri daha soluk göstererek, gürültü terimlerine uygulanacak faz değerleri öne çıkarılmaktadır.

x_1 ile ortalamasını aldığımızda (bu basit bir hüzmeye oluşturma yöntemidir [2]) ise, İGO’lar sırasıyla 10.88 ve 12.94 dB olmuştur. $\hat{s}_0 = (z_0 + x_1)/2$ işaretinin İGO’sunun düşük çıkmasının, z_0 ’daki gürültü teriminin, x_1 ’deki gürültü terimiyle bağımlı hale gelmesinden kaynaklandığını düşünüyoruz. Öte yandan, τ_n ’leri bildiğimiz durumda elde edeceğimiz hizalanmış işaretlerin ortalamasını alarak İGO’nun ancak 3.01 dB artmasını sağlayabilecektik. Dolayısıyla $\hat{s}_+ = (z_+ + x_1)/2$ için elde edilen İGO, z_+ işaretindeki gürültü teriminin, ideal durumda elde edebileceğimiz hizalanmış işaretteki gürültü terimine *istatistiksel* olarak benzediği yönündeki savımızı desteklemektedir. Daha detaylı bir çözümlemenin ilginç olabileceğini düşünüyoruz fakat yer darlığı sebebiyle tartışmayı burada son-



Şekil 4: Önerilen yöntemle hizalanmış işaretler. Şekil 1 ile aynı kesitler gösterilmiştir.



Şekil 5: (a) Klasik ‘cross-correlation’ yöntemi ile kestirilmiş τ_n dizisi. (b) Önerilen yöntemle çıkarılmış faz haritasını kullanarak kestirilmiş τ_n dizisi. Her iki şekilde de gerçek τ_n dizisi kesikli kırmızı çizgiyle gösterilmiştir.

landırıyoruz.

Son olarak, elde edilen faz haritasının τ_n ’i kestirme problemini kolaylaştırıp kolaylaştırmadığını kısaca tartışmak istiyoruz. τ_n dizisini klasik ‘cross-correlation’ [3] yöntemiyle kestirerek Şekil 5a’daki diziyi elde ettik. Elde ettiğimiz faz haritasını $[X_2]$ ile ağırlıklandırarak ‘generalized cross-correlation’ yöntemiyle [3] ise Şekil 5b’deki diziyi elde ettik. Kestirimdeki yüksek hataların sayısının önemli ölçüde azaldığını görüyoruz.

5. SONUÇ

Önerdiğimiz işaret hizalama yöntemi, işaretler arasındaki gecikmeyi hesaplamadan hizalama işlemini yapmaktadır. Bu açıdan, literatürdeki birçok yöntemden ayrılmaktadır. Bununla birlikte, düzenleme sonucu elde edilen faz haritasının gecikme kestirme problemini kolaylaştırdığını gördük. Bu çalışmada, tek bir kaynağın olduğunu varsayarak problem düzenlemesi yaptık. Birden fazla kaynağın olduğu durumda bu yöntemin tüm kaynakları hizalayacağını öngörebiliriz. Bu, tüm kaynakların ilgi dahilinde olduğu (karşılıklı bir konuşmanın kaydı gibi) durumlarda istenen bir durum olsa da, kaynakları ayırtırmayı olanaksız kılabilir. Bu durumda, faz haritasını kullanarak kaynakların yönünü kestirmenin anlamlı olacağını düşünüyoruz. Birden fazla kaynak olduğu durumda, [4]’teki varsayımlar geçerli ise (kaynakların sıfırdan farklı KZFD katsayılarının farklı zaman-frekans bölgelerini işgal etmesi), bu bildiride önerilen faz haritası çıkarma yönteminin kaynak yönlerini kestirme problemini kolaylaştıracığını düşünüyoruz.

6. KAYNAKÇA

- [1] G. C. Carter. Coherence and time delay estimation. *Proc. IEEE*, 75(2):236–255, February 1987.
- [2] S. Gannot and I. Cohen. Adaptive beamforming and postfiltering. In *Handbook of Speech Processing*. Springer, 2008.
- [3] Y. Huang, J. Benesty, and J. Chen. Time delay estimation and source localization. In *Handbook of Speech Processing*. Springer, 2008.
- [4] Ö. Yılmaz and S. Rickard. Blind separation of speech mixtures via time-frequency masking. *IEEE Trans. Signal Processing*, 52(7):1830–1847, July 2004.
- [5] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 1984.
- [6] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 3rd edition, 1991.