

# ENTROPİ TABANLI ÖZNETELİKLER İLE MÜZİK ENSTRUMANLARININ SINIFLANDIRILMASI CLASSIFICATION OF MUSICAL INSTRUMENTS WITH ENTROPY BASED FEATURES

*Mehmet Erdal Özbek*

Elektrik ve Elektronik Mühendisliği Bölümü  
İzmir Yüksek Teknoloji Enstitüsü  
erdalozbek@iyte.edu.tr

## ÖZETÇE

*Zaman ve izge ortamında hesaplanan entropi değerleri kullanılarak müzik işaretlerinin değişimleri de izlenebilmektedir. Bu çalışmada her iki ortamda hesaplanan düzgelemlenmiş entropi değerlerinin öznetelik olarak kullanılması önerilmektedir. Bu entropi tabanlı öznetelikler zaman ve izge ağırlık merkezi, izge düzlüğü ve izge yayılımı gibi benzer öznetelikler ile karşılaştırılmıştır. Daha sonra da müzik enstrumanları sınıflandırılması için başarımları incelenmiştir.*

## ABSTRACT

By using the entropy values computed in temporal and spectral domain, the variations of musical signals can also be followed. In this study, the normalized entropy values computed in both domains are proposed to be used as features. These entropy based features are compared with similar features like temporal and spectral centroid, spectral flatness, and spectral spread. Then, their performances are investigated for classification of musical instruments.

## 1. GİRİŞ

Müzik bilgi çıkarımı çalışmaları, müzik işaretlerinin zamansal, izgesel, algısal, ve benzeri tüm özelliklerinin ortaya çıkarılmasını hedeflemektedir. Bu amaçla bir çok öznetelik kullanılmakta ve tanımlanan bir amaç için belirli özneteliklerin daha iyi sonuç verdiği bilinmektedir [1, 2, 3, 4]. Örneğin özintinti katsayıları işaretin temel titreşim frekansının bulunması için yaygın olarak kullanılırken [5], mel frekansı kepsral katsayıları ise müzik enstrumanları sınıflandırılmasında en iyi sonuçların elde edildiği öznetelik olarak gösterilmektedir [6, 7]. Bu ve benzeri öznetelikler müzik bilgi çıkarımı sistemlerinin belirledikleri çıktılarının elde edilmesi için işaret işleme ve örüntü tanıma teknikleri ile birlikte kullanılmaktadır.

Özneteliklerin çoğu hesaplanırken genellikle kısa ve kısmen örtüşen çerçeveler kullanılmaktadır. Bu şekilde 20-40 ms arasındaki uzunluklarda ses parçacıkları kullanılarak zaman ya da frekans ortamında öznetelikler elde edilmektedir. Bunun yanı sıra, özneteliklerin bütün işaret boyunca aldıkları değerler kullanılarak, ortalama değer, standard sapma, değişinti, eğiklik gibi istatistiksel bilgileri yanında, birinci ve ikinci dereceden

türevleri de kullanılmaktadır. Hatta bazı özneteliklerin bu bilgileri çoğultortam MPEG-7 standardında da bulunmaktadır [8, 9].

Bunların dışında bir diğer önemli gösterim de, işaretin taşıdığı bilginin ve önceden kestirilebilirliğinin ölçüsü olan entropidir [10]. Entropi kavramı bilindiği gibi öncelikle ısı dinamiğinde ve istatistiksel mekanikte kullanılmış, ancak Shannon tarafından bilgi kuramı çerçevesinde temelleri atılmıştır [11]. Entropi, bir rastgele değişkenin durumunu iletmek için gerekli bit sayısının alt sınırını oluşturmaktadır ve bu alt sınırı konuşma ve müzik işaretlerinin daha fazla sıkıştırılmayacağı azaltılamaz bir karmaşıklık belirlemektedir [10]. Sonrasında geliştirilen bağıl entropi ve ortak bilgi kavramlarının birçok örüntü tanıma ve otomatik öğrenme tekniklerinde başarıyla kullanıldığı gösterilmiştir [12].

Boyuttan bağımsız bir ölçü olmasının yanı sıra, sürekli olduğu kadar ayrık veriler için de uygulanabilir olması nedeniyle entropi müzik işaretleri için de yararlı bir öznetelik olarak kabul edilmektedir. Genel olarak konuşma işaretinin izgesel entropisinin müzik işareti izgesel entropisinden daha yüksek olması, konuşma/müzik ayrıştırması için kullanılmasını sağlamıştır [13, 14, 15]. Bunun yanında da entropinin türevinin kodlanarak içerik tabanlı gösterimleri elde edilmiş [16], ya da entropi, doğrudan müzik seslerindeki vuruların izlenmesi ve değerlendirilmesi için kullanılmıştır [17]. Çoğunlukla ise, özneteliğin belirlenmesi için küme seçimi [18], ya da bilgi kazancı, kazanç oranı ve bakışımı belirsizlik [6] gibi, var olan çok sayıda öznetelikten en iyi olanlarının seçilmesine yardımcı olmak amacıyla kullanılmıştır. Bu çalışmada ise entropinin işaretlerdeki değişimleri ifade edebilmesine dayanarak, müzik enstrumanlarının sınıflandırılması amacıyla kullanılması incelenmektedir. Zaman ya da frekans (ya da izge) ortamında elde edilen entropi değerleri düzgelemlenerek elde edilen özneteliğin sınıflandırma başarımı benzer öznetelikler ile karşılaştırılmıştır.

## 2. ENTROPİ

Bir olayın bilgi içeriğinin ölçüsü olayın olma olasılığına dayanmaktadır. Eğer  $X = x_i$  olayı  $p_i$  olasılığı ile gerçekleşmekte ise,  $X$  olayının gözlenmesinden sonra elde edilen bilgi miktarı

$$I(x_i) = \log \left( \frac{1}{p_i} \right) = -\log p_i \quad (1)$$

ile verilmektedir [10]. Shannon entropisi ise olayın  $N$  olası değeri için gözlenmesinden elde edilen bilgi miktarının beklentisi olarak tanımlanmaktadır.

$$H(X) = E[I(X)] = \sum_{i=1}^N p_i I(x_i) = - \sum_{i=1}^N p_i \log(p_i) \quad (2)$$

Bu tanımın önemli bir yansıması en büyük ve en küçük entropi durumlarında ortaya çıkmaktadır. Eğer olayın sadece bir sonucu oluyor ise, olasılık kütle işlevi belirli bir  $k$  değerindeki  $p_i = \delta(k)$  ile verilen birim darbe işlevi olmaktadır. Bu durumda olayın entropisi

$$H = - \sum_i \delta(k) \ln[\delta(k)] = - \ln(1) = 0 \quad (3)$$

olarak bulunur. Eğer olayın sonuçlarının olma olasılığı eşitse, olasılık kütle işlevi düzgün dağılıma sahip olmakta ve en büyük entropi değeri elde edilmektedir.

$$H_m = - \sum_{i=1}^N \frac{1}{N} \ln\left(\frac{1}{N}\right) = - \ln\left(\frac{1}{N}\right) = \ln(N) \quad (4)$$

### 2.1. İzgesel Entropi

Bir işaretin entropisi zaman ya da frekans ortamında, genlik ya da enerji değerlerini esas alarak hesaplanabilir. Zaman ortamında hesaplanan entropi değerleri işaretin zamandaki olasılık kütle işlevinin kestirilmesini gerektirirken, izgesel entropi için işaretin, örneğin kısa-zaman Fourier dönüşümü kullanılarak, izgesinin elde edilmesi gerekmektedir. Sonrasında Denklem (5)'te gösterildiği gibi izge düzgelenecek olasılık kütle işlevi benzeri bir işleve dönüştürülür. Burada  $X_j$ ,  $N$  seleye ayrılmış izgenin  $j$ 'ninci frekans bileşeninin enerjisini ifade etmektedir.

$$\tilde{X}_i = \frac{X_i}{\sum_{j=1}^N X_j} \quad (5)$$

Böylece, izgesel entropi düzgelenecek izge  $\tilde{X}$  kullanılarak elde edilebilir.

$$H = - \sum_{i=1}^N \tilde{X}_i \log(\tilde{X}_i) \quad (6)$$

İzgesel entropi, enerjinin frekans alt-banlarında gruplandırıldığı çoklu-çözünürlük ayrıştırılması kullanılarak da elde edilebilir. Bu durumda elde edilen entropi ise renkli (kromatik) entropi olarak adlandırılmıştır [14].

### 2.2. Düzgelenecek Entropi Ölçüsü

Entropi ifadesinin bir zorluğu, oran bağımlı olması yani değerinin olası durum ya da seleylerin sayısı ile değişmesidir. Bunu önlemek amacıyla önerilen basit bir düzeltme ise düzgelenecek entropi ölçüsünü ortaya çıkarmaktadır.

$$\gamma = \frac{H}{H_m} = \frac{H}{\ln(N)} \quad (7)$$

Bu ölçü her zaman 0 ile 1 arasında değiştiği için, farklı ortamlarda, durum ya da sele sayılarında hesaplanan entropiler karşılaştırılabilmektedir. Çerçeve kullanılarak hesaplanan ses işaretleri için entropi, çerçeve boyutu sayısındaki kadar sele

kullanılarak hesaplanacaktır. Böylece zaman işaretinden ya da izgesel ortamdaki hesaplanan entropi, Denklem (4) ile verilen en büyük entropi değeri ile sınırlandırılmış ve Denklem (7)'de verilen düzgelenecek entropi ölçüsü ile ifade edilebilir duruma getirilmiş olmaktadır.

Düzgelenecek entropi ölçüsü boyutsuzdur ve entropi ölçüsünün verimliliği olarak da ifade edilebilir. Bu sayede ortam, işaret uzunluğu, kullanılan durum ya da sele sayısından bağımsız olarak, evrensel bir ölçüt olarak kullanılabilir.

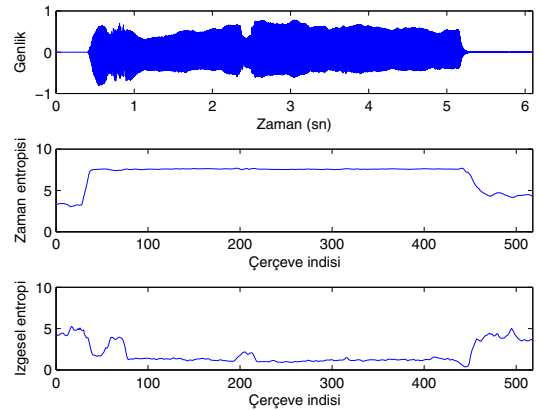
## 3. DENEYLER VE SONUÇLAR

### 3.1. Müzik Enstruman Örnekleri

Entropi ölçülerini değerlendirmek için Iowa Üniversitesi'nin yalıtılmış klasik müzik enstruman ses örneklerini [19] kullanılarak çalışmalar gerçekleştirilmiştir. Bu ses örnekleri 19 adet enstrumanın solo kaydını içermekte ve bu kayıtlar 16 bit çözünürlüklü ve 44100 Hz'de örneklenecek bulunmaktadır. Veritabanında grup halinde bulunan kayıtlar önceki çalışmalarımızdan biri için 2-3 saniye uzunluğundaki notalar bazında ayrıştırılmıştır. Bu çalışma için ise sadece yaylı enstruman örnekleri yani, Kontrbas, Viyolonsel, Viyola, ve Keman seçilmiştir.

### 3.2. Entropi Değerlerinin Hesaplanması

Örneklerin zaman ortamdaki entropileri sele sayısının zaman örneği sayısına eşit olarak seçildiği genlik histogramları kullanılarak hesaplanmıştır. Denklem (6) ile de izgesel entropi, Hanning penceresinin kullandığı hızlı Fourier dönüşümü ile elde edilen izge üzerinden bulunmuştur. Şekil 1 Keman  $G5$  notası örneğini, 4096 örnek uzunluğundaki çerçeve ve %87.5'lik üstüste binme oranı ile hesaplanmış zaman ve izge entropi değerlerini göstermektedir.

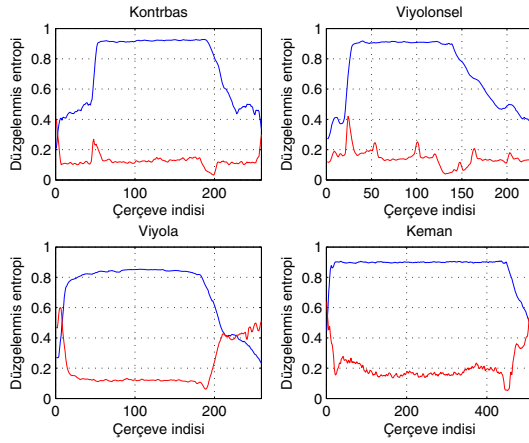


Şekil 1: Keman  $G5$  notası zaman işareti (üstte), zaman entropi değeri (ortada) ve izge entropi değeri (altta).

Şekil 1'den de görüldüğü üzere aynı örnek sayısına sahip çerçeveler kullanılmasına rağmen işaretin zaman entropisi izge entropisinden daha yüksektir. Bir çerçeve süresince yavaşça

değişen genliğe sahip bir işaretin genlik histogramının se-  
leleri benzer kere ortaya çıkan örnekleri içermektedir. Böylece  
olasılık kütle işlevi düzgün dağılıma benzemek ve yüksek ent-  
ropi değerleri elde edilmektedir. Diğer taraftan, bütün frekans  
bileşenlerinin enerjilerinin var olduğu beyaz gürültü tipinde bir  
işaret bulunmadığı sürece, sadece belirli frekanslardaki enerji  
seleleri izge entropisinde ifade edilecektir. Keman örneğinde  
olduğu gibi müzik enstrumanlarında da enerji çoğunlukla temel  
titreşim frekansı ve ona bağlı olarak az sayıdaki harmonik  
frekanslarda bulunmaktadır. Bu durumda da entropi değeri daha  
düşük olmaktadır.

Şekil 2 dört enstrumanın nota örnekleri için düzelenmiş  
zaman ve izgesel entropi değerlerini göstermektedir. Şekil 1'dekine uygun şekilde, düzelenmiş zaman ent-  
ropisinin düzelenmiş izgesel entropiden daha büyük değerler  
aldığı görülmektedir. Zaman entropisinin ortalama değerinin,  
izgesel entropideki ise ortalama değeri etrafındaki değişimlerin  
öznitelik olarak kullanılması durumunda enstrumanlar arasında  
ayırıştırma yapılmasının mümkün olabileceği belirlenmiştir.



Şekil 2: Yaylı enstruman örnekleri için düzelenmiş zaman ve izgesel entropi değerleri.

Müzik enstrumanlarının sınıflandırılması için bir çok  
öznitelik bulunmasına rağmen [3, 6, 7], bu çalışmada öncelikle  
entropi tabanlı öznitelikler ile karşılaştırılabilir olan zaman ve  
izge ağırlık merkezi, izge düzlüğü ve izge yayılımı seçilmiştir.  
Zaman ağırlık merkezi (ZAM),  $x(t)$  ile verilen bir işaretin en-  
erji zarfının ortalaması olarak tanımlanmaktadır.

$$ZAM = \frac{\sum_t x^2(t).t}{\sum_t x^2(t)} \quad (8)$$

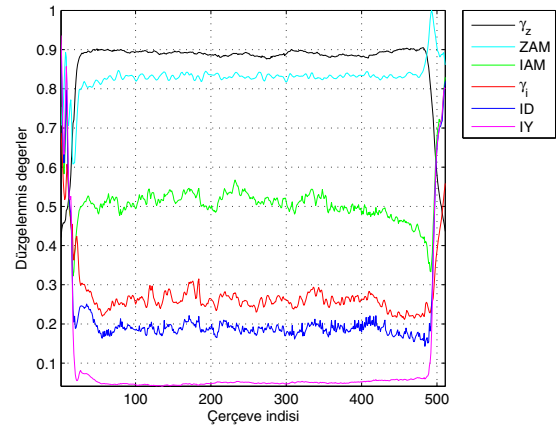
Benzer şekilde izgesel ağırlık merkezi (IAM) de  $f_k$  ilgili  
frekans selesi olmak üzere  $P(f_k)$  ile gösterilen işaretin genlik  
izgesinin ağırlık merkezini tanımlamaktadır.

$$IAM = \frac{\sum_{k=1}^N P(f_k).f_k}{\sum_{k=1}^N P(f_k)} \quad (9)$$

İzgesel ağırlık merkezi, algılanan parlaklık olarak da bilin-  
mektedir ve entropi düzelenmesinde benzer şekilde izge,  
düzelenmiş genliklerin olma sıklıklarının gözlemlendiği dağılım  
olarak da düşünülebilir [2]. İzgesel yayılım (İY) izgenin, izge-  
sel ağırlık merkezi değeri etrafındaki yayılımını ölçer ve izge-  
sel değışıntiye karşılık gelir. İzgesel düzlük (İD) ise izgenin ne  
kadar gürültülü olduğunu ölçmek için iyi bilinen bir yöntemdir  
[20]. İzgenin geometrik ortalamasının aritmetik ortalamasına  
oranı olarak verilir.

$$ID = \frac{\left( \prod_{k=1}^N P(f_k) \right)^{1/N}}{\frac{1}{N} \sum_{k=1}^N P(f_k)} \quad (10)$$

Şekil 3'te bu özniteliklerin zaman ve izgesel entropi  
değerleri ile birlikte, aynı müzik enstrumanı örneği üzerindeki  
değişimleri verilmiştir. Burada  $\gamma_z$  düzelenmiş zaman ent-  
ropisini,  $\gamma_i$  ise düzelenmiş izgesel entropiyi göstermekte-  
dir. Diğer öznitelikler de işaret boyunca aldıkları en yüksek  
değerlerine oranlanarak düzelenmişlerdir. Etiketler ise en  
büyük düzelenmiş değerden en küçüğe doğru sıralanmıştır.  
Benzer değerler alan öznitelikler olarak zaman entropisi ile  
zaman ağırlık merkezi, izgesel entropi ile de izgesel düzlük  
görölmektedir.



Şekil 3: Özniteliklerin karşılaştırılması.

### 3.3. Sınıflandırma Başarımı

Entropi değerlerinin ayrıştırabilme yeteneğini ölçmek için  
her enstrumandan 200 adet ayrıklaştırılmış nota örneği  
seçilmiştir. Bu örneklerin yarısı eğitim kalan yarısı ise  
test amaçlı ayrılmıştır. Herbir örnek için verilen öznite-  
liklerin ortalama ve değışınti değerlerini içeren vektörler  
oluşturulmuştur. Sınıflandırmalar, k-en yakın komşu (k-EYK),  
çok-katmanlı algılayıcı (ÇKA), ve destek vektör makine  
(DVM) sınıflandırıcıları kullanılarak gerçekleştirilmiştir. k-  
EYK sınıflandırıcıda tek bir komşu, ÇKA sınıflandırıcıda tek  
saklı katmanda 20 adet nöron, DVM sınıflandırıcıda radyal  
taban işlevi birim çekirdek değışınti değışkeni ile birlikte  
kullanılmıştır. Sonuçlar k-EYK ve DVM için on kat çapraz

onaylama ile, ÇKA için ise 100 gerçeklemenin ortalaması alınarak elde edilmiştir. Her bir enstrumanın ne kadar başarıyla sınıflandırıldığını incelemek yerine özneliklerin toplamda elde ettikleri ortalama sınıflandırma başarımları ölçü olarak alınmıştır. Tablo 1 elde edilen başarımların yüzdelerini göstermektedir.

Tablo 1: Ortalama sınıflandırma başarımları.

%	k-EYK	ÇKA	DVM
$\gamma_i$	74.28	58.25	81.55
$\gamma_t$	64.72	55.54	72.20
İAM	66.80	62.20	75.18
ID	59.52	54.48	68.15
IY	62.90	57.33	70.83
ZAM	49.25	48.20	57.05

Sonuçlar, zaman entropi değerlerinin benzer değişim gösteren zaman ağırlık merkezinden, izgesel entropi değerlerinin benzer değişim gösteren izgesel düzlükten daha iyi, yaygın olarak kullanılan izgesel ağırlık merkezi kadar da iyi başarımlar elde edildiğini göstermiştir.

#### 4. DEĞERLENDİRME

Bu çalışmada, entropi değerlerinin işaretlerdeki değişimi izlemedeki başarısı gözönüne alınarak düzlenen entropi değerlerinin öznelik olarak kullanılması ve müzik enstrumanlarının sınıflandırılmasındaki başarımların incelenmiştir. Elde edilen sonuçlar, ele alınan yaylı enstruman örneklerinin sınıflandırılması için düzlenen entropi değerlerinin benzer değişimler gösteren öznelikler ile benzer başarımlara sahip olduğunu göstermiştir.

Sınıflandırıcıların en iyi parametre değerlerinde kullanılmaması başarımların değerlerini düşürmektedir. Ancak bütün örnekler için aynı sınıflandırıcı yapılarının kullanılması karşılaştırma sonuçlarındaki değerlerden çok özneliklerin birbirlerine göre başarımların ilişkisini incelemek için yeterli olmaktadır. Bundan sonraki çalışmalarda diğer enstruman seslerini de içerecek örneklerin kullanılmasının yanısıra, farklı çerçeve boyutlarının etkisinin incelenmesi düşünülmektedir.

#### 5. KAYNAKÇA

- [1] Herrera-Boyer, P., Peeters, G., and Dubnov, S., "Automatic classification of musical instrument sounds," *Journal of New Music Research*, 32(1):3–21, 2003.
- [2] Peeters, G., "A large set of audio features for sound description (similarity and classification) in the CUIDADO project," *IRCAM, Tech. Rep.*, 2004.
- [3] Essid, S., Richard, G., and David, B., "Instrument recognition in polyphonic music based on automatic taxonomies," *IEEE Trans. on Audio, Speech and Language Processing*, 14(1):68–80, 2006.
- [4] Klapuri, A. and Davy, M., Eds., *Signal Processing Methods for Music Transcription*, New York, Springer, 2006.
- [5] de Cheveigné, A. and Kawahara, H., "YIN, a fundamental frequency estimator for speech and music," *Journal of Acoustical Society of America*, 111(4):1917–1930, 2002.
- [6] Deng, J. D., Simmermacher, C., and Cranefield, S., "A study on feature analysis for musical instrument classification," *IEEE Trans. on Systems, Man, and Cybernetics - Part B: Cybernetics*, 38(2):429–438, 2008.
- [7] Wiczorkowska A. A. and Kubik-Komar A., "Application of analysis of variance and post hoc comparisons to studying the discriminative power of sound parameters in distinguishing between musical instruments", *Journal of Intelligent Information Systems*, doi:10.1007/s10844-010-0140-5, 2010.
- [8] Peeters, G., McAdams, S., and Herrera, P., "Instrument sound description in the context of MPEG-7," *Proc. of the International Computer Music Conference*, Berlin, Germany, 2000.
- [9] ISO/IEC Working Group, "MPEG-7 overview," <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>, 2004.
- [10] Cover, T. M. and Thomas, J. A., *Elements of Information Theory*, 2nd ed., New Jersey, John Wiley & Sons, 2006.
- [11] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd ed., New York, McGraw-Hill, 1991.
- [12] Bishop, C. M., *Pattern Recognition and Machine Learning*, New York, Springer, 2006.
- [13] Ajmera, J. and Bourlard, I. M. H., "Speech/music segmentation using entropy and dynamism features in a HMM classification framework," *Speech Communication*, vol. 40, pp. 351–363, 2003.
- [14] Misra, H., Ikbali, S., Bourlard, H., and Hermansky, H., "Spectral entropy based feature for robust ASR," *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, Montreal, Quebec, Canada, pp. 193–196, 2004.
- [15] Pikrakis, A., Giannakopoulos, T., and Theodoridis, S., "A computationally efficient speech/music discriminator for radio recordings," *Proc. of the International Conference on Music Information Retrieval and Related Activities*, Victoria, BC, Canada, 2006.
- [16] Ibarrola, A. C. and Chávez, E., "A robust entropy-based audio-fingerprint," *Proc. of the IEEE International Conference on Multimedia and Expo*, Toronto, Ontario, Canada, pp. 1729–1731, 2006.
- [17] Davies, M. E. P. and Plumbley, M. D., "On the use of entropy for beat tracking evaluation," *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, Honolulu, Hawaii, USA, pp. 1305–1308, 2007.
- [18] Çataltepe, Z. and Altinel, B., "Hybrid music recommendation based on different dimensions of audio content and an entropy measure," *Proc. of the European Signal Processing Conference*, Poznań, Poland, pp. 936–940, 2007.
- [19] Fritts, L., "The University of Iowa Electronic Music Studios Musical Instrument Samples," <http://theremin.music.uiowa.edu>, 1997.
- [20] Dubnov, S., "Spectral anticipations," *Computer Music Journal*, 30(2):63–83, 2006.