

# PARALEL EVRİMSEL ENİYİLEME İLE SAYISAL SES SENTEZLEME DIGITAL SOUND SYNTHESIS VIA PARALLEL EVOLUTIONARY OPTIMIZATION

Kamer Ali Yuksel<sup>1</sup>, Aytul Ercil<sup>3</sup>

Yapay Görme ve Örüntü Analiz Laboratuvarı  
Sabancı Üniversitesi  
{kamer, aytulercil}@sabanciuniv.edu

Batuhan Bozkurt<sup>2</sup>

Müzik İleri Araştırmalar Merkezi  
İstanbul Teknik Üniversitesi  
batuhan@batuhanbozkurt.com

## ÖZETÇE

*Bu araştırmada çeşitli ses sentezleme parametrelerinin eniyilemesi için yeni bir paralelleştirilebilir mimari öneriyoruz. Bu mimari farklı ses sentezleme topolojilerinin parametrelerini hedef seslerle eşlemek için genetik algoritmalar kullanmaktadır. Uygunluk fonksiyonu yakınsama süresini azaltmak için paralel olarak hesaplanmaktadır. SuperCollider ses sentezleme ve programlama ortamını kullanarak önerilen mimariye dayalı bir çerçeve gerçekleştirdik ve çok sayıda deney yürüttük. Deneylerin sonuçları çerçevenin ses sentezleme parametrelerinin ümit verici hızlarda ve doğru kestirimi için kullanılabilirliğini göstermiştir.*

## ABSTRACT

*In this research, we propose a novel parallelizable architecture for the optimization of various sound synthesis parameters. The architecture employs genetic algorithms to match the parameters of different sound synthesizer topologies to target sounds. The fitness function is evaluated in parallel to decrease its convergence time. Based on the proposed architecture, we have implemented a framework using the SuperCollider audio synthesis and programming environment and conducted several experiments. The results of the experiments have shown that the framework can be utilized for accurate estimation of the sound synthesis parameters at promising speeds.*

## 1. GİRİŞ

Ses analizi için her girişim aynı zamanda bir tür parametre kestirimi çabasıdır. Analiz işi farklı ses sentezleme yöntemleriyle yeniden sentezlenecek bir kaynak sesin özelliklerini elde etmek için veya bu özellikleri kuramsal bir modele oturtmak amacıyla incelemek için kullanılabilir. Örneğin Roads, Fourier Analizinin bu tip bir analizin döndürdüğü sonuçların (yani analiz edilen işaretin büyüklük ve evre açıklamalarının) kaynağın içeriğine gerçeğe uygun olarak yaklaşacak, sinüs dalgalarıyla bir yeniden sentezleme için parametreler olarak kabul edilebilmesi nedeniyle bir parametre kestirim yöntemi olarak düşünülebileceğine işaret eder [1]. Ancak biz parametre kestirimi problemine elde edilen seslerin çeşitli algısal özelliklerini daha üst düzeyde bir soyutlama bakımından denetleyebilmek ve değiştirebilmek için verilen bir sese değişik sentezleme yöntemleriyle yaklaşmak için gereken veri miktarını etkin bir şekilde azaltmak için sezgisel ve etkileşimli bir biçimde yaklaşıyoruz.

Bu makalede ses sentezleyicilerin parametrelerini eniyilemek

için paralelleştirilebilir evrimsel bir mimarinin kullanımını öneriyoruz. Bu mimari ses sentezi ve algoritmik birleşime özel SuperCollider (SC) programlama dili ile oluşturulan bir modüler evrimsel çerçeve olarak gerçekleştirilmiştir [2]. Çerçeve kullanıcı tarafından yaratılan gelişigüzel ses sentezleyici topolojisini kullanarak verilen her tür hedef sese yaklaşmak için gereken parametre kümesini otomatik olarak eniyilemek için genetik algoritmalar (GA) kullanmaktadır. Çerçevenin verimliliğini sınamak amacıyla çeşitli hedef sesler için bir perküsyon ve çok kipleycili frekans kiplenimi sentezleyici kullanarak deneyler yaptık. Son olarak, evrimsel yöntemlerin yüksek derecede birleştirici gerçek zamanlı ses sentezi ve algoritmik birleşim ortamında etkileşimli hızlarda çalışan yaratıcı kullanımlarına fırsat tanıyacak fikirler tanımladık. Bu çalışmanın en önemli katkısı SC ortamını kullanarak gerçekleştirilen paralel mimariyle ve algısal kaliteyi koruyan basitleştirilmiş uygunluk fonksiyonuyla elde edilen ümit verici yakınsama süresidir.

Bu araştırma kısmen tarafından 109E134 numaralı TUBITAK VIPSAFE projesi tarafından desteklenmiştir.

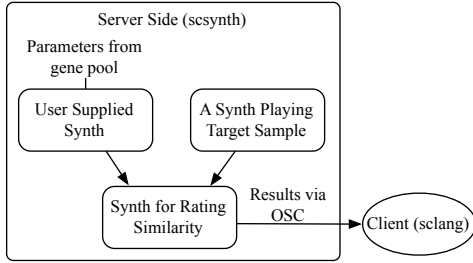
## 2. BENZER ÇALIŞMALAR

Evrimsel metodolojiler daha önce ton eşleme ve çeşitli farklı sentezleyici topolojileri için parametre kestirim problemini çözmek için araştırılmıştır. Manzollı ve arkadaşları dalga biçimlerinin hedef seslerin karşılık gelen psikoakustik özelliklerine evrimini ortaya atmıştır [3]. Garcia [4] çevrimsiz ağaç çizgeleri biçiminde gösterilen ses sentezleyici algoritmalarının tasarımını otomatikleştirmek için genetik programlama yaklaşımını kullanmıştır. Johnson [5] parametre uzayını aramak ve Csound FOF sentezleyicinin parametrelerini yönetmek için kullanıcı tarafından yürütülen etkileşimli bir yaklaşımın araştırıldığı etkileşimli genetik algoritmalarından (EGA) yararlanmıştır. Bu alandaki çalışmaların büyük çoğunluğu önerilen yöntemlerin bahsedilen iş için uygunluğunu ve verimliliğini ispat etmeyi hedeflemektedir. Bu nedenle bu çalışmalarda kullanılan yazılım araçları araştırmanın odaklandığı önermenin özelliklerini göstermeye yöneliktir. Sonuç olarak, diğer sayısal müzik yapma yazılımlarıyla bağlantıları eksiktir; dolayısıyla bunları beste yapma, canlı performans ve diğer yaratıcı amaçlar için pratikte kullanmak zordur.

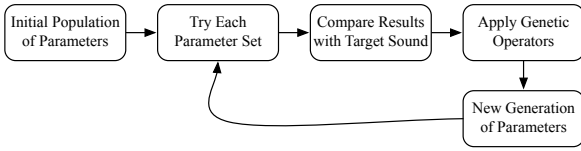
## 3. MİMARİ

Önerilen mimari ses sentezi için bir ana ortamdaki, bir modüler evrimsel çerçeveden (EVWorkBench) ve bir parametre kestirim sisteminden (EVMATCH) oluşmaktadır. Ana ortam ses sentez ve programlama dili SC (sclang olarak

anılacaktır) kullanılarak gerçekleştirilmiştir [2]. Ses sentezi istekleri sunucu tarafında (scsynth) SC ortamı altında derlenmiş birim üretici çizge fonksiyonlar yoluyla ele alınmaktadır (Şekil 1). Bu mimari tek bir slang istemcisinin birden çok yerel ve ağa bağlı scsynth sunucuları örneklerini Açık Ses Denetimi (ASD) protokolü aracılığıyla denetmesini sağlar [6].



Şekil 1: ASD protokolü kullanarak Sunucu tarafı Uygunluk Hesaplaması.



Şekil 2: Genetik Algoritma Döngüsü.

Parametrelerin ilklendirilmesi ve genetik işleçlerin uygulanması sarılmış EVWorkBench sınıfı tarafından üstlenilmiştir. Bu yolla eniyileme iş akışının (ilklendirme, seçme, geçit ve değişim gibi) aşamaları çeşitli evrimsel yaklaşımların kolaylıkla denenebileceği modülerleştirilmiş bir çerçeve oluşturmak için birbirinden ayrılmıştır. Ayrıca (artan miktarda zorlaşan uygunluk fonksiyonlarına sahip olan) katmanlı öğrenme [7] ve değişim işleçlerinin sınırlarının evrim süresince ince ayarının yapılması ana ortamın yorumlanmış ve etkileşimli dili tarafından desteklenmektedir. EVMatch sınıfı kullanıcı tarafından sağlanan sentez tanımının derlenmesinden (SynthDef olarak anılacaktır), derlenen SynthDef'in sağlanan (yerel ve ağa bağlı) sunuculara iletiminden, gen havuzundaki parametrelerin hesaplanması ve eniyilemesinden ve hesaplamının bilişimsel yükünün örneğe kayıtlı sunuculara dağıtılmasından sorumludur. Bu çalışmada eniyileme bir hedef sesin kullanıcı tarafından sağlanan duruk sentezleyici topolojilerinin parametre uzayında yapılmaktadır. Ancak sentezleyici topolojilerini kromozomlar olarak kodlayıp ses sentez algoritmalarını [4]'te tarif edildiği gibi evrimleştirmek de mümkündür. Eniyilemenin hızı farklı arama bölgeleri ve istenen zaman kısıtları için ayarlanabilen seçim baskısını (ör. turnuva büyüklüğü) doğrudan denetleme yoluyla aramanın yakınsama süresini etkileyerek artırılabilir.

### 3.1. GA Eniyileme

Eniyilemeye başlamak için kullanıcı ilk popülasyonu oluşturacak olan istenen özelliklere sahip bir hedef ses dosyası, bir sentezleyici tanımı, parametre adları ve parametrelerin varsayılan değerlerini verir. Bu ilklendirmeden sonra GA döngüsü (Şekil 2) nesilden nesile ilerleyerek daha iyi parametre kümesi çözümlerine ulaşır. GA her yinelemede

her parametre kümesi için sesleri sentezler ve çıkan sesin özelliklerini hedef sesle uygunluğunu hesaplayarak karşılaştırır. Sonra, birden çok birey yürürlükteki popülasyondan yeni bir nesil üretmeye uygunluklarına dayanarak rastgele seçilir. Seçimden sonra geçit ve değişim işleçleri gen havuzuna uygulanır. Bu döngü kullanıcı tarafından belirlenen popülasyon için yeterli uygunluk düzeyine erişilene kadar sürdürülür. Hedef uygunluk değerine ek olarak, kullanıcı dinleyerek son neslin en güçlü bireyinin uygunluğuna karar verebilir veya yinelenen en çok nesil sayısını sınırlayabilir. Turnuva seçimi [8] GA'nın yakınsama hızıyla seçim baskısı arasındaki ilişkiyle ve turnuva seçiminin gürtütlü GA zincirlerine uygunluğu ile ilgili olarak tercih edilmektedir. Turnuva seçimi popülasyondan rastgele seçilen birkaç birey arasında çok sayıda karşılaşma yapılmasını içerir. Her karşılaşmadan sonra en fazla uygunluk gösteren (kazanan) geçit için seçilir. Ardından, bölünme noktası sayısının geçit olasılığıyla orantılı olarak belirlendiği çok noktalı bir geçit işleci uygulanır.

## 4. UYGUNLUK HESAPLAMASI

Uygunluk hesaplaması gen havuzunun her üyesi için seçim aşamasını etkileyecek ve böylece evrimsel sürece yön verecek uygunluk sayıları verir. Parametre kestirimi için kullandığımız uygunluk fonksiyonu sentezleyici tarafından çıktı olarak verilen sesin özelliklerini hedef sesin özellikleriyle gen havuzundaki her parametre kümesiyle yürüterek karşılaştırır. Uygunluk hesaplaması iki ses arasındaki benzerliğin kapsamını açığa çıkaran uygunluk değerlemesiyle sonuçlanır. Uygunluk değerlemesi için hesapladığımız kompleks spektrumlar arasındaki analitik spektral uzaklıkların büyüklükleri kaynak (sentezlenmiş) ve hedef sesler arasındaki benzerliğin bir ölçüsü olarak kullanılır. Bizim gerçekleştirmemizde evre bilgisini dikkate almadan sadece büyüklüklere odaklanarak [4] tarafından önerilen analitik uzaklık metriklerini kullandık. Uygunluk fonksiyonunun sadeleştirilmesi sentezlenen sesin sentezleyici topolojisine dayanmakta olan algısal kalitesini korumakla birlikte yakınsama süresini önemli ölçüde azaltır. Denklem 1 sentezlenen ve hedef seslerin büyüklük spektrogramları arasındaki ortalama hata kare (MSE)'yi hesaplar. Ağırlık matrisi  $W_{mj}$  daha fazla enerjisi olan spektral bölgelerdeki hatanın azaltılmasına yardımcı eder. Ağırlık matrisinin MSE üzerindeki etkisi  $O$  ile ayarlanabilir.

$$MSEMag = \frac{1}{Frames} \sum_{j=1}^{Frames} \sum_{i=1}^{Bins} [ (|X_{ij}| - |T_{ij}|)^2 W_{mj} ]$$

$$W_{mj} = O + (1 - O) \frac{\log |T_{ij}| - \min(\log |T_{ij}|)}{|\min(\log |T_{ij}|) - \max(\log |T_{ij}|)|}$$

Denklem 1: sentezlenen ve hedef seslerin büyüklük spektrogramları arasındaki ortalama hata kare (MSE).

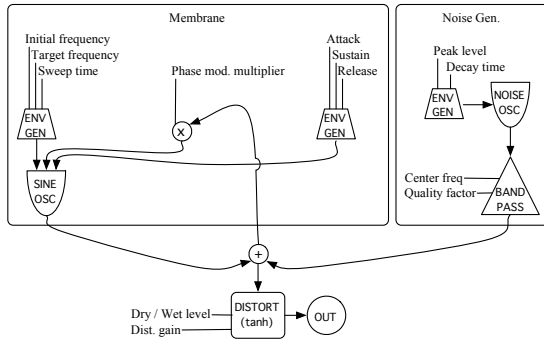
## 5. PARALELLEŞTİRME

Uygunluk fonksiyonu çalışmak için tek bir bireyin kromozomuna ihtiyaç duyduğu için bir nesildeki gen havuzunun her bir üyesi için uygunluk sayılarını belirlemek için yapılacak hesaplamalar aynı anda çalıştırılabileceğinden uygunluk hesaplamaları paralelleştirilebilir. Bizim

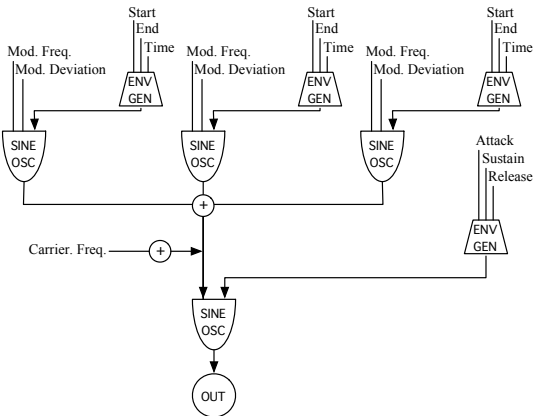
çalışmamızda tüm popülasyonun hesaplamaları parametre kestirim örneğine kayıtlı sunucular arasında bölünmüştür. Böylelikle iş yükü tek bir sistemde bulunan mantıksal işlemciler ve bir ağ üzerindeki bilgisayarlar arasında bağımsız işleme için verimli bir şekilde bölüştürülmüştür. Dolayısıyla kayıtlı sunucular arasında eşzamanlamayı ve yükün dağıtılmasını otomatik olarak yapar. Parametre eşlemenin paralelleştirilmesi kullanıcıların sesle sayısal müzik yapmaya uygun bir şekilde etkileşimli bir hızda çalışmasına olanak tanır. Bu nedenle SC'nin gerçek zamanlı ses sentezleme ve analizi gibi bilişimsel işlerin tek bir makinenin MİB çekirdekleri ve ağa bağlı birden çok bilgisayar arasında paralelleştirilmesine olanak veren (scsynth olarak anılan) istemci-sunucu mimarisini kullandık.

## 6. DENEYLER

Çerçevenin verimliliğini belirlemek için bir perküsyon sentezleyici ve çoklu kipleycili frekans kiplenimi topografisi içeren tipik kullanım senaryoları için çeşitli deneyler yürüttük. Tüm deneyler taşınabilir bir "Intel Core 2 Duo 2.4 Ghz (Penryn)" işlemciyle her iki çekirdeğini de kullanarak yapıldı.



Şekil 3: Elektronik davul ve perküsyon sesleri yaratan PS'nin topolojisi.

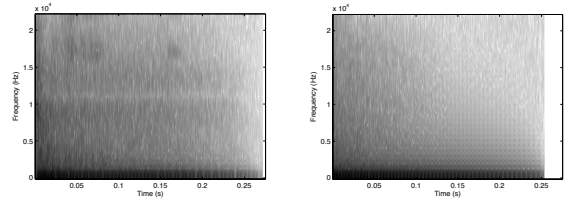


Şekil 4: 3 kipleycili paralel ÇK-FK sentezleyici topolojisi.

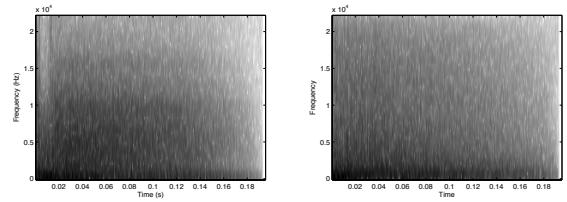
### 6.1. Perküsyon Sentezleyici (PS)

Elektronik davul ve perküsyon sesleri yaratmak için basit bir zar olarak davranan bir frekans zarfı içeren bir sinüs salıncıdan ve zil sesleri veya atak gürültülerinde yardımcı olan bir süzölmüş gürültü üretici zarfından oluşan bir

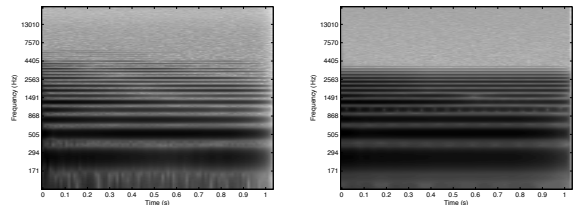
perküsyon sentezleyici (PS) (Şekil 3) gerçekleştirdik. Her iki zarfın toplamı sinüs salıncının evre girişine geri beslenir (EK geri besleme) ve çıkışa bir bozulma aşamasından (tanh bozulma) sonra ulaşır. PS'nin hemen her aşaması (zarf bağımsız değişkenleri, EK miktarı, bozulma miktarı, süzgeç ayarlarını içermek üzere) parametrelendirilmiştir. PS çeşitli elektronik davul ve perküsyon sesleri üretme yeteneğinde olduğu halde parametre arama uzayı ilgili sesler (diğer davul sentezleyiciler tarafından yaratılmış perküsyon sesleri gibi) için çok karmaşık değildir. Ancak akustik kaynaklı sesler daha karmaşık spektrumlara sahiptir; dolayısıyla PS bunların ayırt edici biçimde karmaşık olan spektrumlarını mükemmel bir şekilde üretmez. Yine de PS'nin bizim mimarimizi kullanarak sesler arasında türlerarası çeviriler yapabildiğini gözlemledik. Örneğin PS tipik olarak bir davul sentezleyiciye gerçek zil kayıtları verilip bu ses elektronik haline dönüştürülerek programlanan karakteristik zil seslerine benzer seslere yakınsayabilmektedir. GA'nın uygun bir çözüme yaklaşık olarak 15 saniyede yakınsadığı iki hedef (akustik tom ve trampet örnekleri) ve sentezlenmiş ses çiftinin spektrogram görüntüleri Şekil 5 ve 6'da verilmiştir.



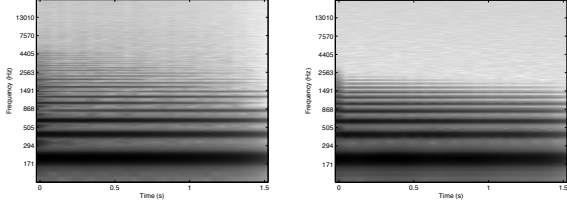
Şekil 5: Akustik bir tom davul örneği için doğrusal frekans ölçek spektrogramı ve perküsyon sentezleyici tarafından eşlenen parametrenin çıktısı.



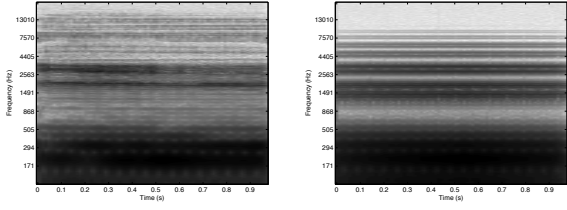
Şekil 6: Bir akustik trampet örneği için doğrusal frekans ölçek spektrogramı çıktısı ve perküsyon sentezleyici tarafından eşlenen parametrenin çıktısı.



Şekil 7: Kaydedilmiş bir piyano örneği için logaritmik frekans ölçekli spektrogram çıktısı ve ÇK-FK sentezleyicinin parametre eşlenmiş sentezleme çıktısı.



Şekil 8: Kaydedilmiş bir rhodes klavye örneği için logaritmik frekans ölçekli spektrogram çıktısı ve ÇK-FK sentezleyicinin parametre eşlenmiş sentezleme çıktısı.



Şekil 9: "ee" sesli harfini çıkaran kaydedilmiş bir insan sesi örneği için logaritmik frekans ölçekli spektrogram çıktısı ve ÇK-FK sentezleyicinin parametre eşlenmiş sentezleme çıktısı.

## 6.2. Çok Kipleyicili Frekans Kiplenimi (Paralel ÇK-FK)

Şekil 4'teki 3 sinüs salınıcısının tek bir taşıyıcı salınıcılı kiplendiği bir paralel ÇK-FK topografisi çerçeveyi daha karmaşık arama uzaylarında sınamak için gerçekleştirilmiştir. Taban frekansları, kipleme miktarları ve indisleri parametrelendirilmiş ve tasarımı taşıyıcı ve kiplenici frekansları arasında tam kat ilişkisi gözetilmemiştir. ÇK-FK sentezleyici çeşitli karmaşık harmonik olan ve olmayan spektrumlar üretme yeteneğine sahiptir. Çeşitli harmonik hedef sesler değişik GA ayarlarıyla denenmiştir; ancak ortalama yakınsama süresi arama uzayının göreceli daha karmaşık olması nedeniyle PS'ye göre oldukça yüksektir (3 dakika civarı). T:K oranı tam kat ilişkisinde (frekans zarfları olmaksızın) kilitlendiğinde yakınsama süresi büyük ölçüde azalmaktadır; ancak üretilen sesler çok daha az ilginç olmaktadır. Çeşitli hedef (piyano, rhodes klavye ve insan sesi örnekleri için) spektrogram görüntüleri ve sentezlenmiş ses çiftleri sırasıyla Şekil 7, 8 ve 9'da verilmiştir.

## 7. TARTIŞMA

Gerçekleştirmenin genel amaçlı arama eniyilemesi işine uygun olması amaçlandı; dolayısıyla parametre kestirim sistemi belli bir sentezleme yönteminin veya sentezleyici topolojisinin işlevsel olmasına dayanmaz. Böylece arama uzayının karmaşıklığı sentezleyicinin hedef sesle bağlantılı olarak tanımlanan parametre aralıklarından doğrudan etkilenir. GA arama verilen bir topoloji için anlaşılmayan parametreleri eleyeceği için olası parametre aralıkları vermeye özellikle dikkat etmek çok gerekli değildir. Ancak algoritmayı birbiriyle ilintili olma olasılığı bulunan parametre aralıkları kümesine yönlendirmek aramanın karmaşıklığını büyük ölçüde azaltacaktır; dolayısıyla sabit bir zaman

penceresinde daha iyi sonuçlar elde edilebilir. Böyle bir kurulumu kullanarak bir enstrümanla üretilen hemen her ses için birkaç iyi ses örneği girildiğinde iyi sonuçlar verecek bir parametre ayarı hesap etmek olası olabilir. Ama pratikte rastgele sağlanan bir yöntemin bir hedef sese tatmin edici bir şekilde yaklaşmasının garantisi yoktur [1] ve genellikle sadece GA olurlu çözümleri göstermeyen parametre kümelerini eleyemeyebilir. Neyse ki SC dağıtımında sağlanan birim üreteçleri sayesinde küçük veya büyük ölçekli sentezleyici topolojileriyle deney yapmak nispeten kolaydır. Dolayısıyla birim üreteç grafik fonksiyonunu tanımladıktan hemen sonra ses üretimi için sentezleyiciler hazırdır. Süreç o kadar gübüzdür ki canlı programlama ve sentezleyicilerin denetimi çeşitli bilgisayar müzisyenleri tarafından canlı kodlama performanslarında gerçek zamanlı olarak yapılabilir [9]. Dinamik, etkileşimli ve neredeyse doğaçlama sentezleyici yaratımındaki sadelik, seslerin türler arası dönüşümü ile kolay deneyler yürütme ve görünüşte uygunsuz sentezleyiciler ve hedefler arasında parametre kestirimine deneysel yaklaşımlar gibi yaratıcı amaçlar için ilginç fırsatlara yol açmaktadır.

## 8. SONUÇLAR

Bu çalışmada sayısal ses sentezleyiciler için esnek parametre kestirimi yapmak için SuperCollider (SC) yazılım platformuna tümleşik genel amaçlı evrimsel bir çerçeve önerdik. Sistem parametre kestirimi işleri için canlı performans ve besteleme iş akışlarının tipik olarak gerektirdiği etkileşimli hızlarda çözümler sağlayabilmektedir. Çerçevenin modüler ve esnek yapısı müzisyenler ve bu alandaki araştırmacılar için geniş olasılıklara yol açabilir.

## 9. KAYNAKÇA

- [1] C. Roads and J. Strawn, The Computer Music Tutorial, 4th ed. Cambridge, MIT Press, 1999.
- [2] J. McCartney, "SuperCollider, a New Real Time Synthesis Language," Proceedings of the International Computer Music Conference, pp. 257–258, 1996.
- [3] J. Manzolli, A. Maia, Jr., J. Fornari, and F. Damiani, "The Evolutionary Sound Synthesis Method," in MULTIMEDIA '01: Proceedings of The Ninth ACM International Conference on Multimedia, New York, NY, USA, 2001, pp. 585–587.
- [4] R. Garcia, "Growing Sound Synthesizers Using Evolutionary Methods," European Conference in Artificial Life ECAL2001. Artificial Life Models for Musical Applications. University of Economics, Prague, Czech Republic., 2001.
- [5] C. Johnson, "Exploring Sound-Space with Interactive Genetic Algorithms," Leonardo, vol. 36, no. 1, 2003.
- [6] M. Wright, "Open Sound Control: An Enabling Technology for Musical Networking," Organised Sound, vol. 10, no. 3, pp. 193–200, 2005.
- [7] J. McDermott, M. O'Neill, and N. J. L. Griffith, "Target-driven Genetic Algorithms for Synthesizer Control," 9th Int. Conference on Digital Audio Effects, 2006.
- [8] B. L. Miller and D. E. Goldberg, "Genetic Algorithms, Tournament Selection, and the Effects of Noise," Complex systems., vol. 9, no. 3, 1995.
- [9] N. Collins, A. McLean, J. Rohrerhuber, and A. Ward, "Live Coding in Laptop Performance," Organised Sound, vol. 8, no. 3, pp. 321–330, 2003.