

ESNEK BİR DALGACIK AİLESİYLE TON KAYDIRMA PITCH SHIFTING USING A FLEXIBLE WAVELET FAMILY

İlker Bayram

Elektronik ve Haberleşme Mühendisliği Bölümü,
İstanbul Teknik Üniversitesi
ibayram@itu.edu.tr

ÖZETÇE

Müzikte transpozisyon parçanın tonlarını, harmoniklerin görelî yapısı korunduğu halde kaydırmaya denktir. Teoride, bu işlem frekansları logaritmik bir ölçekte kaydırarak gerçekleştirilebilir. Fakat pratikte, böyle bir tasarımı gerçekleştirmek için fazlalığı yüksek bir dönüşüm gereklidir. Bu makalede, ton kaydırma için, zaman-sıklık düzlemini esnek bir biçimde örnekleyebilen bir dalgacık ailesini kullanan bir yöntem öneriyoruz. Önerilen yöntem çözümleme ve birleşim için farklı dalgacık çerçeveleri kullanmaktadır. Dalgacık katsayıları üzerinde işlem yapılmaması yöntemin hızlı olmasını sağlamaktadır.

ABSTRACT

Transposing music corresponds to moving a piece up or down in pitch by some amount while maintaining the relative structure of the harmonics. In theory, this operation can be realized by shifting the frequencies on a log-axis. However, in practice, realizing such a scheme requires a very redundant transform. In this paper, we propose a method for pitch-shifting that makes use of a wavelet family that offers flexibility in specifying the time-frequency sampling pattern. The proposed method utilizes different wavelet frames for analysis and synthesis. No processing on the wavelet coefficients is performed, which makes the method fast.

1. GİRİŞ

Tek frekans bileşeninden oluşan

$$x(t) = w(t) \sin(f_0 t) \quad (1)$$

gibi bir işareti düşünelim. Bu işaretin α kadar ton kaydırılmış hali,

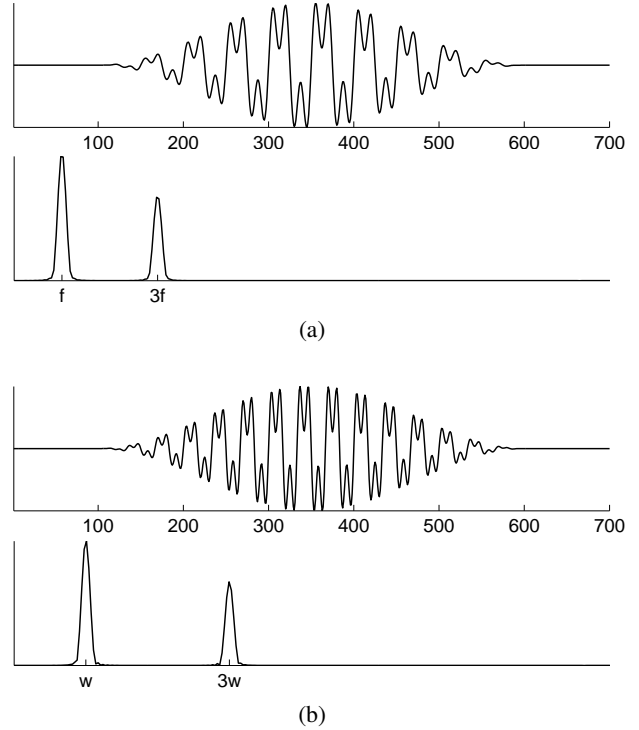
$$x_\alpha(t) = w(t) \sin(\alpha f_0 t) \quad (2)$$

olarak ifade edilebilir. Daha genel bir durumda, K bileşenli

$$x(t) = \sum_{k=1}^K w_k(t) \sin(\phi_k(t)) \quad (3)$$

Bu çalışma TÜBİTAK tarafından 110E240 numaralı proje kapsamında desteklenmektedir.

978-1-4673-0056-8/12/\$26.00 ©2012 IEEE



Şekil 1: (a) İki bileşenli pencerelemiş bir işaret (üst panel) ve spektrogramı (alt panel). Bileşenlerin frekanslarının birbirine oranının 3'e eşit olduğunu görüyoruz. (b) İşaretin ton kaydırılmış hali. Ton kaydırılmış işaretin bileşenlerinin frekansları arasındaki oran korunmuştur.

gibi bir işaretin α kadar ton kaydırılmış hali ise,

$$x_\alpha(t) = \sum_{k=1}^K w_k(t) \sin(\alpha \phi_k(t)) \quad (4)$$

olarak ifade edilebilir. Şekil 1 bu işlemi basit bir örnek üzerinde göstermektedir.

İşaretin (3) gibi bir gösteriminin elde olmadığı durumlarda ise farklı yöntemlere ihtiyaç vardır. Problemin zaman ölçekleme [1] problemiyle olan ilişkisi dolayısıyla iki farklı yaklaşımdan söz edebiliriz. Öncelikle, işareti zaman içinde ölçekleyip, örnekleme frekansını değiştirmek, istediğimiz etkiyi yaratacaktır. Dolayısıyla zaman ölçekleme yapan yöntemler

(bkz. [2, 3, 4, 5, 6, 7, 8]) ton kaydırma için de kullanılabilir. Bir diğer yaklaşım ise zaman ölçekleme yapmak yerine doğrudan ton kaydırmayı hedefler. Bu yöntemleri de kabaca ikiye ayırabiliriz. ‘PSOLA’ [1] türevi, işaretin zaman örnekleri üzerinde çalışan yöntemler [9] ve frekans ortamı yöntemler [10, 1]. Frekans ortamı yöntemler genelde kısa zamanlı Fourier dönüşümünden faydalanmaktadır. Fakat Kısa zamanlı Fourier Dönüşümü (KZFD) frekansları doğrusal olarak örneklemediği için ton kaydırma problemi için çok elverişli değildir. Bu makalede önerdiğimiz yöntem frekans eksenini logaritmik bir ölçekte örneklemediği için katsayılar üzerindeki işlem yükü çok daha düşüktür – benzer yaklaşımlar için [11, 12] incelenebilir.

2. ÖNERİLEN YÖNTEM

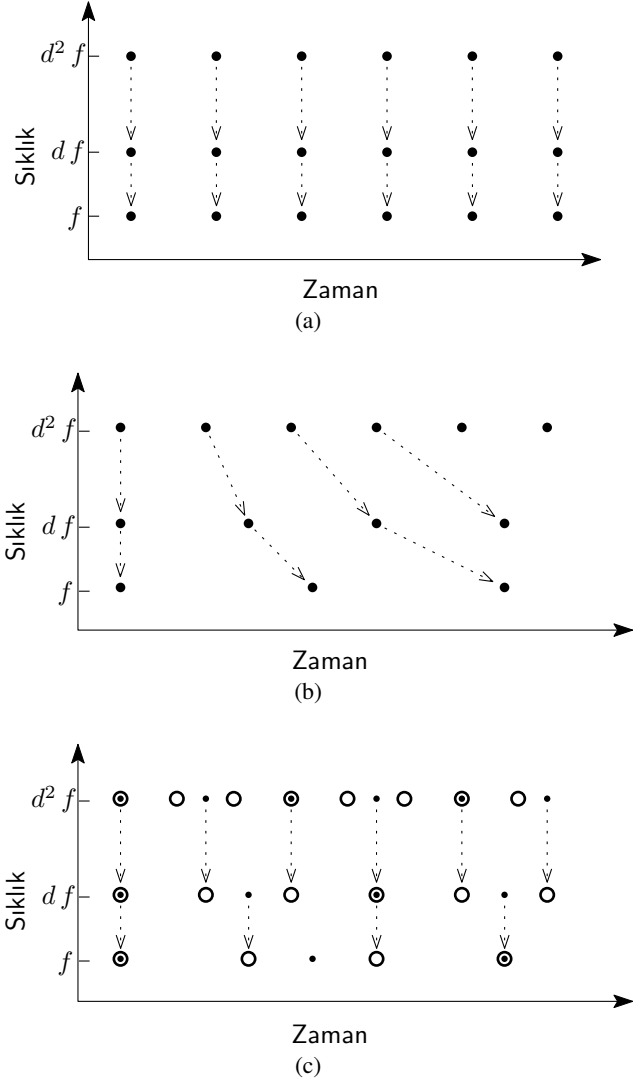
Daha önce de belirttiğimiz gibi, ton kaydırma işlemi, frekansları logaritmik bir ölçekte öteleyerek gerçekleştirilebilir. Bu işlem, işareti zaman-sıklık düzleminde Şekil 2a’daki gibi örnekleddikten sonra örnekleri (yine aynı şekilde oklarla gösterildiği gibi) öteleyerek yapılabilir (bkz. [11]). Bu yöntemin dezavantajı, altbantların genişliği değişmesine rağmen zamandaki örnekleme sıklığının değişmemesi ve bu nedenle örnek sayısının gerekenden çok daha fazla olmasıdır. Bu makalede önerdiğimiz yöntem dalgacık çerçevelerini kullandığı için zaman-sıklık örneklerini daha tutumlu bir şekilde kullanır.

Şekil 2b’yi düşünelim. Burada, noktalar tipik bir dalgacık çerçevesinin atomlarının zaman-sıklık düzlemindeki merkezlerini temsil etmektedir. Bu durumda, (‘ x ’ diye adlandıracağımız) bir işaretin dalgacık katsayılarını, işaretin zaman-sıklık düzleminde gösterilen noktalarındaki örnekleri olarak düşünebiliriz. İşaretin dalgacık dönüşümünde k ’inci altbantın n ’inci katsayısını c_n^k olarak gösterelim. Yeni bir katsayı dizisini $\tilde{c}_{n+1}^k = c_n^k$ şeklinde tanımlayıp, bu katsayıları kullanarak (\tilde{x} olarak göstereceğimiz) yeni bir işaret sentezlediğimizi varsayalım (bkz. Şekil 2b). ‘ d ’ genişleme katsayısı olduğu halde, yeni işaret \tilde{x} , orijinal işaret x ’e göre ‘ d ’ kadar ton kaydırılmış, fakat aynı esnada, zaman içinde d kadar da ölçeklenmiş olacaktır. Elde etmek istediğimiz etki ise, zaman ölçeklemesi yapmadan ton kaydırmaktır.

Şimdi de Şekil 2c’ye bakalım. Bir önceki şekildeki gibi, x işaretinin noktalarla gösterilen zaman-sıklık örnekleri elimizde olsun. Bu örneklerin oklarla gösterilen çemberlere ait olduğunu varsayarak yeni bir işaret sentezleyelim ve elde ettiğimiz işarete \tilde{x} diyelim. Yeni işaret, \tilde{x} , orijinal işaretin ton kaydırılmış hali olacaktır. Dahası, bir önceki durumun aksine, işaretin zamanda işgal ettiği bölge değişmeyecektir. Bu fikri hayata geçirmek için zaman-sıklık düzlemindeki örnekleme örüntüsünü kolaylıkla değiştirebildiğimiz bir dalgacık ailesine ihtiyacımız vardır. Bir sonraki bölümde bu aileyi tanıtaçaz.

3. ESNEK BİR DALGACIK AİLESİ

Bölüm 2’de anlatılan yöntemi gerçekleştirmek için [13]’de sunulan dalgacık ailesinden faydalanacağız. Bu aile, Şekil 3’te gösterilen (i) genişleme katsayısı ‘ d ’, (ii) Q-katsayısı ‘ $\Delta\omega/f$ ’, (iii) örnekleme parametresi ‘ Δt ’ gibi üç ana parametrenin kolaylıkla değiştirilmesine izin verir. Bu dalgacık ailesi Şekil 4’te gösterilen yinelenmiş süzgeç kümelerinden faydalanılarak tanımlanmıştır. Gerçeklemede ise bu sisteme denk

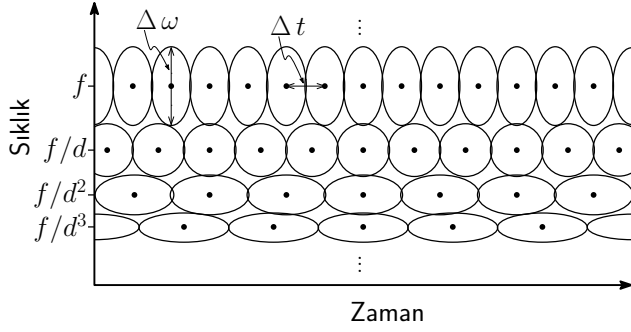


Şekil 2: (a) Ton kaydırma işlemi, işaretin logaritmik ölçekteki frekans altbantlarının kaydırılmasıyla gerçekleştirilebilir. (b,c) Noktalar (ve çemberler) bir dalgacık çerçevesinin atomlarının zaman-sıklık düzlemindeki merkezlerini göstermektedir. (b) Eğer katsayıları bir altbant kaydırırsak işareti hem zaman içinde ölçeklenmiş hem de ton kaydırmış oluruz. (c) Zaman ekseninde daha sık bir çerçeveden gelen örneklerin (noktalar) altbantları kaydırılır. Yeni altbantlar zamanda daha seyrek bir dalgacık çerçevesinin örnekleriymiş gibi işaret sentezlenir. Bu şekilde ton kaydırma yapıldığı halde zamanda ölçekleme gerçekleşmez.

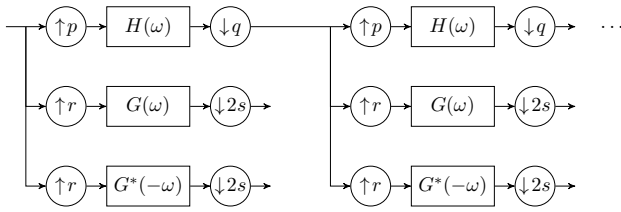
olan paralel süzgeç kümesi kullanılarak hızlı bir dönüşüm elde edilmiştir [13].

Dalgacık ailesinden iki örnek Şekil 5’te gösterilmiştir. Dalgacık çerçevesi birbirine Hilbert transformuyla ilintili dalgacıkları barındırdığından, ses ve bir boyutlu biyomedikal işaretler gibi çok salımlı işaretlerin işlenmesi için elverişlidir.

Uygulamamıza dönecek olursak, Şekil 2’de kullanılan dalgacık çerçeveleri aynı Q katsayısına ve genişleme katsayısına sahiptirler, fakat Δt parametrelerinin birbirine oranı ‘ d ’, yani genişleme katsayısına eşittir [13].



Şekil 3: Bir dalgacık çerçevesinin atomlarının zaman-sıklık düzlemindeki dağılımı yukarıdaki gibidir. Burada üç parametreye vurgu yapılabilir : (i) genişleme katsayısı ‘ d ’, (ii) Q-katsayısı ‘ $\Delta\omega/f$ ’, (iii) örnekleme parametresi ‘ Δt ’. Bu parametreleri kolayca değiştirmemize olanak sağlayan esnek bir dalgacık ailesi [13]’de önerilmiştir.



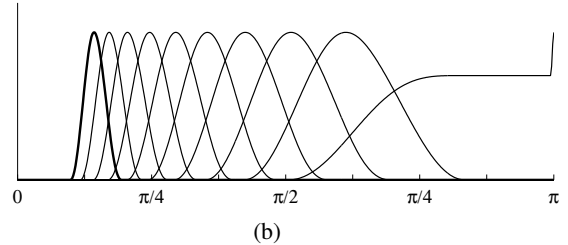
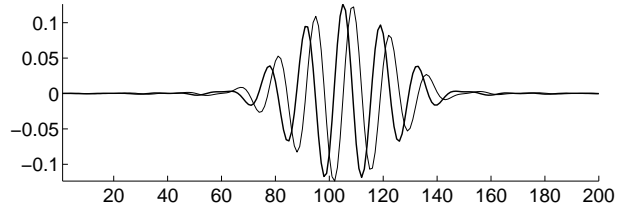
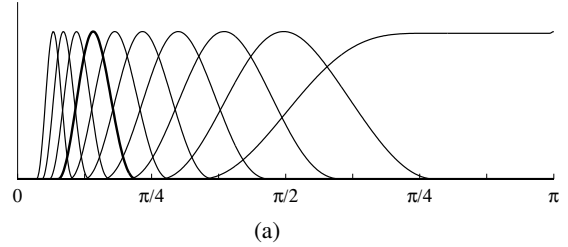
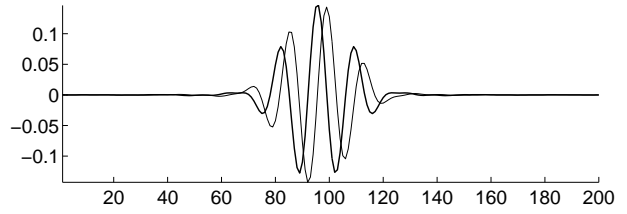
Şekil 4: Bu makalede kullandığımız dalgacık dönüşümü yukarıdaki yinelenmiş süzgeç kümesi kullanılarak gerçekleştirilmiştir. Süzgeç kümesindeki ‘yüksek geçiren’ altbantlar ($G(\omega)$ ve $G(-\omega)$ süzgeçlerini kullanan) sırasıyla işaretin pozitif ve negatif frekansları üzerine etki etmektedir. Dolayısıyla yapı, dalgacıkların Hilbert dönüşümleriyle çalışmaya elverişlidir. Daha fazla detay için bkz. [13].

4. DENEY

Önerdiğimiz yöntemi bir keman parçası üzerinde test ettik. İşareti analiz etmek için, genişleme katsayısı $d = 275/279 \approx 2^{1/48}$, Q katsayısı 41.1 ve artıklık (redundancy) katsayısı $R = 3.49$ olan bir dalgacık çerçevesi kullandık. Elde edilen katsayılar Şekil 6’nın sol panelinde gösterilmektedir. İşareti bir tam ses kaydırmak için f frekansındaki içeriği $2^{1/6} f$ frekansına taşımak gerekmektedir. $d \approx 2^{1/48}$ olması nedeniyle, işareti bir tam ses kaydırmak için, işaretin altbant katsayılarını 8 altbant kaydirdık ($d^8 \approx 2^{1/6}$). Daha sonra bu katsayıları, Q ve genişleme katsayıları aynı fakat artıklık katsayısı $d^8 R$ olan bir dalgacığın katsayıları gibi alarak yeni bir işaret sentzedik. Elde edilen işaretin ilk dalgacık çerçevesine göre katsayılarını yeniden hesapladığımızda elde ettiğimiz skalogram Şekil 6’nın sağ panelinde gösterilmiştir. Görüleceği gibi, her iki skalogram dikey eksendeki kayma dışında yaklaşık olarak aynıdır. Algısal olarak da, ton kaydırılmış işaret, keman sesi özelliğini yitirmemiştir.

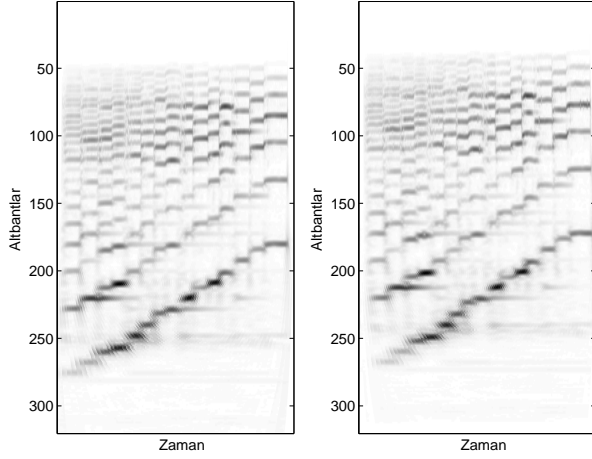
5. TARTIŞMA VE SONUÇ

Bu makalede ton kaydırma için yeni bir yöntem önerdik. Yöntem, esnek bir dalgacık ailesinden faydalanmakta,

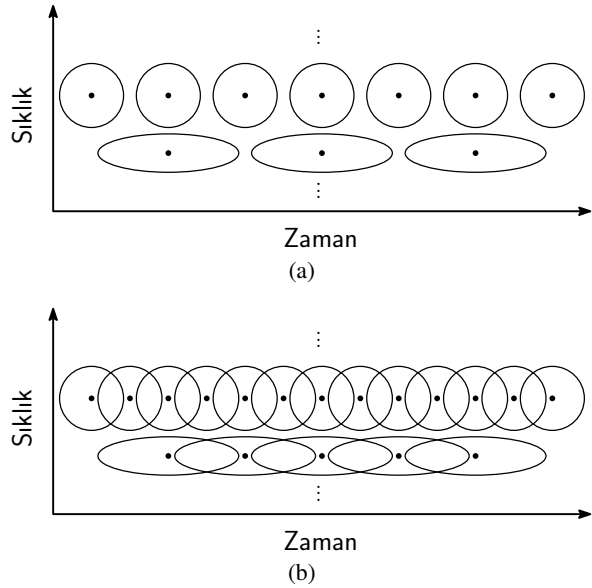


Şekil 5: Şekil 4’te gösterilen yinelenmiş süzgeç kümesinden elde edilen dalgacıklar. Üst panellerde dalgacıklar ve Hilbert dönüşümleri (sırasıyla kalın ve ince çizgilerle), alt panellerde ise dalgacık çerçevesinin frekans çözümlenmeleri gösterilmiştir. Dalgacık çerçevesinin parametreleri şöyledir : (a) Q katsayısı = 2, artıklık (redundancy) katsayısı $R = 3$; (b) Q = 3, $R = 3$. Daha fazla detay için bkz. [13].

çözümleme ve sentez için farklı dalgacık çerçevesi kullanılmaktadır. Yöntem, enstruman seslerini birkaç ‘ses’ (bkz. Kısım 4) kaydirdığımızda önemli bir bozulmaya yol açmamaktadır. Fakat yöntemin sınırlarını anlamak için bir noktaya dikkat çekmek istiyoruz. Şekil 2, dalgacık atomlarının sadece zaman-sıklık merkezlerini göstermektedir. Şekil 7, atomların zaman-sıklık desteklerini de göstermektedir. Burada, Q ve genişleme katsayısı aynı olan fakat artıklık katsayısı farklı olan dalgacık çerçevesinin atomları arasında birbirinden farklı ilişkiler olduğunu görebiliriz. Özellikle, bu farklılığı göz önünde bulundurarak dalgacık katsayılarını basit bir işleme ile düzletecek bir yöntemin başarımı artırmakta etkili olacağını düşünüyoruz. Bu iyileştirmeyi gelecekte daha detaylı olarak çalışmayı planlıyoruz.



Şekil. 6: Sol panel: Orijinal işaretin skalogramı. Sağ panel : Ton kaydırılmış işaretin skalogramı. İdealde sağdaki skalogram soldakinin dikey ekseninde kaydırılmış hali olmalıdır. Önerdiğimiz yöntem bunu pratikte de yaklaşık olarak sağlamıştır.



Şekil. 7: Kısım 2’de dalgacık çerçevelerinin atomlarının zaman-sıklık düzlemindeki merkezleri üzerinden bir yöntem önerilmiştir. Fakat bu atomlar aslında zaman-sıklık düzleminde belirli bir alan işgal ederler. Zaman içindeki örnekleme arttırarak elde ettiğimiz çerçevedeki atomların birbiriyle ilişkisi (b), orijinal çerçevenin atomlarının birbiriyle ilişkisinden (a) farklıdır.

6. KAYNAKÇA

[1] U. Zölzer, editor. *DAFX: Digital Audio Effects*. Wiley, 2011.

[2] F. J. Charpentier, M. G. Stella. Diphone synthesis using an overlap-add technique for speech waveforms concatenation Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP) (1986).

[3] J. Laroche, M. Dolson. Phase-vocoder: About this phasiness business. Proc. IEEE Workshop on Appl. Signal Proc. to Audio and Acoustics (1997).

[4] J. Laroche, M. Dolson. Improved phase vocoder time-scale modification of audio. IEEE Trans. Signal Processing, 7(3):323–332 (1999).

[5] M. R. Portnoff. Time-scale modification of speech based on short-time Fourier analysis. IEEE Trans. Acoust., Speech, and Signal Proc., 29(3):374–390 (1981).

[6] M. Puckette. Phase-locked vocoder. Proc. IEEE Workshop on Appl. Signal Proc. to Audio and Acoustics (1995).

[7] İ. Bayram, I. W. Selesnick. A Dual-Tree Rational-Dilation Complex Wavelet Transform. IEEE Trans. Signal Proc., 59(12):6251–6256 (2011).

[8] T. F. Quatieri. Discrete-Time Speech Signal Processing, Prentice Hall (2001).

[9] H. Valbret, E. Moulines, J. P. Tubach. Voice transformation using PSOLA technique. Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP) (1992).

[10] J. Laroche, M. Dolson. New phase-vocoder techniques for pitch-shifting, harmonizing and other exotic effects. Proc. IEEE Workshop on Appl. Signal Proc. to Audio and Acoustics (1999).

[11] M. Dörfler, N. HoligHaus, T. Grill, and G. Velasco. Constructing an invertible constant-Q transform with nonstationary Gabor frames. In Proc. Int. Conf. on Digital Audio Effects (DAFx), 2011.

[12] J. E. Youngberg. Rate/Pitch modification using the constant-Q transform. In Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP), 1979.

[13] İ. Bayram. An Analytic Wavelet Transform with a Flexible Time-Frequency Covering. IEEE Transactions on Signal Processing’e gönderildi.