

آشکارسازی آغازه در تکنوازی تار بر پایه انرژی و فرکانس پایه

بهرز فرخی و احسان‌اله کبیر

انرژی در هر باند فرکانسی از تفاضل لگاریتمی بین دو نمونه زمانی متوالی استفاده می‌شود. بر اساس آنچه کلاپوری گزارش کرده است، مزیت استفاده از تفاضل لگاریتمی بر تفاضل معمولی دقت بالایی آن در شناسایی زمان آغازها است. در [۵] تغییرات انرژی سیگنال با استفاده از تبدیل بسته ویولت^۱ محاسبه شده و در نهایت با استفاده از شبکه عصبی بازخورددار، محل آغازها تعیین می‌شود.

کارایی اغلب روش‌های انرژی پایه با نرم‌تر شدن آغازها کاهش می‌یابد. برای رفع این مشکل، بلو در [۶] روشی بر مبنای فاز پیشنهاد کرد. این روش که حساسیت بالایی در برابر نویز دارد در مواردی که سیگنال تمیز نباشد در شناسایی آغازها دچار مشکل می‌شود. برای حل این مشکل در کارهای بعدی از جمله [۷] و [۸] ترکیبی از ویژگی‌های انرژی و فاز برای آشکارسازی آغازها پیشنهاد شد. این روش‌ها ضمن برخورداری از مزیت روش فازپایه، حساسیت کمتری در برابر نویز دارند.

در اکثر سازها، تغییر نت با تغییر فرکانس پایه سیگنال همراه است. این ویژگی اساس کار دسته دیگری از روش‌های آشکارسازی آغازها است [۹] تا [۱۲]. در این روش‌ها با شناسایی ساختار فرکانسی سیگنال، تغییرات آن که احتمالاً متناظر با نواخته شدن یک نت جدید است بررسی می‌شود. از نظر تئوری، روش‌های فرکانس پایه قادر نیستند چند نت هم‌فرکانس و متوالی را تشخیص دهند. با این حال در بسیاری از سازها از جمله تار به دلیل اعوجاجی که برخورد مضراب با سیم در طیف سیگنال ایجاد می‌کند، احتمال تشخیص آغازها، هر چند با دقتی پایین‌تر وجود خواهد داشت [۱۳] و [۱۴]. برای بررسی بیشتر این سه دسته روش و مقایسه کمی و کیفی آنها می‌توانید به [۸]، [۱۵] و [۱۶] مراجعه کنید.

در کارهایی که در سالیان اخیر ارائه شده است از جمله [۱۷] تا [۱۹] با ترکیب ویژگی‌های انرژی، فاز و فرکانس، تلاش برای ارائه روشی بهینه در جریان است. در این روش‌ها سعی شده که با ترکیب ویژگی‌های سه‌گانه فوق، بهترین نتیجه درباره شناسایی نت‌های سخت و نرم به دست آید. در کنار روش‌هایی که از ویژگی‌های سه‌گانه گفته شده استفاده کرده‌اند، روش‌هایی نیز ارائه شده که به عنوان مثال از مدل‌های آماری [۲۰] و نیز ساختار ضرب‌آهنگی قطعه موسیقی [۲۱] کمک گرفته‌اند. تعیین نوع ساز پیش از آشکارسازی آغازها^۲ رویکردی تازه است که در [۲۲] به آن توجه شده است. در این مقاله با استفاده از مدل پنهان مارکوف، نوع ساز مشخص شده و بر اساس آن آغازها تعیین می‌شود. از آنجا که در هر ساز آغازهای سخت یا نرم غالب‌تر است، این روش می‌تواند در آشکارسازی آغازها رویکردی متناسب با نوع آن را در پیش گیرد. با وجود تلاش‌های فراوانی که تا به امروز انجام شده است تا رسیدن به نتایج قابل اعتماد به خصوص در مورد آغازهای نرم در ساز و آواز فاصله زیادی وجود دارد.

هدف از این تحقیق، پیشنهاد روشی تازه برای آشکارسازی آغازها در تکنوازی ساز تار است. موسیقی ایرانی به دلیل برخورداری از پشتوانه

چکیده: در این مقاله روشی تازه برای آشکارسازی آغازها در تکنوازی ساز تار ارائه شده است. در تعریف تابع آشکارساز آغازها، از دو ویژگی انرژی و فرکانس پایه استفاده کرده‌ایم که نتیجه آن به کمک یک الگوریتم قله‌یاب سازگار برای آشکارسازی اولیه آغازها استفاده می‌شود. با استفاده از یک روش تطبیق کلیشه‌ای بهبود یافته، فرکانس‌های پایه تعیین شده و از تطبیق زمانی آنها با نتایج اولیه، آشکارسازی نهایی آغازها انجام می‌شود. تابع آشکارساز پیشنهادی به دلیل استفاده هم‌زمان از دو ویژگی انرژی و فرکانس پایه، قابلیت تشخیص نت‌های نرم و سخت را داراست. برای آزمایش، این سیستم را روی مجموعه‌ای از قطعات تکنوازی تار با آغازهای معلوم به کار گرفته و نتایج آن را در مقایسه با روش مبتنی بر ویژگی انرژی و بر اساس ملاک F تحلیل کرده‌ایم. نتایج به دست آمده نشان‌دهنده بهبود کارایی سیستم در شناسایی آغازهای ساز تار است.

کلیدواژه: آشکارسازی آغازها، پردازش صوت، ساز تار، موسیقی ایرانی، نت.

۱- مقدمه

موسیقی از توالی خردمندانه و خلاقانه نت‌ها شکل می‌گیرد و آغازها به لحظه نواخته شدن نت‌ها گفته می‌شود. با شناسایی سیگنال موسیقی، اغلب با شناسایی نت‌ها و لحظه نواخته شدنشان شروع شده و نتیجه برای مراحل بعد همچون تشخیص ریتم و آوانگاری استفاده می‌شود. به طور کلی آغازها در دو شکل "سخت" و "نرم" قابل مشاهده است. نت سخت با تغییر ناگهانی انرژی در لحظه نواخته شدن نت همراه است در حالی که برای نت نرم این جهش ناگهانی چشم‌گیر نیست. با توجه به این تقسیم‌بندی، سه رویکرد اساسی را در آشکارسازی آغازها می‌توان مشاهده کرد: انرژی پایه، فاز پایه و فرکانس پایه.

مشاهده تغییرات ناگهانی انرژی که در آغازهای سخت مشهودتر است، اساس روش‌های انرژی پایه را تشکیل داده است. در نخستین کارها با مشاهده پوش سیگنال و یافتن جهش‌های ناگهانی با استفاده از یک تابع قله‌یاب، محل آغازها تعیین می‌شد. این روش‌ها جز در سازهای کوبه‌ای، ضعف و عدم دقت فراوان دارد. محققان به تدریج دریافته‌اند که تفکیک پهنای فرکانسی سیگنال به چندین بازه و بررسی یک به یک آنها برای شناسایی آغازها کارا تر است. از نخستین کارها که با الهام از این ایده انجام شد و کارایی خوبی دارد، روش پیشنهادی در [۱] است. کلاپوری در [۲] با الهام از مدل سیستم شنوایی گوش انسان روشی تازه را ارائه کرد که در بسیاری از مقالات بعدی از جمله [۳] و [۴] به آن توجه شد. در این روش برای تحلیل فرکانسی از بانک فیلتری که سیستم شنوایی انسان را مدل می‌کند، استفاده شده است و برای بررسی تغییرات

این مقاله در تاریخ ۱۳ تیر ماه ۱۳۹۳ دریافت و در تاریخ ۶ تیر ماه ۱۳۹۴ بازنگری شد. این مقاله از پایان نامه کارشناسی ارشد بهرز فرخی در دانشگاه تربیت مدرس استخراج شده است.

بهرز فرخی، دانشکده مهندسی برق، دانشگاه علم و صنعت ایران، تهران، (email: behraz.84@gmail.com)

احسان‌اله کبیر، دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران، (email: kabir@modares.ac.ir)

1. Wavelet Packet Transform
2. Onset Detection

$$X_{rms} = \sqrt{2} \cdot \log_{10} \frac{\sqrt{\sum_{i=0}^N x[i]^2}}{\sqrt{N} \times 2 \times 10^{-5}} \quad (1)$$

$$A_{norm} = 10^{\frac{X_{rms}}{20}} \quad (2)$$

$$x'[n] = A_{norm} \times x[n] \quad (3)$$

در این روابط، $x[n]$ و $x'[n]$ به ترتیب نمونه n ام از سیگنال اولیه و نرمال شده آن بوده و N تعداد نمونه‌ها است.

۲-۱ تحلیل زمان-فرکانس CQT

موسیقی ایرانی ربع پرده‌ای است، یعنی هر اکتاو در آن به ۲۴ قسمت تقسیم می‌شود. با فرض این که f فرکانس بم‌ترین نت در گستره فرکانسی نت‌ها باشد، فرکانس K امین نت طبق قاعده اعتدال در موسیقی از (۴) محاسبه می‌شود

$$f_k = f \times 2^{\frac{k}{24}} \quad (4)$$

چنانچه نسبت فرکانس هر نت به فاصله آن از نت بعد را با Q نشان دهیم، این مقدار که ضریب کیفیت نامیده می‌شود همواره ثابت و مستقل از فرکانس است. با در نظر گرفتن این موضوع برای تحلیل فرکانسی سیگنال موسیقی باید به دنبال تبدیلی بود که دقت فرکانسی مناسب را فراهم کند. تبدیلی را که برای این منظور در نظر گرفته‌ایم تبدیل با ضریب کیفیت ثابت CQT نام دارد که نخستین بار توسط براون در [۲۳] و [۲۴] ارائه شد. تبدیل CQT در کارهای متعددی که در زمینه پردازش سیگنال موسیقی انجام شده کارایی خود را نشان داده است. رابطه (۵) عنصر فرکانسی k ام از تبدیل CQT سیگنال گسسته در زمان $x[i]$ با فرکانس نمونه برداری f_s را نشان می‌دهد. در این رابطه $w[k, i]$ پنجره متغیر با فرکانس به طول $N[k]$ است. رابطه (۶) و (۷) به ترتیب پنجره همینگ استفاده شده در این تبدیل و نحوه محاسبه طول آن را نشان می‌دهد

$$x[k] = \frac{1}{N[k]} \times \sum_{i=0}^{N[k]-1} w[k, i] x[i] \cdot \exp \frac{-j2\pi Q i}{N[k]} \quad (5)$$

$$w[k, i] = \alpha + (1 - \alpha) \cos \frac{2\pi i}{N(k)}, \quad \alpha = \frac{25}{46} \quad (6)$$

$$N[k] = \frac{f_s}{f_k} \times Q \quad (7)$$

برای این کار هر اکتاو مطابق (۸) به ۹۶ قسمت تقسیم و در کل از ۷۱۰ نمونه فرکانسی برای پوشش پهنای ۱۳۰ Hz تا ۲۱ KHz استفاده شده است

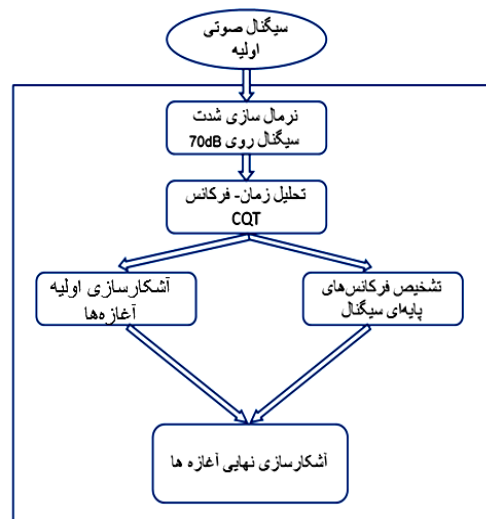
$$f_k = f \times 2^{\frac{k}{96}}, \quad 0 \leq k \leq 709 \quad (8)$$

در این تقسیم‌بندی، ضریب کیفیت تقریباً ۱۳۸ به دست می‌آید

$$Q = \frac{f_k}{\Delta f_k} \approx 138 \quad (9)$$

برای کاهش حجم محاسبات، تبدیل هر ۵ ms یک بار روی سیگنال اعمال و اندازه آن به شکل زیر در تابع CQT' و CQT' ذخیره می‌شود

$$CQT'(f_k, n) = |x(k)| \quad (10)$$



شکل ۱: نمودار مراحل آشکارسازی آغازه در تار.

تاریخی و فرهنگی، از ویژگی‌های ممتازی برخوردار است. جدای از فاصله ربع پرده که مشخصه ظاهری موسیقی شرقی و ایرانی است، تکنیک‌های متعدد نوازندگی، مقامات و گوشه‌ها، بافت غنی موسیقی ایرانی را شکل می‌دهد. تار به عنوان یکی از سازهای اصلی موسیقی ایران، دارای ظرایف و تکنیک‌های متعدد اجرایی است که بسته به مهارت نوازنده، هر دو نوع آغازه نرم و سخت در نوای آن دیده می‌شود. روشی را که برای آشکارسازی آغازه در ساز تار پیشنهاد کرده‌ایم از ترکیب دو مشخصه انرژی و فرکانس برای تعیین دقیق‌تر آغازه بهره می‌گیریم. با ترکیب این دو ویژگی، تابع آشکارساز قابلیت تفکیک و جداسازی بهتر نت‌های مجاور هم را یافته است. این ویژگی به خصوص در مواردی که فاصله زمانی اجرای دو نت پی در پی بسیار کوتاه است، اهمیت زیادی دارد. نکته دیگری که در شناسایی آغازه در تار مد نظر بوده، شناسایی نت ریز است. نت ریز از چند نت هم‌فرکانس، متوالی و با زمان اجرایی کوتاه تشکیل شده است. در آوانگاری موسیقی، تنها اولین نت از این مجموعه اهمیت داشته و دیگر نت‌ها در نوشتار موسیقی نمی‌آید. برای آشکارسازی نت ریز، هم‌زمان با آشکارسازی اولیه آغازه‌ها با استفاده از یک روش تطبیق کلیشه‌ای بهبودیافته، فرکانس پایه سیگنال تعیین و از تطبیق آغازه با فرکانس پایه نظیر، آشکارسازی نهایی آغازه‌ها انجام می‌شود.

بخش‌های مختلف این مقاله به این ترتیب آرایش یافته است: در بخش ۲ روش پیشنهادی برای آشکارسازی آغازه تار شرح داده می‌شود. در بخش ۳ کارایی روش پیشنهادی در مواجهه با مجموعه داده‌ای که برای این منظور گرد آورده‌ایم، سنجیده می‌شود. در این مجموعه از نواخته‌های استادان بزرگ موسیقی ایرانی همچون هوشنگ ظریف، غلامحسین بیگچه‌خانی و علی‌اکبرخان شهنازی که هر یک دارای سبک و شیوه نوازندگی منحصر به فردی هستند، استفاده شده است. نحوه انتخاب قطعات نیز بر اساس تنوع نت‌های ریز و معمولی در طول اجرا بوده است. بخش ۴ از این مقاله به جمع‌بندی مطالب اختصاص دارد.

۲- روش پیشنهادی برای آشکارسازی آغازه ساز تار

شکل ۱ مراحل به کار رفته برای آشکارسازی آغازه در تکنوازی تار را نشان می‌دهد که در ادامه، جزئیات آن توضیح داده می‌شود.

مشخصات سیگنال صوتی که در این سیستم پردازش می‌شود، تک‌کاناله با فرکانس نمونه‌برداری ۴۴۱۰۰ Hz است که شدت آن به روش

زیر روی ۷۰ dB نرمال می‌شود [۴]

ضرایب وزنی *Brown* به کار رفته در (۱۴) برای کاهش خطای اکتاو در تشخیص فرکانس پایه است. مقادیر این ضرایب که به صورت خطی از ۰٫۶ تا ۱ تنظیم شده، در جدول ۱ آمده است. برای کاهش نویز، طیف *Z* را با استفاده از پنجره میانگین گیر 5×5 هموار کرده و طیف حاصل را *Y* نامیده ایم

$$Y(f_k, n) = \frac{1}{25} \times \sum_{i=-2}^2 \sum_{j=-2}^2 Z(f_{k+i}, n+j) \quad (15)$$

در این مرحله، طیف $const(f_k, n)$ به شرح زیر محاسبه می‌شود

$$N(f_k, n) = norm\{Y(f_k, n)\} \quad (16)$$

$$const(f_k, n) = (N(f_k, n))^r \quad (17)$$

در (۱۷)، وظیفه عملگر $norm\{\}$ نرمال کردن تابع *Y* در لحظه دلخواه *n* نسبت به مقادیر ستون فرکانس است. شکل ۲ نمایش دوبعدی از طیف *N* است که برای بخش کوتاهی از یک تکنوازی تار به دست آمده است. خطوط افقی تیره رنگ نشان‌دهنده فرکانس‌های غالب هستند و محل‌ها در شکل با خطوط عمودی سفید رنگ مشخص شده است.

توابع $const(f_k, n)$ و $CQT'(f_k, n)$ که در این مرحله به دست آمد به عنوان ورودی برای دو مرحله بعد یعنی آشکارسازی اولیه آغازها و تعیین فرکانس‌های پایه سیگنال استفاده می‌شود.

۲-۲ آشکارسازی اولیه آغازها

تابع آشکارساز آغازها تابعی از زمان است و آغازها بر اساس جهشی که روی دامنه آن ایجاد می‌کند تعیین می‌شود. برای تعریف آن بر اساس آنچه در مقدمه گفته شد، بسته به جنس صوتی ساز، از سه ویژگی انرژی، فاز و فرکانس استفاده می‌شود. تابع پیشنهادی ما برای آشکارسازی آغازهای ساز تار، ترکیبی انرژی پایه و فرکانس پایه است. نخست تابع *D* به شرح زیر تعریف می‌شود

$$D(f_k, n) = H\{CQT'(f_k, n) - CQT'(f_k, n-1)\} \quad (18)$$

عملگر *H* شرطی بوده و به شکل زیر تعریف می‌شود

$$H(x) = \begin{cases} x & , x \geq 0 \\ 0 & , x < 0 \end{cases} \quad (19)$$

از محاسبه میانگین تابع *D*، تابع آشکارساز انرژی پایه به شکل زیر تعریف می‌شود

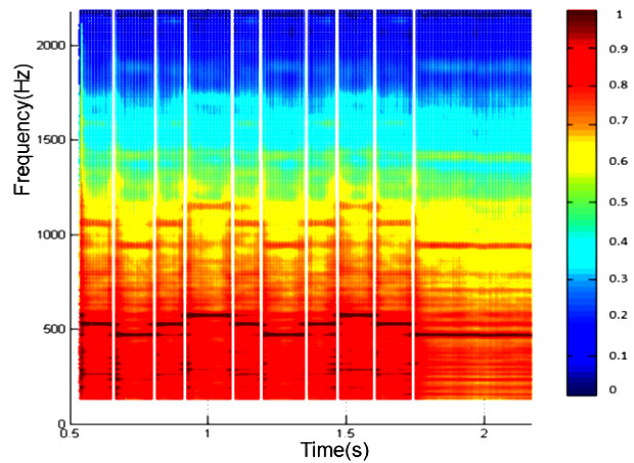
$$ED(n) = \sum_{k=1}^N D(f_k, n) \quad (20)$$

همان گونه که در این تعریف مشخص است، وزن در این رابطه یکسان و ثابت است. روش پیشنهادی ما، محاسبه میانگین وزن دار تابع *D* است به طوری که وزن‌ها در هر لحظه، متناسب با آرایش فرکانس‌های پایه‌ای سیگنال و هارمونیک‌های آنها تنظیم شود.

تابع آشکارساز آغازها پیشنهادی که انرژی و فرکانس پایه نامیده می‌شود به شکل زیر تعریف می‌گردد

$$EDP(n) = \sum_{k=1}^N Const(f_k, n) \times D(f_k, n) \quad (21)$$

ضرایب وزنی که در تعریف تابع آشکارساز استفاده شده است بر اساس الگوی فرکانس‌های غالب در طیف *Const* سیگنال آرایش یافته است. بنابراین با تأثیر آن روی تابع *D* که بر اساس ویژگی انرژی تعریف شد،



شکل ۲: طیف *N* از یک قطعه تار؛ خطوط قائم سفید رنگ برای مشخص شدن محل‌ها به شکل افزوده شده است.

$$CQT(f_k, n) = dB\{x(k)\} \quad (11)$$

در این روابط، *n* شماره پنجره زمانی به کار رفته در تبدیل و وظیفه عملگر $\{dB\}$ تبدیل مقدار به دسی‌بل است.

درک گوش انسان از بلندی صدا، علاوه بر انرژی به فرکانس آن نیز بستگی دارد. این موضوع حتی در نوازندگی سازها تأثیرگذار است. به این ترتیب که نوازنده برای ایجاد درک یکنواختی از بلندی صدا، ناخواسته، نت‌های مختلف را با شدت‌های متفاوت اجرا می‌کند. ارتباط میان فرکانس و شدت صوت برای ایجاد درک یکنواختی از بلندی صدا توسط منحنی‌های هم‌شدت شنوایی^۱ بیان می‌شود. برای کاهش تأثیر این پدیده ادراکی از منحنی هم‌شدت شنوایی ۷۰ dB ارائه شده در [۲۵] استفاده شده است. چنانچه اندازه منحنی هم‌شدت شنوایی ۷۰ dB در فرکانس *f* را با $Moor(f)$ نشان دهیم، طیف *A* ناشی از تأثیر این منحنی روی طیف *CQT* به شکل زیر به دست می‌آید

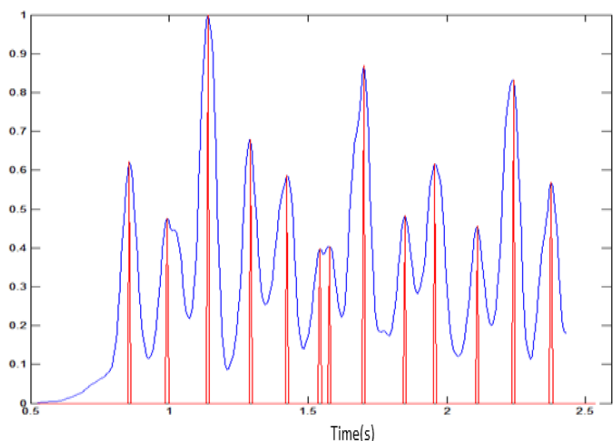
$$A(f_k, n) = CQT(f_k, n) - Moor(f_k) \quad (12)$$

در مقیاس لگاریتمی، فاصله بین هر فرکانس و هارمونیک‌هایش عددی مستقل از فرکانس و ثابت است. برای مثال هارمونیک سوم هر فرکانس همواره در فاصله ثابت $\log(3f/f) = \log(3/1)$ از آن قرار دارد. بر این اساس می‌توان الگوی یکنمایی از *m* هارمونیک نخست هر نت در مقیاس لگاریتمی تهیه کرد و با محاسبه همبستگی این الگو با فرکانس‌های مختلف طیف، به ازای بیشترین همبستگی، فرکانس پایه سیگنال را پیدا کرد [۲۶]. برای کار خود از ایده فوق استفاده کرده، الگوی از ۱۰ هارمونیک نخست را برای گروه‌بندی هارمونیک طیف سیگنال موسیقی تعریف کرده ایم. با استفاده از (۸) فاصله هر نت از هارمونیک *m* ام آن به طریق زیر محاسبه می‌شود

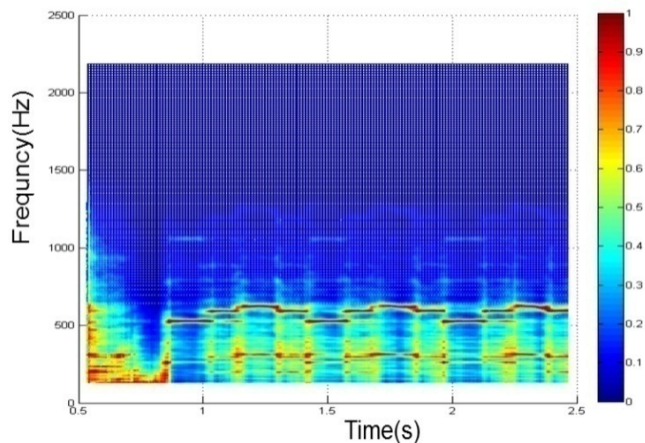
$$\Delta k(m) = 96 \log_2 m \quad (13)$$

برای ۱۰ هارمونیک نخست هر نت، گردشده مقادیر به دست آمده از رابطه فوق به همراه درصد خطایی که این تقریب ایجاد می‌کند در جدول ۱ آمده است. با محاسبه همبستگی این الگو با طیف *A*، طیف *Z* سیگنال به طریق زیر به دست می‌آید

$$Z(f_k, n) = \sum_{i=1}^{10} Brown(i) \times A(f_{k+\Delta k(i)}, n) \quad (14)$$



شکل ۴: تابع آشکارساز آغازه و آغازه‌های تعیین شده با استفاده از الگوریتم قله‌یاب سازگار برای طیف فرکانسی شکل ۳.



شکل ۳: طیف Const بخشی از یک قطعه تار.

جدول ۱: مشخصات الگوی به کار رفته برای گروه‌بندی هارمونیک طیف CQT.

m	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
$\Delta k(m)$	۰	۹۶	۱۵۲	۱۹۲	۲۲۳	۲۴۸	۲۷۰	۲۸۸	۳۰۴	۳۱۹
$\left \frac{mf_k - f_{k+\Delta k(m)}}{mf_k} \right \times 100$	۰	۰	۰.۱	۰	۰.۰۶	۰.۱	۰.۳	۰	۰.۲	۰.۰۶
$Brown(m)$	۱	۰.۹۵	۰.۹۱	۰.۸۶	۰.۸۲	۰.۷۷	۰.۷۳	۰.۶۸	۰.۶۴	۰.۶

قله‌یاب فوق برای تابع آشکارساز EPD به دست آمده است.

۲-۳ تعیین فرکانس‌های پایه سیگنال

در موسیقی تک‌صدایی، احتمال حضور یک نت در هر لحظه وجود دارد. در چنین مواردی همان گونه که پیشتر توضیح داده شد، ماکسیمم لحظه‌ای طیف Y مشخص‌کننده فرکانس پایه سیگنال خواهد بود. در ساز تار، قابلیت اجرای چند نت هم‌زمان وجود دارد و این موضوع دقت این روش را کاهش می‌دهد زیرا تنها یک فرکانس را پیدا کرده و در مواردی که چند نت هم‌زمان وجود داشته باشد، دچار اعوجاج خواهد شد. برای تشخیص تمام فرکانس‌های پایه سیگنال از ایده [۱۵] استفاده کرده‌ایم. برای تعیین فرکانس‌های پایه، ابتدا با استفاده از شرط (۲۶) تقاطعی از طیف $Const$ را که از آستانه α_1 بزرگ‌ترند، مشخص می‌کنیم

$$Const(f_k, n) \geq \alpha_1 \quad (26)$$

فرض کنید n_1 و n_2 نقاط مرزی از بازه $[n_1, n_2]$ بوده که از شرط (۲۶) برای فرکانس f_m به دست آمده است. در این صورت اگر دو شرط بعدی برقرار باشد، این فرکانس به عنوان فرکانس غالب بازه زمانی مفروض پذیرفته خواهد شد

$$\max \{ (Const(f_k, n))_{k=m, n \in [n_1, n_2]} \} > \alpha_2 \quad (27)$$

$$\max \left\{ \left(\sum_{n=n_1}^{n_2} Const(f_k, n) \right)_{k=1:N} \right\} = \sum_{n=n_1}^{n_2} Const(f_m, n) \quad (28)$$

پارامترهای α_1 و α_2 را به ترتیب ۰.۹۵ و ۰.۹۸ انتخاب کرده‌ایم.

۲-۴ آشکارسازی نهایی آغازه‌ها

هدف از این مرحله، برچسب گذاری و تأیید نهایی آغازه‌ها است. با در نظر گرفتن نتایج قبل یعنی آغازه‌هایی که در مرحله قبل آشکار شدند و فرکانس‌های پایه نظیر، آغازه‌هایی که چهار شرط زیر به ترتیب در مورد آنها برقرار شود تأیید خواهند شد:

دقت تابع آشکارساز بیشتر شده و هم‌زمان تأثیر دو مؤلفه انرژی و فرکانس را خواهیم داشت.

از آنجا که نت‌ها در طول اجرا با شدت‌های متفاوتی نواخته می‌شوند، الگوریتم قله‌یابی برای شناسایی آغازه‌ها، کارا خواهد بود که سازگار با شرایط متغیر در موسیقی باشد. الگوریتم قله‌یابی که در اینجا استفاده شده است روش ذکرشده در [۸] بوده و انطباق‌پذیری بالایی با شرایط متغیر اجرا در موسیقی دارد. برای این کار ابتدا تابع آشکارساز را به ازای هر پنج نمونه متوالی، هموار کرده و سپس نرمال آن محاسبه می‌شود. لحظاتی از تابع حاصل که هم‌زمان سه شرطی را که در ادامه آمده است، برقرار سازند به عنوان آغازه‌های اولیه انتخاب می‌شود. در این روابط برای راحتی از نام عمومی $detector(n)$ برای مشخص کردن هر دو نوع تابع ED و EPD استفاده شده است

$$detector(n) \geq detector(k) \quad , \quad n-a \leq k \leq n+a \quad (29)$$

$$detector(n) \geq \frac{\sum_{k=n-b}^{n+c} detector(k)}{c+b+1} + \delta \quad (30)$$

$$detector(n) \geq g_\alpha(n-1) \quad (31)$$

در دو شرط نخست، پارامترهای بازه a ، b و c به ترتیب ۴، ۸ و ۵ انتخاب شده‌اند. در شرط سوم، $g_\alpha(n)$ تابع آستانه به پارامتر α است و از (۲۵) به دست می‌آید

$$g_\alpha(n) = \max \{ detector(n), \alpha g_\alpha(n-1) + (1-\alpha) detector(n) \} \quad (25)$$

در این روابط، پارامترهای قله‌یابی δ و α به ترتیب ۰.۱۰۳ و ۰.۱۰۱ انتخاب شده است.

شکل ۳ طیف Const بخشی از یک قطعه تار است. شکل ۴ آغازه‌های نظیر که با الگوریتم قله‌یاب پیدا شده است را نشان می‌دهد. در این شکل، خطوط عمودی نشان‌دهنده آغازه‌هایی است که توسط الگوریتم

تفکیک نت‌های متوالی با فرکانس‌های برابر و دیرندهای بسیار کوتاه است. در مورد نت‌های تشکیل‌دهنده نت ریز، به دلیل فاصله زمانی کوتاه اجرای نت‌ها، جهش انرژی جز در مورد نت نخست چشم‌گیر نیست. مشکل دیگر شناسایی این نت‌ها، هم‌فرکانس بودن آنها است که شناسایی آغازها را از طریق روش‌های فرکانس پایه مشکل می‌کند. تابع آشکارساز آغاز پیشنهادی با ترکیب دو ویژگی انرژی و فرکانس این مسائل را تا حدود زیادی حل کرده است.

۳- آزمایش سیستم آشکارساز آغاز

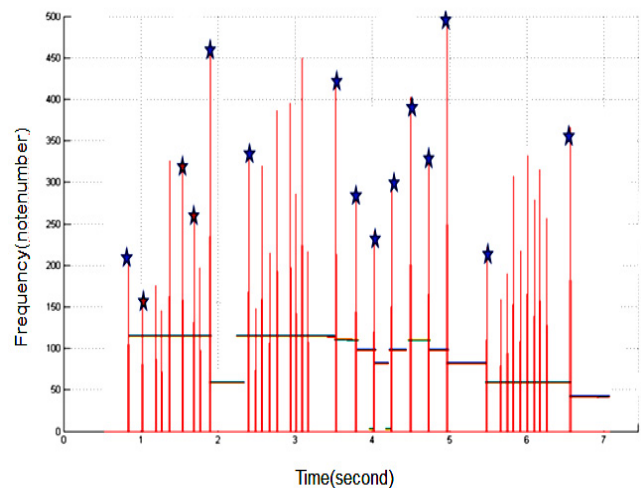
در بررسی قطعات و نواخته‌های اساتید موسیقی ایرانی، آنچه جلب توجه می‌کند، حضور تکنیک‌های متعدد نوازندگی است. این تکنیک‌ها و به کارگیری آنها تا حدود زیادی به سبک نوازندگی بستگی دارد. با این حال در میان تکنیک‌های متعددی که هر یک در آشکارسازی دقیق آغازها تأثیرگذار است، تکنیک‌های ریز و تک‌ریز حجم گسترده‌ای از اجرای نوازندگان تار را به خود اختصاص داده است. به طوری که توسط تمام نوازندگان پذیرفته و اجرا می‌شود و علائم مخصوصی برای آوانگاری آنها ابداع شده است. با توجه به این موضوع در این تحقیق تمرکز خود را بر شناسایی این بخش عمده از بافت موسیقی ایرانی گذاریم و شرط‌های بخش پس‌پردازش و به تبع آن گردآوری مجموعه دادگان نیز با این ملاک انجام شد. یعنی بررسی عملکرد سیستم در مواجهه با قطعاتی با تنوع نت‌های ریز، تک‌ریز و معمولی. برای کارایی‌سنجی، این سیستم را روی مجموعه‌ای از قطعات تکنوازی تار که توسط جمعی از اساتید برجسته موسیقی ایرانی اجرا شده است، آزموده‌ایم. ملاک دیگر که در انتخاب این مجموعه مد نظر قرار گرفت، انتخاب قطعات با سرعت‌های اجرایی از بسیار تند تا بسیار کند بود. به عنوان مثال قطعه شماره ۱۰ از جدول نتایج، سرعت اجرایی بسیار بالایی داشته و تعداد آغازهای زیادی در مدت زمان کوتاه در آن وجود دارد، در حالی که قطعه شماره ۵ که پیش‌درآمد شور نام دارد، از سکوت‌های با زمان‌های طولانی برخوردار است. جدول ۲، نام و آدرس قطعات و بازه زمانی استفاده‌شده از آنها برای آماده‌سازی مجموعه آزمون را نشان می‌دهد. در انتخاب داده‌های آموزشی که از آنها برای تنظیم پارامترها استفاده شده است از قطعاتی استفاده کرده‌ایم که در کلاس‌های آموزش تار استاد امیرحسین رضا در آموزشگاه موسیقی ظریف در تهران اجرا و توسط نویسندگان ضبط شده است. برای برآورد نتایج، طبق روال اکثر مقالاتی که در زمینه آشکارسازی آغاز ارائه شده است از ملاک F استفاده کرده‌ایم. اگر تعداد آغازهای صحیح آشکارشده را با CD نشان دهیم و تعداد آغازهایی را که به اشتباه آشکارشده یا آشکارشده هستند به ترتیب با FN و FP نشان دهیم، مقیاس F و دو پارامتر آن به کمک روابط زیر به دست می‌آید

$$P = \frac{CD}{CD + FP} \quad (29)$$

$$R = \frac{CD}{CD + FN} \quad (30)$$

$$F = \frac{2P \times R}{P + R} \quad (31)$$

برای تأیید انتخاب سیستم در آشکارسازی درست آغازها، حداکثر اختلاف زمانی ۵۰ میلی‌ثانیه از محل دقیق آغاز، مجاز فرض شده است. علت این انتخاب، عدم قطعیتی است که در هر اندازه‌گیری از جمله تعیین زمان دقیق نواخته‌شدن نت اجتناب‌ناپذیر است. جدول ۳ نتایج حاصل از آزمایش



شکل ۵: نتیجه مرحله پس‌پردازش روی آغازهای یک قطعه تار که با ستاره مشخص شده است.

(۱) اگر حداقل سه آغاز هم‌فرکانس و متوالی در فاصله کمتر از ۱۲۰ ms از یکدیگر قرار داشتند، تنها نخستین آغاز و به عنوان نت ریز پذیرفته می‌شود.

(۲) اگر یک نت ریز و آخرین نتی که پیش از آن اجرا شده است هم‌فرکانس بوده و در فاصله زمانی کمتر از ۱۵۰ ms از هم قرار داشته باشند، از این دو نت تنها نخستین آن در آوانگاری نهایی پذیرفته می‌شود. به این گونه نت‌ها، تک‌ریز می‌گویند که در آنها بین اولین و دومین زیرنت یک نت ریز، بیش از حد معمول تأخیر زمانی وجود دارد.

(۳) اگر دو نت ریز متوالی شناسایی‌شده توسط شرط نخست به گونه‌ای آرایش یافته باشند که آخرین نواخت نت ریز اول با فاصله کمتر از ۲۴۰ ms از اولین نواخت نت ریز دوم باشد، نت ریز دوم پذیرفته نمی‌شود.

(۴) اگر دو آغاز با فاصله کمتر از ۹۰ ms از یکدیگر قرار داشتند، آغاز با دامنه بزرگ‌تر پذیرفته می‌شود.

شرط سوم برای تصحیح اشتباهات احتمالی سیستم در شناسایی بعضی از زیرنت‌های تشکیل‌دهنده نت ریز است. در مواردی که از زیرنت‌های تشکیل‌دهنده نت ریز حتی یکی شناسایی نشود، تنها با در نظر گرفتن شرط اول، بیش از یک نت ریز شناسایی خواهد شد. بنابراین با کمک شرط سوم، امکان تصحیح این خطا وجود خواهد داشت.

شرط چهارم با این فرض به کار می‌رود که هیچ نوازنده‌ای قادر به نواختن دو نت متوالی با دیرند کمتر از ۹۰ میلی‌ثانیه نمی‌باشد.

شکل ۵ نتیجه نهایی سیستم آشکارساز آغاز را روی بخشی از یک قطعه تکنوازی تار نشان می‌دهد. خطوط عمودی نشان‌دهنده زمان و شدت نسبی آغازهای بالقوه هستند که توسط تابع آشکارساز آغاز و الگوریتم قله‌یاب به دست آمده است. بزرگی این خطوط برای تطبیق با منحنی دیگر بزرگ‌نمایی شده است. خطوط افقی، فرکانس‌های پایه نظیر را بر حسب شماره نت نشان می‌دهد. آغازهایی را که به کمک چهار قید بالا برجسب می‌خورند با علامت ستاره مشخص کرده‌ایم. آغازهای دوم، سوم و چهارم از سمت راست به عنوان خطای شناسایی سیستم هستند. این سه نت بر اساس مفهوم نت ریز که پیشتر توضیح داده شد، جزوی از مجموع نت‌های تشکیل‌دهنده یک نت ریز می‌باشند و نباید مستقلاً در آوانگاری موسیقی آورده شوند. بر اساس آنچه در این چهار شرط آمده است، دقت سیستم در شناسایی نت ریز تا حد بالایی وابسته به قدرت

جدول ۲: نام و آدرس قطعات و بازه زمانی استفاده شده در آزمون سیستم آشکارساز آغازه؛ در ستون مرجع، بازه زمانی به شکل (ثانیه: دقیقه) ذکر شده است.

شماره قطعه	مرجع	زمان (ثانیه)	تعداد نتها
۱	[۲۷] قطعه اول؛ ۲۳:۴۵-۲۲:۵۰	۵۵	۲۸۳
۲	[۲۷] قطعه دوم؛ ۵:۰۴-۵:۲۹	۲۵	۱۱۲
۳	[۲۸] سی دی ۲؛ قطعه ۱۸؛ ۰:۲۶-۰:۰۲	۳۴	۱۱۵
۴	[۲۸] سی دی ۴؛ قطعه ۲۹؛ ۰:۴۲-۰:۰۲	۳۳	۱۱۶
۵	[۲۹] کاست ۱؛ پیش درآمد شور؛ ۰:۲۶-۰:۰۰	۲۶	۷۲
۶	[۲۹] کاست ۱؛ رنگ شور؛ ۰:۲۱-۰:۰۰	۲۱	۷۲
۷	[۲۹] کاست ۱؛ رنگ بیات ترک؛ ۰:۴۵-۰:۰۰	۴۵	۱۶۲
۸	[۲۸] سی دی ۴؛ قطعه ۷؛ ۱:۰۴-۰:۰۹	۵۵	۱۳۴
۹	[۳۰] کاست ۲؛ اجرای داریوش پیرنیاکان؛ ۰:۵۱-۰:۰۰	۵۱	۴۱۵
۱۰	[۳۰] کاست ۲؛ اجرای غلامحسین بیگجه‌خانی؛ ۱:۱۰-۰:۳۰	۴۰	۲۱۹
مجموع		۶ دقیقه و ۱۵ ثانیه	۱۷۰۰

جدول ۳: نتایج به دست آمده از آزمون سیستم پیشنهادی.

شماره قطعه	ملاک F برای روش ED (%)	ملاک F برای روش EPD (%)
۱	۸۶,۲۲	۸۹,۹۰
۲	۸۳,۷۶	۸۷,۱۱
۳	۹۱,۲۲	۹۶,۱۰
۴	۸۵,۳۶	۸۵,۲۵
۵	۷۷,۲۷	۷۹,۶۸
۶	۸۵,۵۳	۸۷,۱۸
۷	۸۱,۱۸	۸۳,۳۳
۸	۸۳,۲۷	۸۸,۵۸
۹	۸۴,۳۵	۸۴,۱۰
۱۰	۸۴,۹۵	۸۷,۴۴
میانگین	۸۴,۳۱	۸۶,۹

جدول ۴: نتایج سیستم آشکارساز آغازه روی مجموعه داده‌ها که بر اساس شفافیت در اجرا به دو دسته تقسیم شده است: دسته نخست قطعات شماره ۴-۸ و ۱۰ را شامل می‌شود که قطعاتی با اجرای شفاف هستند، در حالی که دسته دوم شامل قطعات ۷-۵ و ۹ است که به وضوح دسته نخست اجرا نشده‌اند.

شماره دسته	۱	۲
میانگین ملاک F برای روش ED	۸۹,۰۶	۸۳,۵۷
میانگین ملاک F برای روش ED	۸۵,۷۹	۸۲,۰۸

علی‌اکبرخان شهنازی و داریوش پیرنیاکان اجرا شده است. این نواخته‌ها به شفافیت و وضوح قطعات دسته نخست نیست و بنابراین شناسایی نتها در این قطعات می‌تواند برای سیستم آشکارساز آغازه مشکل‌تر باشد. متوسط ملاک F برای هر دو دسته قطعه در جدول ۴ آمده است.

۴- جمع‌بندی

در این تحقیق روشی تازه برای آشکارسازی آغازه در ساز تار ارائه شد. روش فوق از دو بخش اصلی تشخیص فرکانس پایه و آشکارسازی اولیه آغازه‌ها تشکیل شده است که نتایج آنها در مرحله پس‌پردازش برای آشکارسازی نهایی آغازه‌ها استفاده می‌شود. برای تحلیل زمان-فرکانس سیگنال اولیه از تبدیل CQT استفاده شد و از نتیجه آن، تابع انرژی و فرکانس-پایه را برای آشکارسازی آغازه تعریف کردیم و در نهایت کارایی آن را در مواجهه با داده‌های آزمونی که برای این منظور تهیه و محل آغازه‌ها در آن از پیش مشخص شده بود، سنجیدیم. نتایجی که در این مرحله حاصل شد نشانگر بهبود عملکرد سیستم پیشنهادی در استفاده هم‌زمان از دو ویژگی انرژی و فرکانس در مقایسه با استفاده از ویژگی انرژی برای آشکارسازی آغازه‌های تار است. از مزایای تابع آشکارساز پیشنهادی، دقت بالای آن در تفکیک آغازه‌های بسیار نزدیک به هم است. علاوه بر این با تأثیر ضرایب وزنی‌ای که بر اساس الگوی فرکانسی سیگنال تنظیم می‌شود، حساسیت سیستم در برابر نویز کاهش یافت. در مرحله پس‌پردازش از چند ویژگی مهم در چیدمان نتها برای سازه‌های مضرابی ایران از جمله تار و سه‌تار استفاده کردیم.

مراجع

- [1] P. Marsi, Computer Modelling of Sound for Transcription and Synthesis of Musical Signal, Ph.D. Dissertation, University of Bristol, UK, 1996.
- [2] A. Klapuri, "Sound onset detection by applying psychoacoustic knowledge," in *IEEE Int. Conf. Acoust., Speech, and Sig. Proc.*,

سیستم پیشنهادی را برای هر دو نوع تابع آشکارساز آغازه انرژی پایه ED و تابع آشکارساز انرژی و فرکانس پایه EPD نشان می‌دهد. همان‌گونه که مشخص است، روش انرژی و فرکانس پایه در مقایسه با روش انرژی پایه پاسخ بهتری در شناسایی آغازه‌های ساز تار دارد. با مقایسه توابع آشکارساز ED و EPD می‌توانیم دلیل این قابلیت را بررسی کنیم. در تابع آشکارساز انرژی پایه، تغییرات محتوای انرژی در تمام نمونه‌های فرکانسی به دست آمده و از میانگین آنها با ضرایب وزنی یکسان برای تعیین تابع آشکارساز آغازه استفاده می‌شود. در حالی که در روش انرژی و فرکانس پایه، ضرایب وزنی به شکل لحظه‌ای و بر اساس فرکانس‌های قالب سیگنال، محاسبه و استفاده می‌شود. بر این اساس تغییرات انرژی در هر فرکانس بر اساس الگوی فرکانس‌های غالب و هارمونیک‌های آنها که احتمالاً متناظر با نواخته‌شدن نتی با همان فرکانس است ارزش‌گذاری شده و در محاسبه تابع آشکارساز آغازه استفاده می‌شود.

در این قسمت می‌خواهیم که برای دست‌یافتن به دید بهتر، نتایج به دست آمده را با دقت و جزئیات کامل‌تری بررسی کنیم. همان‌گونه که پیشتر گفته شد، در قطعات اجراشده با ساز تار، تکنیک‌ها و عناصر زیبایی‌شناختی متعددی بسته به نوازنده و سبک او مشاهده می‌شود. بر این اساس مجموعه داده‌ای که در این مقاله گردآوری شده است را به دو دسته تقسیم می‌کنیم: دسته نخست قطعات شماره ۴-۸، ۱۰ را شامل می‌شود و توسط استادان هوشنگ ظریف و غلامحسین بیگجه‌خانی اجرا شده است. یکی از مشخصه‌های این نوازندگان اجراهای تمیز و زخمه‌های شفاف آنها است. دسته دوم شامل قطعات ۷-۵ و ۹ است که توسط اساتید

- [19] P. M. Kumar, J. Sebastian, and H. A. Murthy, "Musical onset detection on carnatic percussion instruments," in *Proc. IEEE 21st National Conf. on Communications, NCC'15*, Mumbai, India, 6 pp., 27 Feb.- 1 Mar., 2015.
- [20] H. Thornburg, R. J. Leistikow, and J. Berger, "Melody extraction and musical onset detection via probabilistic models of framewise STFT peak data," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1257-1272, May 2007.
- [21] N. Degara, M. E. P. Davies, A. Pena, and Mark D. Plumbley, "Onset event decoding exploiting the rhythmic structure of polyphonic music," *IEEE J. of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1228-1239, Oct. 2011.
- [22] B. Zhu, J. Gan, J. Cai, Y. Wang, and H. Wang, "Adaptive onset detection based on instrument recognition," in *Proc. 12th IEEE Int. Conf. on Signal Processing, ICSP'14*, pp. 2416-2421, Hangzhou, China, 19-23 Oct. 2014.
- [23] J. C. Brown, "Calculation of a constant Q spectral transform," *J. Acoust. Soc. Am.*, vol. 89, no. 1, pp. 425-434, Jan. 1991.
- [24] J. C. Brown and M. S. Puckette, "An efficient algorithm for the calculation of a constant Q transform," *J. Acoust. Soc. Am.*, vol. 92, no. 5, pp. 2698-2701, Nov. 1992.
- [25] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 224-240, Apr. 1997.
- [26] J. C. Brown, "Musical fundamental frequency tracking using a pattern recognition method," *J. Acoust. Soc. Am.*, vol. 92, no. 3, pp. 1394-1402, Sep. 1992.
- [۲۷] تار هوشنگ ظریف، همراهی تنبک سیامک بنایی، انتشارات ماهور، سی دی شماره ۱۰۷.
- [۲۸] دستور مقدماتی تار، اجرا هوشنگ ظریف، انتشارات ماهور، سی دی شماره ۲۳۵.
- [۲۹] ردیف دوره عالی استاد علی اکبرخان شهنازی، انتشارات ماهور، مجموعه کاستی.
- [۳۰] "صد سال تار"، انتشارات ماهور، کاست شماره ۱۱۸.
- بهرآز فرخی** در تیرماه سال ۱۳۶۳ در زنجان به دنیا آمد. او کارشناسی و کارشناسی ارشد خود را در رشته مهندسی برق و الکترونیک، به ترتیب از دانشگاه زنجان و دانشگاه تربیت مدرس تهران و در سال‌های ۱۳۸۶ و ۱۳۹۰ دریافت کرد. او هم‌اکنون در مقطع دکتری مهندسی برق و الکترونیک، در دانشگاه علم و صنعت ایران مشغول به تحصیل است. زمینه‌های پژوهشی مورد علاقه او، پردازش سیگنال‌های صوتی و سیستم‌های واسط مغز و کامپیوتر است.
- احسان‌اله کبیر** در آبان ۱۳۳۷ در تهران به دنیا آمد. او کارشناسی ارشد پیوسته خود را در مهندسی برق و الکترونیک از دانشکده فنی دانشگاه تهران و دکترای خود را در مهندسی سیستم‌های الکترونیک از دانشگاه اسکس در انگلستان، به ترتیب در سال‌های ۱۳۶۴ و ۱۳۶۹ دریافت کرد. او اکنون استاد دانشکده مهندسی برق و کامپیوتر دانشگاه تربیت مدرس است. زمینه پژوهشی مورد علاقه او تحلیل و بازشناسی مستندات است.
- ICASSP'99*, vol. 6, pp. 3089-3092, Phoenix, AZ, USA, 15-19 Mar. 1999.
- [۳] ب. فرخی و ا. کبیر، "اشکارسازی آغاز در تکنوازی تار"، *اولین کنفرانس بین‌المللی آکوستیک و ارتعاشات، دانشگاه صنعتی امیرکبیر*، ۸ صص، ۳۰ آذر ۱ دی ۱۳۹۰.
- [4] B. Thoshkanna and K. R. Ramakrishnan, "A psychoacoustics based sound onset detection algorithm for polyphonic audio," in *Proc. Int. Conf. on Signal Processing, ICSP'08*, pp. 1424-1427, Beijing, China, 26-29 Oct. 2008.
- [5] E. Marchi, G. Ferroni, F. Eyben, S. Squartini, and B. Schuller, "Audio onset detection: a wavelet packet based approach with recurrent neural networks," in *Proc. IEEE Int. Joint Conf. on Neural Networks, IJCNN'14*, pp. 3585-3591, Beijing, China, 6-8 Jul. 2014.
- [6] J. P. Bello and M. Sandler, "Phase-based note onset detection for music signals," in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing, ICASSP'03*, vol. 5, pp. 441-444, 6-10 Apr. 2003.
- [7] J. Bello, C. Duxbury, M. Davies, and M. Sandler, "On the use of phase and energy for musical onset detection in the complex domain," in *IEEE Sig. Proc. Letters*, vol. 11, no. 6, pp. 553-556, Jun. 2004.
- [8] S. Dixon, "Onset detection revisited," in *Proc. 9th Int. Conf. on Digital Audio Effects, DAFX'06*, pp. 133-137, Montreal, Canada, 18-20 Sep. 2006.
- [9] R. Zhou, M. Mattavelli, and G. Zoia, "Music onset detection based on resonator time frequency image," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1685-1695, Nov. 2008.
- [10] N. Collins, "Using a pitch detector as an onset detector," in *Proc. Int. Soc. Music Inf. Retrieval Conf., ISMIR'0*, pp. 100-106, London, UK, 11-15 Sep. 2005.
- [11] R. Zhou, *Feature Extraction of Musical Content for Automatic Music Transcription*, Ph.D Dissertation, Swiss Federal Inst. of Technol., Lausanne, Oct. 2006.
- [12] H. Heo, D. Sung, and K. Lee, "Note onset detection based on harmonic cepstrum regularity," in *Proc. IEEE Int. Conf. on Multimedia and Expo., ICME'13*, 6 pp., San Jose, CA, Jul. 2013.
- [۱۳] پ. جعفرزاده، *آوانگاری اصوات موسیقی ایرانی*، دانشگاه سمنان، پایان‌نامه کارشناسی ارشد، اسفند ۱۳۸۰.
- [۱۴] پ. حیدریان، *بازشناسی نت‌های موسیقی برای سنتور*، پایان‌نامه کارشناسی ارشد، دانشگاه تربیت مدرس، بهار ۱۳۷۹.
- [15] J. Bello, et al., "A tutorial on onset detection in musical signals," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 5, pp. 1035-1047, Sep. 2005.
- [16] N. Collins, "A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions," in *Proc. 118th AES Conv.*, Barcelona, Spain, pp. 28-31, May 2005.
- [17] A. Holzapfel, Y. Stylianou, A. C. Gedic, and B. Bozkurt, "Three dimensions of pitched instrument onset detection," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1517-1527, Aug. 2010.
- [18] E. Benetod and Y. Stylianou, "Auditory spectrum-based pitched instrument onset detection," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 1968-1977, Nov. 2010.