

به نام خدا



دانشگاه صنعتی شاهرود

دانشکده مهندسی برق و رباتیک  
گروه الکترونیک

کلاس‌بندی دستگاه‌های موسیقی ایرانی به کمک ویژگی پرو فایل پیچ

دانشجو: ابراهیم گواهیان

استاد راهنما: دکتر حسین مروی

استاد مشاور: دکتر علی سلیمانی

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

تیرماه ۱۳۸۹



## تقدیر و تشکر:

با تشکر از استاد عزیزم جناب آقای دکتر مروی و همچنین جناب آقای دکتر احمدی فرد که مرا در انجام این مهم یاری کردند.

و همچنین با تقدیر ویژه از دوست همیشگی ام کوروش قنبری که در سخت‌ترین لحظات، امید و دلخوشی را به من ارزانی داشت.

## تقدیم به:

پدر و مادر عزیزم که صبورانه این فرصت را به من دادند که در جهت خواسته ها و علایقم گام بردارم و به دوست عزیزم یاسین امینزاده که با ایثار و فداکاری زمینه را برای پیشرفت کار و ایجاد انگیزه و امید در من فراهم کرد.



## چکیده:

مفهوم فواصل در موسیقی ایرانی و اثرات آن بر ساختار و کلاسه‌بندی ردیف موسیقی ایرانی یکی از مهمترین مسائلی است که موسیقی‌دانان به آن توجه می‌کنند. آنالیز مفهومی موسیقی همچنین، در سال‌های اخیر بسیار مورد توجه پژوهشگران و محققان در زمینه پردازش سیگنال قرار گرفته است. بحث تشخیص مقام یا الگو (در موسیقی غربی)، از روی سیگنال صوتی، ارتباط مستقیم با مفهوم فواصل دارد. یکی از ویژگی‌های اصلی که در بیان مفهوم فواصل نقش دارد پیچ است که در کاربردهای کلاسه‌بندی و آنالیزهای مربوط به بررسی خواص تشابه در حوزه بازیابی اطلاعات موسیقی، بسیار مورد استفاده قرار می‌گیرد.

در این پایان نامه، با مطالعه روشهای موجود آنالیز سیگنال‌های موسیقی غربی و مقایسه آنها، نهایتاً یک روش جدید برای تشخیص مقام‌های مختلف موسیقی ایرانی پیشنهاد شده است که هیچ محدودیتی برای سازهای موجود در قطعه موسیقی مورد آنالیز نداشته و نسبت به روشهای مشابه از سرعت بالاتری نیز برخوردار است. در تشخیص مقام، یک پروفایل پیچ ۲۴ تایی که شامل اطلاعات اساسی در مورد مقام قطعه مورد نظر است از سیگنال صوتی قطعه استخراج می‌شود. به علاوه، در بخش استخراج ویژگی، یک مرحله کاهش بعد در الگوریتم اعمال می‌شود که محاسبات را به طور قابل توجهی کاهش داده و موجب افزایش سرعت الگوریتم می‌شود. نهایتاً با اعمال الگوریتم پیشنهادی به دیتابیس نتایج چشمگیری مشاهده شده است.

واژگان کلیدی:

موسیقی ایرانی، مقام، بازیابی اطلاعات موسیقی، فاصله، پروفایل پیچ.

## لیست مقاله‌های استخراج شده از پایان نامه

[1] E. Gavahian and H. Marvi, "Mode detection of Persian music from Musical Audio", Digital Signal Processing Journal, Elsevier, under review.



# فهرست مطالب

## فصل اول:

۱	مقدمه ای بر تئوری موسیقی، مفاهیم بنیادی موسیقی
۱	۱-۱ مقدمه
۲	۲-۱ صدا
۳	۳-۱ ویژگی های فیزیکی
۵	۴-۱ صدا (Tone)، نغمه (Note) و نغمه نگاشت (Score)
۷	۵-۱ پیچ
۸	۶-۱ فاصله
۸	۱-۶-۱ فواصل مهم در موسیقی ایرانی
۸	۲-۶-۱ هنگام (Octave)
۹	۳-۶-۱ پرده و نیم پرده
۱۱	۴-۶-۱ تعدیل فواصل
۱۲	۵-۶-۱ فاصله چهارم و پنجم
۱۳	۶-۶-۱ سینت
۱۳	۷-۱ الگو (Scale)
۱۴	۱-۷-۱ گام (Gamut)
۱۴	۲-۷-۱ الگوی دیاتونیک (Diatonic Scale)

۱۶	۳-۷-۱ مد (Modes)
۱۶	۴-۷-۱ الگوی کروماتیک موسیقی ایرانی
۱۶	۵-۷-۱ مقام
۱۷	۸-۱ دستگاه
۱۹	۹-۱ صداهاى هماهنگ و غير هماهنگ
۱۹	۱۰-۱ طيف
۲۱	۱-۱۰-۱ طيف ديناميك و طيف استاتيك
۲۳	۱۱-۱ پوش دامنه
۲۴	۱۲-۱ ساختار پايان نامه

## فصل دوم:

۲۶	<b>بازيابی اطلاعات موسيقي: تعاريف، کاربردها، رويکردها و</b>
----	---

### چالشها

۲۶	۱-۲ مقدمه
۲۹	۲-۲ تعاريف
۲۹	۱-۲-۲ بازيابی
۲۹	۲-۲-۲ بازيابی اطلاعات
۳۰	۳-۲-۲ بازيابی مفهومی اطلاعات موسيقي
۳۱	۱-۳-۲-۲ موارد استفاده
۳۳	۲-۳-۲-۲ رويکردها
۳۳	۱-۲-۳-۲-۲ متا-ديتا
۳۷	۲-۲-۳-۲-۲ استخراج ویژگی های سطح بالای موسيقي (High_Level Features)

۳۹ ۳-۲-۳-۲ استخراج ویژگی های سطح پایین (Low-Level Audio Features)

۴۰ ۳-۲ بازیابی اطلاعات موسیقی، موسیقی غیر غربی و چالشهای روبرو

## فصل سوم:

۴۲ تبدیل با Q ثابت (Constant-Q-Transform)

۴۲ ۱-۳ مقدمه ای بر آنالیز سیگنالهای موسیقایی در حوزه فرکانس

۴۵ ۲-۳ محاسبه تبدیل Constant-Q

۴۶ ۳-۳ مقایسه تبدیل فوریه و تبدیل Constant-Q

۴۸ ۴-۳ کاربردها

## فصل چهارم:

۵۲ مروری بر برخی کارهای انجام شده در حوزه بازیابی  
اطلاعات موسیقایی جهت کلاسه بندی موسیقی

۵۲ ۱-۴ مقدمه

۵۳ ۲-۴ روش اول:

تشخیص کلید با استفاده از پروفایل پیچ

۵۴ ۱-۲-۴ الگوریتم کلی

۶۰ ۳-۴ روش دوم:

بازیابی اطلاعات موسیقی ترکی با استفاده از هیستوگرام پیچ

۶۱ ۱-۳-۴ محاسبه هیستوگرام پیچ

۶۳ ۲-۳-۴ تشخیص نغمه تونیک

۶۶ ۳-۳-۴ تشخیص مقام

۶۸	۴-۴ روش سوم:
	بررسی سیستم Tuning برای سیگنال های صوتی
۶۸	۱-۴-۴ ویژگی های استفاده شده
۶۸	۱-۱-۴-۴ میانگین طیفی
۶۹	۲-۱-۴-۴ کروما
۷۰	۲-۴-۴ نتایج

## فصل پنجم:

۷۲	کلاسه بندی دستگاه های موسیقی ایرانی با روش
----	--

## پیشنهادی

۷۲	۱-۵ مقدمه
۷۲	۲-۵ روش پیشنهادی برای کلاسه بندی دستگاه های موسیقی ایرانی
۷۳	۱-۲-۵ محاسبه اسپکتروگرام (تبدیل CQT)
۷۸	۲-۲-۵ کاهش بُعد
۸۰	۳-۲-۵ تعیین پیچ میزان سازی
۸۰	۴-۲-۵ استخراج بین های فرکانسی
۸۳	۵-۲-۵ استخراج پروفایل پیچ
۸۵	۶-۲-۵ اندازه گیری میزان
۸۵	۳-۵ نتایج
۸۶	۱-۳-۵ تئوری موسیقی مبنا
۸۷	۲-۳-۵ دیتابیس
۸۹	۳-۳-۵ الگوریتم های مورد مقایسه

۸۹	۴-۵ مقایسه نتایج
۹۰	۱-۴-۵ مقایسه اول
۹۰	۲-۴-۵ مقایسه دوم
۹۳	۳-۴-۵ مقایسه سوم
۹۵	۵-۵ نتیجه گیری
۹۶	۶-۵ پیشنهاد برای کارهای آینده

۹۷

**مراجع**

**چکیده انگلیسی**

## فصل اول:

### مقدمه ای بر تئوری موسیقی، مفاهیم بنیادی موسیقی ایرانی و

### معرفی برخی اصول مقدماتی پردازش سیگنال

#### ۱-۱ مقدمه:

موسیقی ایرانی به همراه موسیقی عربی و ترکی خانواده بزرگی را تشکیل می‌دهد که بیش از هزار سال قدمت دارد. این خانواده قرن‌ها دارای اصول و قواعد تقریباً واحدی بوده است و همه فرهنگ‌های تشکیل دهنده‌ی آن در تدوین این قواعد نقش داشته‌اند. از حدود سه، چهار قرن پیش، رفته‌رفته اعضای این خانواده بزرگ هرکدام راه جداگانه‌ای در پیش گرفته و مسیرهای متفاوتی برای تحول خود برگزیدند. این تجزیه منجر به پیدایش خانواده‌های موسیقایی کوچکتری مثل خانواده موسیقی عربی، خانواده موسیقی آسیای میانه و غیره شده است که با وجود پاره‌ای اختلافات، در بسیاری از مبانی نظری با هم اشتراک دارند. در این میان، موسیقی ایرانی به همراه موسیقی آذربایجانی، خانواده واحدی را تشکیل می‌دهد که نکات مشترک فراوانی با موسیقی عراقی دارد.

در میان انواع موسیقی غیر غربی و در مقایسه با موسیقی غربی، موسیقی‌های شرقی، به خصوص موسیقی ایرانی، از اهمیت ویژه‌ای برخوردارند. در این فصل هدف ما ارائه مقدمه‌ای در مورد صدا و موسیقی به همراه نگاهی اجمالی به تفاوت‌ها و اشتراکات موجود در موسیقی ایرانی و غربی است. در کنار

مباحث تئوریک مطرح شده، برخی اصول اولیه پردازش سیگنال نیز در مواجهه با کاربردهای موسیقایی آنها و جهت آشنایی معرفی می‌شوند.

تمامی شکل‌ها به جز شکل ۲ و ۶ و برخی از تعاریف موسیقی غربی، از مرجع [۱] گرفته شده و در برخی تعاریف موسیقی ایرانی نیز از مرجع [۲] استفاده شده است. مابقی توضیحات و اظهار نظرها، صرفاً برداشت و تحلیل نگارنده می‌باشد.

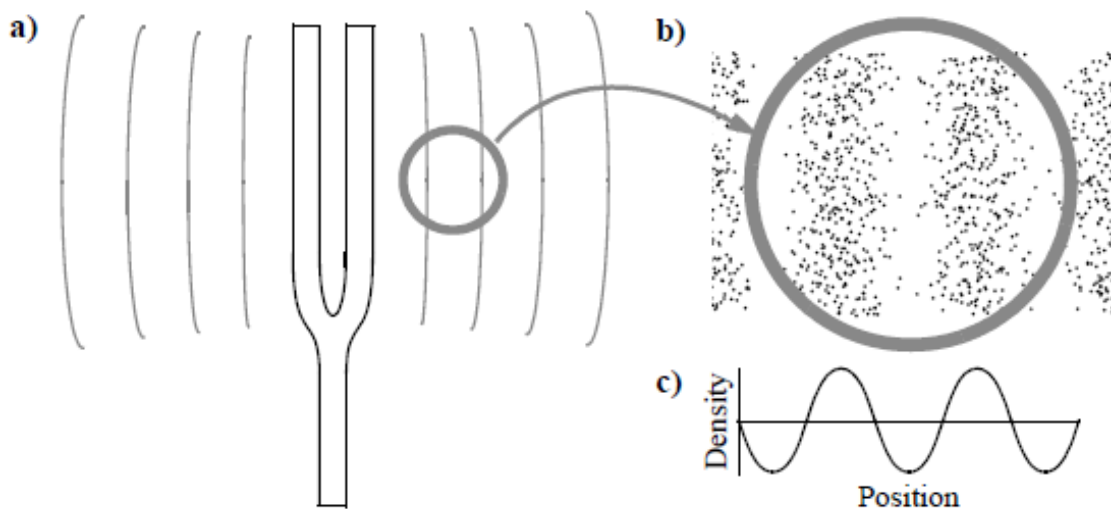
## ۱-۲ صدا:

صدا از ارتعاش اجسام به وجود می‌آید. هنگامی که جسمی مرتعش می‌شود هوای اطراف خود را نیز به ارتعاش در می‌آورد. به این ترتیب صدایی تولید می‌کند که توسط همین هوا منتقل می‌شود و به گوش ما می‌رسد. بدون هوا یا هر سیال دیگری صدایی وجود نخواهد داشت، کما اینکه اگر جسمی در خلا مرتعش شود صدایی تولید نمی‌کند.

بعضی صداها دارای ارتعاش منظم‌اند. یعنی کاملاً معلوم است که اجسامی که این صداها را تولید می‌کنند در یک ثانیه چند بار نوسان می‌کنند، مثل سازهای موسیقی. اما بعضی صداها دارای ارتعاش منظمی نیستند. یعنی معلوم نیست اجسامی که این صداها را بوجود می‌آورند چند بار در ثانیه نوسان می‌کنند، مثل صدای شکستن شیشه، بسته شدن در، شر شر آب و ...

اگر یک کوکه ی شاخه‌ای (دیپازون) را به صدا درآورده و نزدیک گوش خود بیاوریم یکی از خالص‌ترین و در عین حال ساده‌ترین صداهای طبیعت را خواهیم شنید. آنچه که می‌شنویم حاصل تغییرات متناوب فشار هوا در صماخ گوش می‌باشد که شاخه‌های دیپازون در هوا ایجاد میکنند. در شکل زیر نمایی از مولکولهای هوا در مجاورت دیپازون نشان داده شده است که مناطقی از هوا را که دارای فشار کمتر و بیشتر می‌باشند نشان می‌دهد. می‌توان این تغییرات در هوا را به تغییرات حاصل از پرتاب یک

سنگ در آب تشبیه کرد. حساسیت گوش انسان نسبت به شکل موجی که آن را تحریک می کند بسیار بالاست. چنانکه دقیقا قادر است منابع صوتی، موقعیت نسبی صدا و نیز مشخصات مهمی در مورد محیط پیرامون را از شکل موج آنها تشخیص دهد. اگر میانگین شدت صوت یک دیافازون را نمایش دهیم به شکل موجی شبیه آنچه در شکل ۱-۱C نشان داده شده می رسیم.



شکل 1-1: صدای تولید شده توسط یک دیافازون و تغییرات حاصل از آن در هوا.

### ۳-۱ ویژگی های فیزیکی:

نرخ تغییرات متناوب فشار، فرکانس است و شدت نوسانات فشار را شدت صدا<sup>۱</sup> می گویند. شدت صدا کیفیتی است که صدای قوی را از صدای ضعیف متمایز می کند. مشخصه ای که تغییرات شدت صدا در طول زمان را نشان می دهد پوش است. لحظه شروع صدا را آغاز<sup>۲</sup> می گویند و مدت زمانی که این صدا

<sup>1</sup> Intensity

<sup>2</sup> Onset



ادامه دارد، مادامی که ما آن را می‌شنویم، دیرند یا کشش<sup>۳</sup> نامیده می‌شود. از ویژگی‌های دیگر صدا طنین<sup>۴</sup> یا رنگ صداست. اگر یک صدای یکسان از دو ساز مختلف تولید شود آن خاصیتی از صوت که ما را قادر به تفکیک آن دو از هم می‌کند طنین است.

آخرین ویژگی فیزیکی یک صوت شکل موج آن است. سیستم شنوایی ما برای تشخیص کیفیت صدا (زیر و بمی، کر و یا تیز بودن صدا) از شکل موج صدا استفاده می‌کند. همچنین برای تشخیص منبع صدا (ویولن، سنتور و یا تار) نیز از شکل موج استفاده می‌کند. البته مشخصه‌های مهم دیگری نیز برای صوت وجود دارد که ما به معرفی فرکانس، شدت، آغاز، دیرند، پوش و شکل موج به عنوان مقدمه اکتفا می‌کنیم. فرکانس، تعداد نوسانات در ثانیه است و واحد آن هرتز می‌باشد. گوش انسان قادر است فرکانس‌های در محدوده ۱۷ هرتز تا ۱۷۰۰۰ هرتز را بشنود. البته فرکانس‌های بالا مربوط می‌شود به صدای کودکان و خانم‌ها و فرکانس‌های پایین به اجراهای برخی از انواع موسیقی مثل موسیقی راک. اما در موسیقی این گستره به فرکانس کمینه ۲۷.۵ هرتز و فرکانس بیشینه ۳۵۲۰ هرتز، آن هم برای ساز پیانو که یکی از کامل‌ترین سازهاست محدود می‌شود. شدت صوت با دسی بل اندازه‌گیری می‌شود. شدت صوت از ۴۰ دسی بل در محیط‌های خیلی آرام شروع شده و تا ۱۲۰ دسی بل که سقف شنوایی انسان است ادامه پیدا می‌کند. واحد اندازه‌گیری دیرند نیز ثانیه می‌باشد.

گوش انسان فرکانس تغییرات فشار هوای ناشی از ایجاد صدا را به عنوان پیچ به مغز می‌رساند. تغییرات شدت فشار<sup>۵</sup> تحت عنوان بلندی<sup>۶</sup> توسط مغز ادراک می‌شود. چنانکه در صورت عدم وجود تغییرات فشار هوا در اطراف گوش، چیزی که مغز ادراک می‌کند سکوت است. در مفهوم موسیقایی، آغاز و دیرند به

---

<sup>3</sup> Duration

<sup>4</sup> Timbre

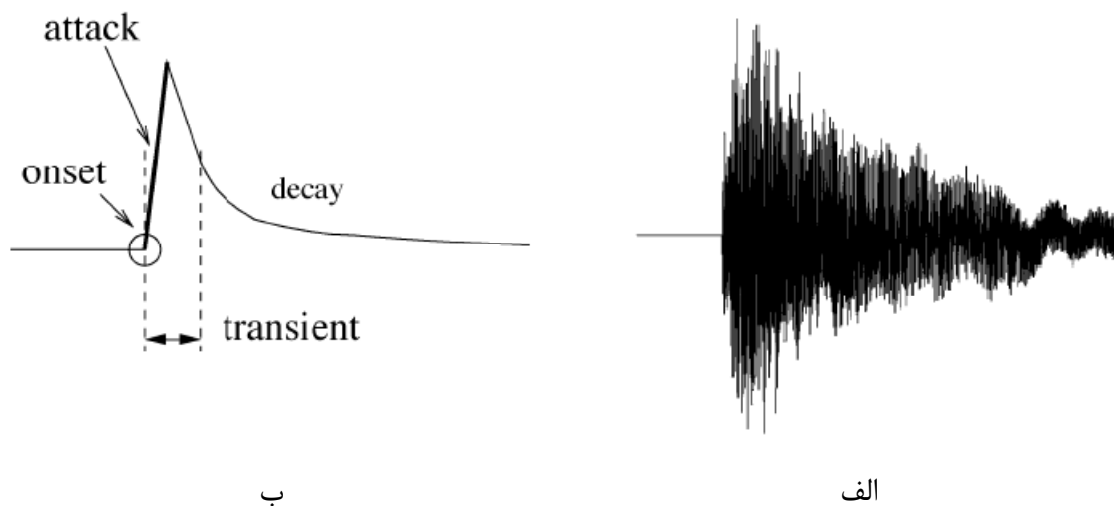
<sup>5</sup> Intensity

<sup>6</sup> Loudness

عنوان المان‌های ریتم شناخته می‌شوند. فرکانس، آغاز و دیرند جنبه‌هایی از صوت می‌باشند که بر مبنای تغییرات زمانی صوت شکل می‌گیرند و شدت صدا مقیاسی از تغییرات انرژی صوت است.

## ۱-۴ صدا<sup>۷</sup>، نغمه<sup>۸</sup> و نغمه نگاشت<sup>۹</sup>:

صدا با سه کیفیت شنیداری معنی پیدا می‌کند: پیچ، بلندی و طنین. وقتی این سه کیفیت اصلی صدا با دو پارامتر زمانی دیگر یعنی آغاز و دیرند ترکیب می‌شود نغمه پدید می‌آید. در واقع نغمه صدایی است که در قالب زمان قرار گرفته باشد. شماتیک یک نغمه واحد موسیقایی، به همراه نقطه آغاز و دیرند در شکل ۱-۲ نشان داده شده است. زمان شروع نغمه (آغاز) تا صفر شدن مجدد انرژی، دیرند نغمه می‌باشد.



شکل ۱-۲: نمایی از یک تک نغمه ایده آل، به همراه نقطه آغاز و تاخیر.

با قرار دادن نغمه‌ها در یک ترتیب زمانی، نغمه نگاشت پدید می‌آید که نحوه اجرای صحیح نغمه‌ها توسط سازهای مختلف را نشان می‌دهد. به عمل نوشتن نغمات موجود در یک قطعه موسیقی نغمه نگاری

<sup>7</sup> Tone  
<sup>8</sup> Note  
<sup>9</sup> Score

گفته می‌شود. امروزه تقریباً در همه جای دنیا از شیوه نغمه نگاری اروپایی، به همراه اضافاتی جهت بیان جزئیات موسیقی مربوطه، برای ثبت موسیقی استفاده می‌شود. در شکل ۱-۳ نمونه ای از یک قطعه موسیقی نغمه‌نگاری شده نشان داده می‌شود. به پنج خط موازی که نغمه های موسیقی بر روی آن نگاشته می‌شوند خطوط حامل می‌گویند.

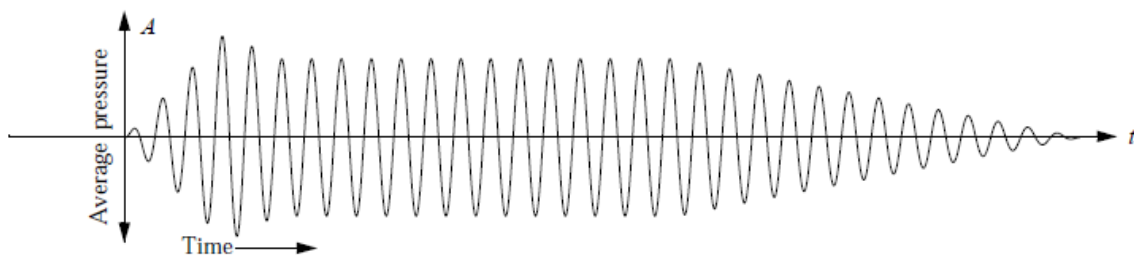


شکل ۱-۳: یک قطعه موسیقی نغمه نگاری شده برای گوشه درآمد عراق.

اگر یک نغمه تکی به صورت نشان داده شده در شکل الف ۱-۴، نغمه‌نگاری شود نمایش فیزیکی یا شکل موج آن به صورت شکل ب ۱-۴ خواهد بود.



الف



ب

شکل ۱-۴: الف: یک تک نغمه موسیقایی، ب: شکل موج آن

شکل ۱-۴ دو نمایش مختلف که حاوی اطلاعاتی یکسان از یک نغمه موسیقی می‌باشند را نشان می‌دهد. در شکل الف ۱-۴، شکل نمادین یک نغمه موسیقی نشان داده شده است و به نوازنده این امکان را می‌دهد که چگونه قطعه نت‌نگاری شده را بنوازد در حالی که در قسمت ب جنبه‌ی فیزیکی آن به نمایش درآمده است و در کاربردهای پردازش سیگنال و آنالیز سیگنال‌های موسیقی مورد استفاده قرار می‌گیرد.

### ۱-۵ پیچ:

اگر مقیاس فیزیکی ارتعاشات در ثانیه را فرکانس تعریف کنیم، پیچ مقیاسی برای درک مفهوم فرکانس خواهد بود. اصطلاحاً پیچ را شریک مفهومی فرکانس می‌نامند. در مقایسه فرکانس و پیچ این نکته قابل ذکر است که پیچ به گستره شنوایی انسان محدود می‌شود در حالی که برای فرکانس محدودیتی وجود ندارد. نکته دیگری که این دو مفهوم را از هم جدا می‌کند این است که غالباً اتفاق می‌افتد که برای بسیاری از نغمه‌های موسیقی فرکانس اصلی دارای انرژی بسیار اندک و یا حتی صفر می‌باشد اما مغز انسان پیچ آن را تشخیص می‌دهد؛ علی‌رغم اینکه فرکانس اصلی‌ای وجود ندارد. این تئوری که تحت عنوان تئوری تناوب [۳] مطرح می‌شود با ترکیب چند هماهنگ با انرژی بالا به شکل موجی می‌رسد که حاوی ویژگی‌های زمانی برجسته بوده و دوره تناوب آن نشان دهنده فرکانس اصلی مورد نظر است. بدین ترتیب یک نوع توانایی اندازه‌گیری دوره تناوب در مغز وجود دارد که می‌تواند اطلاعات یکسانی را از یک نغمه موسیقی چه حاوی فرکانس اصلی و چه فاقد آن بدست آورد. مفهوم دقیق‌تر این مسئله در فصل مربوط به تبدیل CQT با توضیحاتی که در مورد نحوه عملکرد این تابع ارائه می‌شود بیشتر درک خواهد شد.

یکی از مشکلاتی که در بررسی موسیقی فرهنگ های مختلف با آن روبرو هستیم نبود یک پیچ مرجع<sup>۱۰</sup> برای کوک کردن سازهاست که مشکلات فراوانی را در آنالیز سیگنال های موسیقی بوجود می آورد. این مشکل تنها در موسیقی غربی با قرار دادن نغمه  $A=440$  Hz مرتفع شده است و نوازندگان، بخصوص در اجراهای ارکستری، عموماً سازهایشان را مطابق با این فرکانس مرجع کوک می کنند. اما در موسیقی های غیر غربی اختلاف کوک گاهی تا ۲ نغمه بالاتر یا پایین تر از پیچ مرجع نیز می رسد.

## ۱-۶ فاصله:

اختلاف موجود در پیچ دو نغمه را فاصله می گویند. به عبارتی وقتی پیچ دو نغمه با هم متفاوت باشند آن دو از هم فاصله دارند. حساسیت گوش ما به فواصل اساس ملودی و علم هماهنگی است. موسیقی هر فرهنگی برای خود فواصل مشخص و منحصر به فردی دارد.

## ۱-۶-۱ فواصل مهم در موسیقی ایرانی:

در ادامه به معرفی فواصل مهم موسیقی ایرانی می پردازیم که برخی از آنها نیز با موسیقی غربی مشترک است.

## ۱-۶-۲ هنگام<sup>۱۱</sup>:

در تعریف موسیقایی، زمانی که دو نغمه کاملاً به هم شبیه اند، اما یکی زیرتر و دیگری بم تر است فاصله میان آنها یک یا چند هنگام (اکتاو) است. این فاصله از جمله فواصل مشترک بین موسیقی ایرانی،

---

<sup>10</sup> Reference Pitch

<sup>11</sup> Octave

موسیقی غربی و بسیاری دیگر از انواع موسیقی هاست. از دید ریاضی، اگر فرکانس یک صدا  $f_R$  باشد آنگاه صدایی که فرکانس آن  $f_R \cdot 2^1$  باشد را یک هنگام (اکتاو) بالاتر می‌دانند. و در حالت کلی اگر فرکانس صدایی  $f_R \cdot 2^x$  باشد آنرا  $x$  هنگام بالاتر می‌دانند.

### ۱-۶-۳ پرده و نیم پرده:

کوچکترین فاصله در موسیقی غربی فاصله نیم پرده است. چنان که در شکل ۱-۵ نشان داده شده است، از بین هفت نغمه اصلی که در موسیقی غربی وجود دارد پنج تای آنها دارای فاصله تمام پرده یا به اختصار پرده و دو تای آنها دارای فاصله نیم پرده می‌باشند.

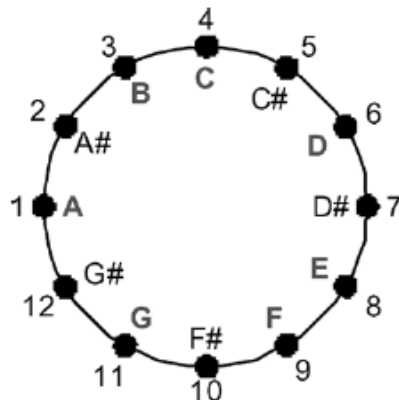
Degree:	1	2	3	4	5	6	7
Letter:	C	D	E	F	G	A	B (C)
Note:							
Syllable:	do	re	mi	fa	sol	la	ti (do)

Whole step:       Half step: 

شکل ۱-۵: فواصل بین نغمه های موسیقی غربی. برگرفته از [۱]

فواصل نیم پرده را دوم کوچک و فواصل تمام پرده را دوم بزرگ می‌گویند. با احتساب نغمه های موجود بین فواصل یک پرده ای دوازده نغمه جداگانه در محدوده یک هنگام برای موسیقی غربی خواهیم داشت. این دوازده نغمه و اسامی آنها به همراه نشانه های مورد استفاده برای هر کدام در شکل ۱-۶ نشان داده شده است.

در موسیقی ایرانی دو فاصله مهم دیگر نیز وجود دارد که در موسیقی غربی وجود ندارد و آن فواصل سه ربع پرده (۳/۴ پرده) و ۵/۴ پرده است. مفهوم ربع پرده که اساس تشکیل این دو فاصله مهم در موسیقی ایرانی را تشکیل می‌دهد خود بحثی جداگانه می‌طلبد که در این مجال نمی‌گنجد. تنها چیزی که ذکر آن



La	La#	Si	Do	Do#	Re	Re#	Mi	Fa	Fa#	Sol	Sol#
1	2	3	4	5	6	7	8	9	10	11	12

شکل ۱-۶: دوازده نغمه موسیقی ایرانی به همراه اسامی و نشانه‌های آنها. علامت #، دیز خوانده می‌شود.

لازم است اینکه ربع پرده تنها جنبه نظری داشته و در عمل به آن دقتی که نیم پرده و تمام پرده محاسبه می‌شوند، قابل محاسبه نیست. چرا که در موسیقی ایرانی ربع پرده هیچ‌گاه به طور دقیق و به شکل ۱/۴ پرده (یا نصف نیم پرده) اجرا نمی‌شود و همواره نسبت به موقعیت دقیق خود انحراف دارد. مفهوم ربع پرده در موسیقی ایرانی بسیار مورد بحث بوده و متغیر بودن آن، کار آنالیز سیگنال را نیز تا حدود زیادی مشکل می‌کند.

برای محاسبه فواصل با دقت ربع پرده به مفهوم ریاضی و نظری آن از رابطه زیر استفاده می‌شود:

$$f_x = f_R \cdot 2^{(x/24)} \quad (1-1)$$

طبق رابطه (۱-۱) در هر هنگام ۲۴ نغمه مختلف را با فواصل ربع پرده از هم خواهیم داشت.  $x = 0$  فرکانس نغمه اول و  $x = 24$  فرکانس نغمه یک هنگام بالاتر را نشان می‌دهد. با احتساب ۷ نغمه اصلی در

فاصله یک هنگام و در نظر گرفتن فواصل نیم و ربع پرده، جمعا ۲۴ نغمه مختلف در فاصله یک هنگام خواهیم داشت. نغماتی که به اندازه ربع پرده و نیم پرده کمتر از نغمه اصلی باشند به ترتیب با علامات کرن و بمل، و نغماتی که به اندازه ربع پرده و نیم پرده از نغمه اصلی بیشتر باشند به ترتیب با علامات سُری و دیز شناخته می‌شوند.

نکته مهمی که در فصل بازیابی اطلاعات موسیقی به تفصیل به آن اشاره می‌شود بحث تعدیل فواصل موسیقی غربی است که از چالش‌های اصلی در بازیابی اطلاعات موسیقی ایرانی (کلا موسیقی غیر غربی) بر اساس روش‌های موجود برای آنالیز سیگنال‌های غربی است. چرا که در موسیقی‌های غیر غربی از جمله موسیقی ایرانی فواصل تعدیل شده نبوده و هیچ فرکانس مرجع و استاندارد برای نغمه‌ها موجود نمی‌باشد. البته از آنجا که کمترین فاصله ای که گوش انسان قادر به تشخیص آن است حدود ۸ سنت می‌باشد [۱]، این ایراد به موسیقی غربی وارد است که تعدیل فواصل از توانایی‌ها و ظرافت‌های موسیقی غربی کاسته است. ظرایفی که موسیقی‌دانان ایرانی هرگز حاضر به فدا کردن آن برای رسیدن به یک گام تعدیل شده نشده‌اند و کماکان به تحرک فواصل معتقدند.

### ۱-۶-۴ تعدیل فواصل:

در علم فیزیک فاصله یک پرده را به ۹ قسمت مساوی تقسیم می‌کنند. به این فواصل ۱/۹ پرده‌ای کوما می‌گویند. در گذشته نیم پرده‌ها را به فاصله چهار کوما از نت می‌گرفته‌اند و بدین ترتیب بین دیز یک نت و بمل نت بعدی یک کوما فاصله وجود داشته و بر هم منطبق نبوده‌اند. اما بعدها، بنا بر قاعده ای که به آن تعدیل فواصل می‌گویند، از این اختلاف ۱/۹ پرده‌ای (۲۲ سنتی) صرف نظر شد و نتیجتاً نیم پرده‌های کروماتیک و دیاتونیک بر هم منطبق شدند [۴]. بدین ترتیب، در فواصل تعدیل شده موسیقی غربی، فرکانس همگی فواصل بر اساس یک فاصله واحد نیم پرده‌ای است. هر نغمه دارای فرکانس مشخص بوده و



نغمه های دیگر طبق رابطه زیر بر اساس یک فرکانس مرجع که معمولاً فرکانس نغمه  $A_4 = 440\text{Hz}$  است محاسبه می‌شوند. در رابطه زیر  $f_R$  همان  $A_4$  است.

$$f_x = f_R \cdot 2^{(x/12)} \quad (2-1)$$

طبق این رابطه فرکانس نغمه‌های مختلف در فاصله یک هنگام به صورت نشان داده شده در جدول زیر می‌باشند.

Frequencies of the Equal-Tempered Scale

$k$	Name	Frequency (Hz)	$k$	Name	Frequency (Hz)
0	A	440.000	7	E	659.255
1	$A_{\sharp}, B_{\flat}$	466.163	8	F	698.456
2	B	493.883	9	$F_{\sharp}, G_{\flat}$	739.988
3	C	523.251	10	G	783.990
4	$C_{\sharp}, D_{\flat}$	554.365	11	$G_{\sharp}, A_{\flat}$	830.609
5	D	587.329	12	A	880.000
6	$D_{\sharp}, E_{\flat}$	622.253			

جدول ۱-۱: فرکانس ۱۲ نغمه موجود در فاصله یک هنگام با فواصل نیم پرده. برگرفته از [۱].

### ۱-۶-۵ فاصله چهارم و پنجم:

اگر در شکل ۴، فاصله بین هر نغمه (درجه اول) و چهارمین نغمه بعد از آن (درجه پنجم) را بر حسب نیم‌پرده بشماریم، ۷ خواهد شد. به این فاصله ۷ نیم‌پرده‌ای که بین درجه اول و درجه پنجم وجود دارد در اصطلاح موسیقایی فاصله پنجم می‌گویند که یکی از فواصل مهم در انواع موسیقی است. به همین ترتیب فاصله بین هر نغمه (درجه اول) و سومین نغمه بعد از آن (درجه چهارم) نیز به فاصله چهارم معروف می‌باشد که یک فاصله ۶ نیم‌پرده ایست.

### ۱-۶-۶ سنت:

چنانکه دیدیم فاصله یک هنگام را می توان به ۱۲ نیم پرده تقسیم کرد. حال اگر بخواهیم با رزولوشن بالاتری این فاصله یک هنگامی را نشان دهیم می توانیم از تقسیمات سنت استفاده کنیم. بدین ترتیب که اگر فاصله یک هنگامی بین دو نغمه را به ۱۲۰۰ قسمت مساوی تقسیم کنیم، به هر کدام از واحد های جداگانه یک سنت می گویند و طبق رابطه (۱-۳) این تقسیمات را تشکیل می دهند:

$$f_x = f_R \cdot 2^{(x/1200)} \quad (۳-۱)$$

به هر کدام از واحدهای  $2^{(1/1200)} = 1.0005778$  یک سنت می گویند.

این واحد می تواند برای نشان دادن فاصله بین نغمات در هر نوع موسیقی اعم از موسیقی غربی یا غیرغربی مورد استفاده قرار گیرد. اما نکته ای که در موسیقی غربی وجود دارد این است که فواصل در موسیقی غربی تعدیل شده است یعنی به عنوان مثال، علی رغم اینکه طبق رابطه ۲ فاصله پنجم  $2^{(7/12)} = 701.955$  سنت می باشد اما در فواصل تعدیل شده این فاصله را ۷۰۰ سنت می گیرند. به این عمل رُند کردن فواصل بین نغمات که اولین بار توسط موسیقیدان بزرگ آلمانی، باخ، پیشنهاد شد تعدیل می گویند و تنها در موسیقی غربی وجود دارد و موسیقیدانان شرقی تاکنون حاضر به چنین کاری نشده اند.

### ۱-۷ الگو<sup>۱۲</sup>:

یک الگوی موسیقایی<sup>۱۳</sup> از یک سری پیچ های مرتب شده که از یک فرمول خاص پیروی می کنند تشکیل می شود. به هر کدام از پیچ های موجود در الگو یک درجه می گویند. به عنوان مثال برای یک الگو با  $n$  پیچ،  $n$  درجه مختلف وجود دارد که هر کدام نام درجه  $n$  ام را به خود می گیرند. یکی از الگوهای مهم در

<sup>12</sup> Scale

<sup>13</sup> Musical Scale

موسیقی ایرانی دانگ<sup>۱۴</sup> و گام است که هرکدام در جای خود بررسی می‌شوند. یک معادل خوب برای Scale در موسیقی ایرانی واژه تسلسل است که در برخی از منابع موسیقی ایرانی پیشنهاد شده است.

### ۱-۷-۱ گام<sup>۱۵</sup>:

هرگاه یک الگوی موسیقایی (برای موسیقی ایرانی، هندی، غربی و یا هر نوع موسیقی دیگر) در فاصله یک هنگام نوشته شود یک گام موسیقایی تشکیل می‌شود. برای موسیقی غربی در فاصله یک هنگام، ۱۲ نغمه‌ی مختلف با فواصل نیم‌پرده و در موسیقی ایرانی ۲۴ نغمه با فواصل ربع پرده موجود است که تمامی قطعات موسیقی از آنها تشکیل می‌شود.

### ۱-۷-۲ الگوی دیاتونیک<sup>۱۶</sup>:

الگوی پایه برای تمام سیستم‌های الگوی موسیقی غربی الگوی دیاتونیک است. این الگو در هر اکتاو ۷ پیچ دارد که با حروف A, B, C, D, E, F, G نمایش داده می‌شوند. الگوی دیاتونیک از دو نوع فاصله تشکیل شده است؛ فواصل نیم‌پرده<sup>۱۷</sup> و فواصل تمام‌پرده<sup>۱۸</sup>. همانطور که در شکل ۱-۷ نشان داده شده است در فاصله یک هنگام، برای موسیقی غربی، ۵ فاصله تمام‌پرده و ۲ فاصله نیم‌پرده وجود دارد. اگر برای فواصل نیم‌پرده عدد ۱ را اختصاص دهیم برای فواصل تمام‌پرده عدد ۲ را خواهیم داشت؛ شکل ۱-۸. دو الگوی مهم در موسیقی غربی الگوی بزرگ<sup>۱۹</sup> و الگوی کوچک<sup>۲۰</sup> می‌باشند که اولی از درجه اول الگوی دیاتونیک و دومی از درجه ششم آن آغاز می‌شود.

<sup>14</sup> Tetrachord

<sup>15</sup> Gamut

<sup>16</sup> Diatonic Scale

<sup>17</sup> Semi tone

<sup>18</sup> Whole tone

<sup>19</sup> major

Degree:	1	2	3	4	5	6	7
Letter:	C	D	E	F	G	A	B (C)
Note:							
Syllable:	do	re	mi	fa	sol	la	ti (do)
Whole step:							

شکل ۷-۱: شماتیک الگوی دیاتونیک غربی

	Starting degree						
	↓						
Degree:	1	2	3	4	5	6	7
Interval size:	2	2	1	2	2	2	1

شکل ۸-۱: اندازه فواصل و درجه های آنها

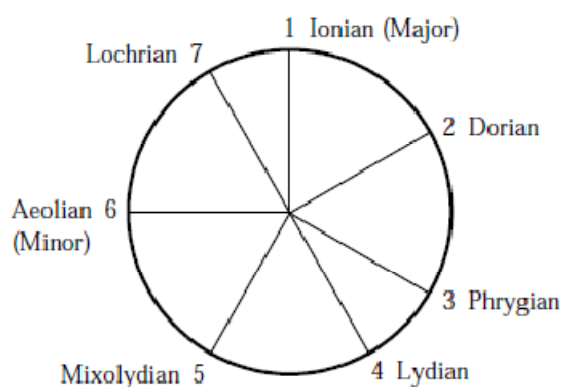
در موسیقی غربی اگر فواصل تمام پرده را به دو فاصله نیم پرده بشکنیم به الگویی دست می یابیم که آن را الگوی کروماتیک<sup>۲۱</sup> می نامند که یک الگوی بالقوه برای الگوهای مختلف موسیقی غربی از قبیل: الگوی دیاتونیک (۲،۲،۱،۲،۲،۲،۱)، الگوی پنتاتونیک (۲،۳،۲،۲،۳)، الگوی کوچک ملودیک (۲،۱،۲،۲،۱،۳،۱) و ... می باشد.

<sup>20</sup> minor

<sup>21</sup> Chromatic Scale

### ۱-۷-۳ مد ۲۲:

در موسیقی غربی اگر الگویی را از درجه ای غیر از درجه اول یا ششم الگوی دیاتونیک آغاز کنیم به الگوهای می‌رسیم که به آنها مد می‌گویند؛ شکل ۱-۹.



شکل ۱-۹: الگوی دیاتونیک و درجه‌های مختلف آن که تولید مد می‌کنند. برگرفته از [۱]

### ۱-۷-۴ الگوی کروماتیک موسیقی ایرانی:

الگوی کروماتیک موسیقی ایرانی که در واقع یک الگوی بالقوه برای انواع الگوهای موسیقی ایرانی است از ۲۴ نغمه با فواصل پی‌درپی ربع پرده‌ای در فاصله یک هنگام تشکیل می‌شود که اساس تشکیل مقام در موسیقی ایرانی است.

### ۱-۷-۵ مقام:

همانگونه که ذکر شد گام بالقوه اساس تشکیل مقام در موسیقی ایرانی می‌باشد. در تشکیل مقام سه عامل اصلی نقش دارند: ۱. گام بالفعل ۲. نظام سلسله مراتب درجه‌ها (نت شاهد و ایست) و ۳. سرمشق

<sup>22</sup> Modes

لحنی (سیر ملودیک). [۲] از آنجا که تمامی نغمه های موجود در گام بالقوه در تشکیل الگوهای موسیقی ایرانی نقش ندارند، ترکیب تعدادی از این نغمه ها و قرار گرفتنشان در فاصله یک هنگام تولید گام بالفعل می کند. چهار گام بالفعل موسیقی ایرانی از این قرارند: (۳،۳،۴،۴،۳،۳،۴)، (۴،۴،۲،۴،۴،۴،۲)، (۳،۴،۳،۴،۳،۴،۳) و (۳،۵،۲،۴،۳،۵،۲). در مقایسه الگوهای موسیقی ایرانی با الگوهای موسیقی غربی به این نکته پی می بریم که در موسیقی غربی به جز الگوی دیاتونیک (۲،۲،۱،۲،۲،۲،۱) تقریباً تمامی الگوهای دیگر منسوخ شده اند در حالی که در موسیقی ایرانی کماکان ۴ الگوی اصلی مورد استفاده قرار می گیرد. در موسیقی ایرانی واژه مقام تقریباً دارای مفهومی معادل با مد در موسیقی غربی است، با این تفاوت که چنان که اشاره کردیم، در تشکیل مقام های موسیقی ایرانی اصولاً به جز الگو، دو پارامتر دیگر سلسله مراتب درجه ها و سرمشق لحنی نیز نقش داشته اند.

نکته ای که در تکمیل بحث مقام حائز اهمیت است اینکه ملودی در موسیقی ایرانی نه در محدوده هنگام، بلکه در محدوده یک دانگ حرکت می کند، یعنی چهار یا حداکثر پنج نغمه را به کار می گیرد. اما هارمونیک های نغمه های اجرا شده در فاصله یک دانگ، گام بالفعل مربوطه را در محدوده یک هنگام تایید می کند.

## ۸-۱ دستگاه:

یکی از مفاهیم بنیادی در موسیقی ایران مفهوم دستگاه است. هر دستگاه در موسیقی ایرانی مجموعه ای از قطعات موسیقی است که به آنها گوشه می گوئیم. گوشه ها در مقام های مختلف ساخته شده اند و بر اساس نظم خاصی به هم پیوند خورده اند. یکی از نتایج این نظم خاص این است که گذر از یک مقام به مقامی دیگر در درون دستگاه به شکل خوشایند صورت می گیرد. در همه دستگاه ها یک مقام مادر وجود

دارد که نام خود را به دستگاه می‌دهد [۲]. به تعبیری می‌توان موسیقی ایرانی را به مثابه درختی در نظر گرفت که دستگاه، تنه این درخت و گوشه‌ها، شاخ و برگ این درخت را تشکیل می‌دهند.

در موسیقی ایرانی هفت دستگاه به نام‌های "شور، همایون، نوا، راست پنجگاه، سه‌گاه، چهارگاه و ماهور" و پنج آواز به نام‌های "دشتی، بیات زنده، بیات اصفهان، افشاری و ابوعطا" وجود دارد. هر کدام از دستگاه‌ها یا آوازه‌ها به منزله یک کلاس جداگانه است که تعدادی از گوشه‌های موسیقی ایرانی را در خود جای داده است. سه ویژگی اصلی در چگونگی کلاسه‌بندی ردیف دستگاهی موسیقی ایرانی نقش دارند: (۱) مفهوم فواصل؛ (۲) نغمه شاهد و ایست و (۳) سیر ملودیک. سیر ملودیک یا سرمشق لحنی شامل اطلاعاتی در مورد تغییرات زمانی قطعه است. اهمیت نغمه شاهد و ایست به منزله نغمه‌های تونیک و دامیننت در موسیقی غربی است، هرچند که مفاهیم آنها با یکدیگر متفاوت است. مفهوم فواصل و نغمه شاهد و ایست مربوط به ویژگی‌های فرکانسی صدا بوده و در ساخت الگوی فرکانسی قطعه نقش دارند که تعریف آنها به تفصیل بیان گردید. اما آنچه که معمولاً مورد بحث می‌باشد نحوه کلاسه‌بندی و چینش گوشه‌ها در هر یک از این دوازده کلاس جداگانه است. در [۵]، برای اولین بار و متفاوت با آنچه که تاکنون مطرح بوده است، چهار گام اصلی برای موسیقی ایرانی معرفی شده‌اند که تمامی این دوازده کلاس را می‌توان تنها به کمک این چهار گام اصلی معرفی کرد. از دیدگاه پردازش سیگنال آنچه که برای ما اهمیت دارد آن خواصی از صوت است که موجب اشتراکاتی در هر یک از گوشه‌های موجود در این مجموعه شده و آن‌ها را در یک کلاس خاص جای می‌دهد. به عبارتی بحث تشابه در کلاسه‌بندی دستگاه‌های موسیقی ایرانی بیانگر همین تشابهات و اختلافات موجود در گوشه‌ها و قطعات موسیقی ایرانی است. استخراج ویژگی مناسب که بتواند این تشابهات و اختلافات را نشان دهد از جمله کارهای اساسی است که در بررسی خواص تشابه در کلاسه‌بندی موسیقی ایرانی باید مد نظر قرار گیرد و بخش قابل توجهی از کار ما در مجموعه پیش رو را نیز به

خود اختصاص می‌دهد. در این پایان نامه تلاش ما بر استخراج آن ویژگی از موسیقی ایرانی است که مفهوم فواصل را بیان می‌کند و ما را به یکی از این چهار گام اصلی موسیقی ایرانی می‌رساند.

## ۹-۱ صداهای هماهنگ و غیرهماهنگ:

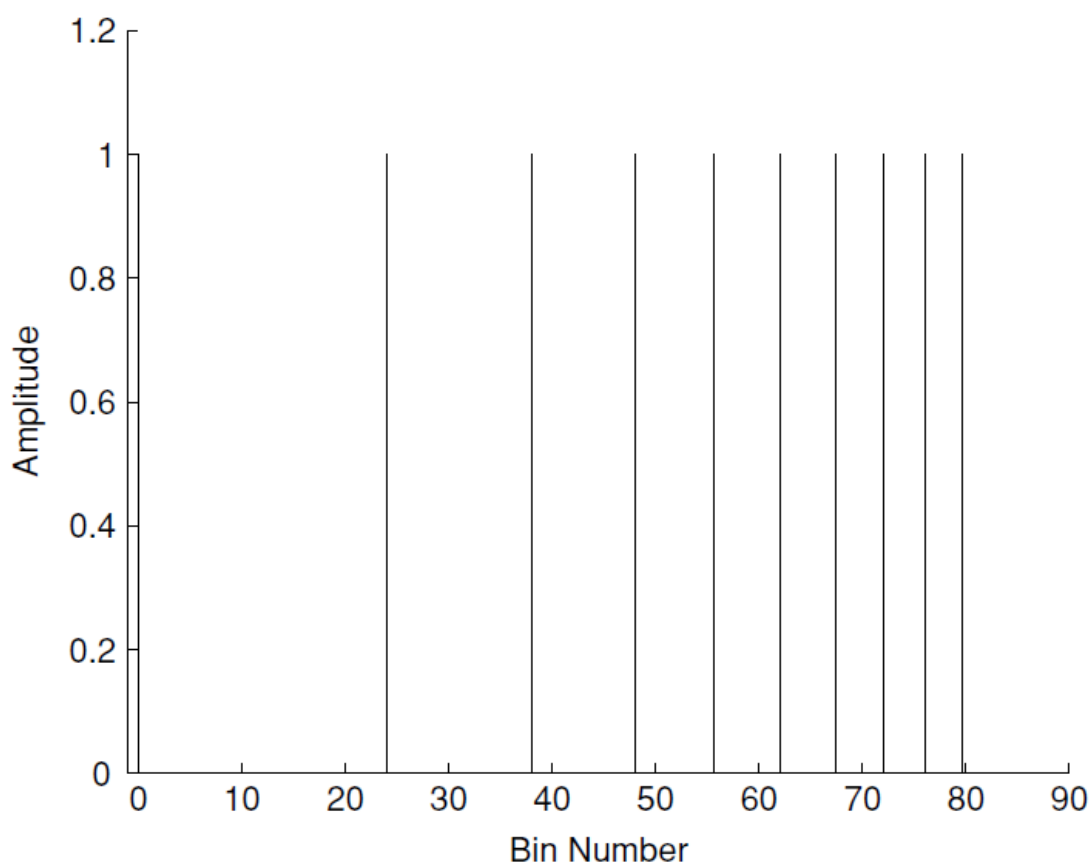
صداهای موسیقایی را در حالت کلی به دو دسته ی صداهای هماهنگ و صداهای غیرهماهنگ تقسیم می‌شوند. در مقابل سازهای ملودیک که صداهای هماهنگ تولید می‌کنند سازهای کوبه‌ای هستند که صداهای غیر هماهنگ تولید می‌کنند. سازهای ملودیک سازهایی هستند که اجزای فرکانسی آنها با یکدیگر هماهنگ بوده و به نوعی با نغمه‌های موجود در الگوی کروماتیک مطابقت دارند. سازهای کوبه‌ای یا غیر هماهنگ سازهایی هستند که اجزای فرکانسی آنها با نغمه‌های موجود در الگوی کروماتیک مطابقت ندارند.

برای صداهای هماهنگ اولین جزء فرکانسی فرکانس اصلی خوانده می‌شود. این فرکانس همان فرکانسی است که گوش آن را به عنوان پیچ صدا تشخیص می‌دهد. بقیه هماهنگ‌های موجود را اصطلاحاً **Overtone** می‌نامند. چنان که در فصل‌های بعد اشاره خواهیم کرد، اجزای فرکانسی صداهای هماهنگ ایجاد یک الگوی واحد می‌کنند، شکل ۱-۱۰، که مستقل از فرکانس اصلی صدای ایجاد شده است.

## ۱۰-۱ طیف:

اگر نور خورشید را به درون یک منشور بتابانیم، طیف سیگنال حاصل رنگ‌های موجود در نور خورشید را به صورت جداگانه نشان می‌دهد. در واقع منشور نور را به یک سری از فرکانس‌های خطی از کم به زیاد تفکیک می‌کند. شدت هر رنگ در رنگین کمان حاصل، نشان دهنده میزان مشارکت آن نور در کیفیت نور





شکل ۱-۱۰: الگوی واحد برای صداهای هماهنگ که نسبت به تغییر فرکانس اصلی مقاومند و تنها موقعیت بینها (الگوی فوق) بر روی محور افقی تغییر می کند.

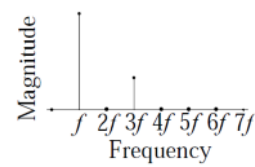
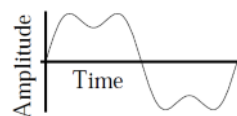
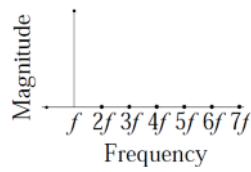
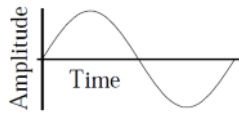
خورشید است. به همین ترتیب، در یک صوت نیز طیف صدا بیانگر شدت و فرکانس سینوسی‌هایی است که در تشکیل صدا نقش دارند. در واقع طیف صدا توزیع انرژی در شکل موج را نشان می‌دهد. طیف سیگنال شامل تمام فرکانس‌های ممکن از  $-\infty$  تا  $+\infty$  هرتز و شدت‌های  $0$  تا  $+\infty$  دسی‌بل می‌باشد. طیف یک صدای خاص زیر مجموعه‌ای از این فضای دو بعدی نامحدود می‌باشد. به عنوان مثال، در شکل ۱-۱۱، چهار شکل موج با طیف مربوطه اش نشان داده شده است. شکل موج الف ۱-۱۱ یک سینوسی واحد است که طیف آن یک خط عمودی تکی را نشان می‌دهد. خط افقی فرکانس سینوسی‌ها و خط عمودی شدت آنها را نشان می‌دهد. طیف دوم بیانگر این مطلب است که شکل موج مربوطه شامل دو

سینوسی است، فرکانس اصلی که در  $f$  اتفاق می‌افتد و هماهنگ سوم آن در  $3f$ . با توجه به کوتاه‌تر بودن اندازه شدت بین هماهنگ، می‌توان فهمید که انرژی آن کمتر از انرژی فرکانس اصلی بوده است. در دو شکل موج آخر طیف سیگنال برای هماهنگ‌های اضافه شده به سیگنال نشان داده شده است. در همگی آنها انرژی هارمونیک‌ها از انرژی فرکانس اصلیشان کمتر است. البته این قضیه عمومیت ندارد و با توجه به نوع ساز یا نغمه‌ی مربوطه هر کدام از هارمونیک‌ها یا فرکانس اصلی می‌توانند انرژی بیشتری را اختیار کنند. اگر می‌توانستیم صدای مربوط به شکل موج آخر را بشنویم، صدایی شبیه به صدای ساز کلارینت به گوش می‌رسید. همانگونه که اشاره کردیم، از آنجا که همگی بینهای فرکانسی مضارب صحیحی از فرکانس اصلی‌اند، به آنها طیف‌های هماهنگ می‌گوییم. اما چنان که در شکل ۱۱ نیز نشان داده شده است، این طیف یک طیف هماهنگ نمی‌باشد، چرا که اجزای فرکانسی مضارب صحیحی از فرکانس اصلی نمی‌باشند. اصطلاحاً به این طیف یک طیف ناهماهنگ می‌گویند. سازهای کوبه‌ای از قبیل تنبک، طبلا یا دف همچنین طیفی ایجاد می‌کنند.

### ۱-۱۰-۱ طیف دینامیک و طیف استاتیک:

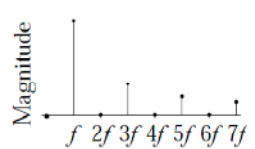
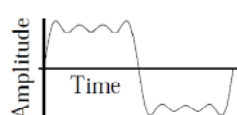
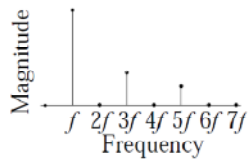
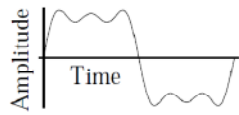
یک طیف استاتیک میانگین توزیع انرژی اجزای فرکانسی را بر روی یک بازه نسبتاً طولانی از زمان نشان می‌دهد. شکل بالا طیف استاتیک نغمه‌ها را نشان می‌دهد، چرا که المان زمان در طیف حاصل حذف شده و میانگین شدت اجزای فرکانسی برای نغمه مربوطه نشان داده شده است. اگر تغییرات استاتیک طیف سیگنال را در پنجره‌های زمانی کوچک محاسبه کرده و به یک طیف متغیر با زمان برسیم، طیف حاصل را طیف دینامیک سیگنال می‌گویند.

فرض کنید یک نغمه موسیقی داریم و می‌خواهیم توزیع انرژی در طول زمان آن را ببینیم، در این صورت:



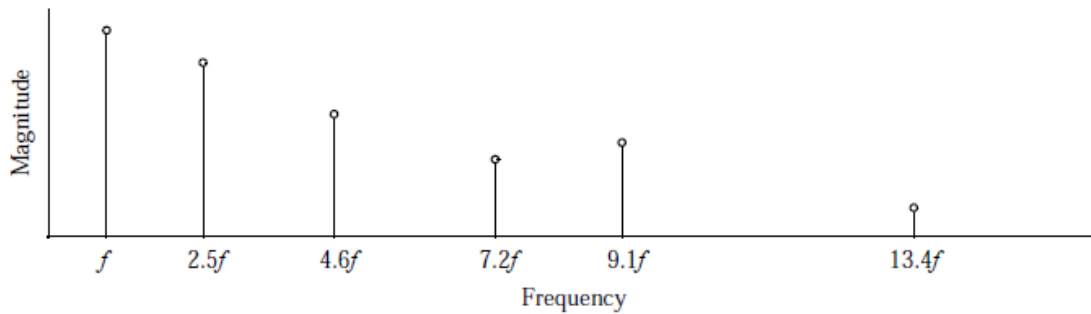
الف

ب



ج

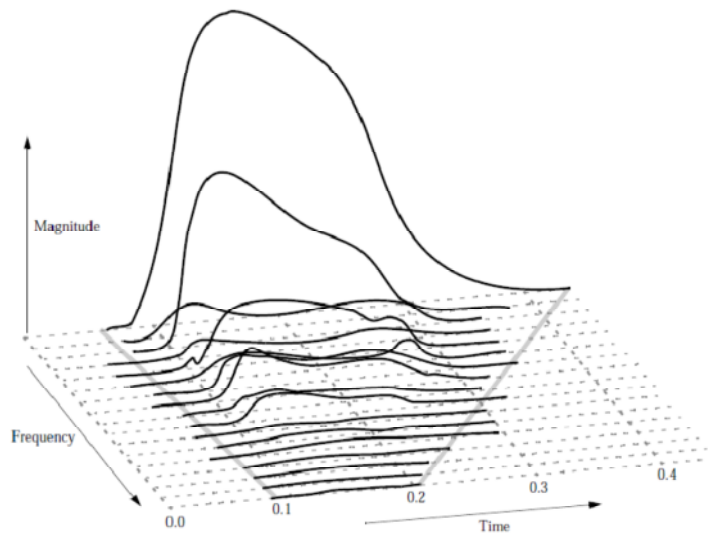
د



و

شکل ۱-۱: شکل موج سیگنال های هماهنگ (الف تا د) و ناهماهنگ (و) و طیف آنها. برگرفته از [۱]

۱. سیگنال را به یک سری از پنجره های کوچک زمانی تقسیم می کنیم.
  ۲. برای هر کدام از بخش های سیگنال طیف استاتیک را جداگانه محاسبه می کنیم.
  ۳. طیف های حاصل را با توجه به تغییرات زمان مرتب می کنیم.
- در شکل ۱-۱۲ طیف دینامیک سیگنال برای یک نغمه موسیقایی نشان داده شده است. از این خاصیت در فصل آخر برای استخراج طیف استاتیک از طیف دینامیک و کاهش محاسبات استفاده شده است.



شکل ۱-۱۲: طیف دینامیک سیگنال برای یک نغمه موسیقایی. برگرفته از [۱]

با دیدن طیف دینامیک فوق و با توجه به موقعیت قرار گرفتن بین‌های فرکانسی به خوبی مشخص است که طیف حاصل از روی یک سیگنال هماهنگ بدست آمده است. تمرکز انرژی بر روی هماهنگ‌های اولیه و کاهش چشمگیر آن در هماهنگ‌های بالاتر همگی نشان از هارمونیک بودن صدا دارند. هماهنگ‌های اولیه زودتر به صدا در آمده و دیرتر از هماهنگ‌های بعدی میرا می‌شوند.

### ۱-۱۱ پوش دامنه:

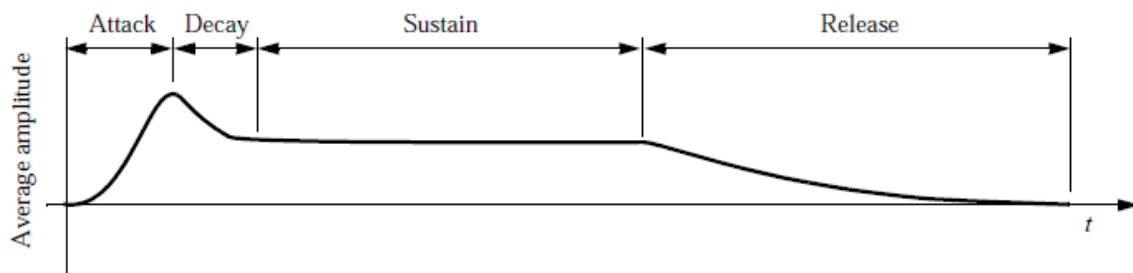
چنان که گفتیم اجزای فرکانسی صدا تنها به کمک سه ویژگی نشان داده می‌شوند؛ دامنه، فرکانس و زمان. تا کنون دو نمایش مختلف از اجزای فرکانسی صدا را نشان دادیم:

۱. طیف دینامیک سیگنال که یک نمایش سه بعدی از اجزای فرکانسی صوت مربوطه است.

۲. طیف استاتیک صوت که با میانگین گرفتن از دامنه هر یک از اجزای فرکانسی سیگنال در طول زمان

به طور جداگانه و حذف پارامتر زمان حاصل می‌شود.

اما با میانگین گرفتن از کل اجزای فرکانسی در طول زمان و حذف پارامتر فرکانس یک نمایش دو بعدی دیگر که در آن دامنه بر حسب زمان نشان داده شده است می‌رسیم. به این نمایش دو بعدی پوش دامنه می‌گویند که در شکل ۱-۱۳ نشان داده شده است.



شکل ۱-۱۳: پوش دامنه یک صدای موسیقایی. برگرفته از [۱]

## ۱-۱۲ ساختار پایان نامه:

در فصل اول مقدمه‌ای در مورد صدا، تئوری موسیقی و مفاهیم بنیادی موسیقی ایرانی ارائه شده است. به علاوه برخی توابع و آنالیزهای مربوط به پردازش سیگنال که در آنالیز مفاهیم موسیقی و استخراج ویژگی‌های مورد استفاده در این حوزه نقش دارند نیز معرفی می‌شوند.

فصل دوم به معرفی و شناسایی شاخه‌های اطلاعات موسیقی پرداخته و موارد کاربرد و رویکردهای مختلف آن را بیان می‌کند.

در فصل سوم یکی از تبدیل‌های اساسی در حوزه پردازش سیگنال‌های موسیقی که از اهمیت بسیار بالایی برخوردار است معرفی شده و با تبدیل معروف فوریه مقایسه می‌شود.

برخی کارهای انجام شده در حوزه بازیابی اطلاعات موسیقی غربی و شرقی که با این پایان نامه شباهت و هم‌خوانی قابل توجهی دارند در فصل چهارم معرفی و بررسی می‌شوند.

نهایتاً فصل پنجم به معرفی روش پیشنهادی پرداخته و نتایج حاصل را به تفصیل بیان می‌کند.

## فصل دوم:

### بازیابی اطلاعات موسیقی: تعاریف، کاربردها، رویکردها و چالشها

#### ۱-۲ مقدمه:

امروزه به لطف سیستم‌های دیجیتال دسترسی به فایل‌های موسیقی آنچنان آسان گشته که در یک مجموعه شخصی ساده هزاران قطعه موسیقی به راحتی و بدون در نظر گرفتن محدودت‌های زمانی لازم برای گوش دادن به آنها (به طوری که برای یک مجموعه چند هزار قطعه‌ای روزها زمان لازم است) یافت می‌شوند. توزیع آثار جدید موسیقی ساده‌تر شده و افزایش چشمگیری در حجم کلکسیون‌های موسیقی دیده می‌شود.

روشهای سنتی گوش دادن به موسیقی و راه‌های در دسترس بودن آن از قبیل شبکه‌های رادیویی یا فروشگاه‌های بزرگ نوارهای کاست جای خود را به سیستم‌های دیجیتالی داده و بیشتر جنبه شخصی و انفرادی پیدا کرده‌اند. ظهور وب سایت‌های شبکه‌های اجتماعی<sup>۲۳</sup> یک جهش سریع در ایجاد راه‌های جدید برای بدست آوردن موسیقی را موجب شده که اساس روشهای توزیع و پخش موسیقی را به کلی دگرگون کرده و شرکتهای بزرگ انتشار موسیقی را بر آن داشته که در سیاستهای کاری خود تجدید نظر کنند. در سال ۲۰۰۵ یک رشد سه برابری در دانلود قانونی موسیقی و صدای زنگ تلفنهای همراه به ارزش ۱.۱ میلیارد دلار گزارش شده است که طبیعتاً کاهشی معادل را در فروش CD موجب می‌شود [۶].

<sup>23</sup> social networking Web sites

بنابر این لازم است که راه‌حلهایی برای دستیابی به این مجموعه‌های موسیقی با در نظر گرفتن انتظارات کاربران جهت دانلود و آپلود ارائه شود. به مجموعه این راه‌کارها بازیابی اطلاعات موسیقی<sup>۲۴</sup> گفته می‌شود که اخیراً بخش قابل توجهی از مقالات تحقیقی را به خود اختصاص داده است. بازیابی اطلاعات موسیقی یک گرایش نوپا و بسیار فعال در عرصه فعالیت‌های تحقیقاتی است. این قضیه به روشنی در مقالات منتشر شده در کنفرانس‌های بین‌المللی سالانه بازیابی اطلاعات موسیقی<sup>۲۵</sup> قابل مشاهده است. همچنین انتشار تحقیقات مربوط به بازیابی اطلاعات موسیقی در مجلات علمی معتبر از قبیل ACM، IEEE و conference on Computer Music Modeling and Retrieval (CMMR) و جزآن گواه دیگری بر این مدعی است..

در حال حاضر بیشتر روشهای معمول برای دستیابی به موسیقی بر استفاده از متا-دیتا استواراند، که در بخشهای بعدی در مورد آن توضیح داده خواهد شد. از آنجایی که متا-دیتا غنی و گویا است به تعبیر بسیاری، جهت کاربردهای جستجو و ارائه موسیقی کفایت می‌کند. بیشتر سرویس‌های دانلود موسیقی از روشهای مبتنی بر تنها متا-دیتا استفاده می‌کنند که البته یک موفقیت نسبی نیز برای آنها به ارمغان داشته است. اما بایستی به این نکته توجه کرد که برای کلکسیون‌های بسیار بزرگ (حاوی بیش از صدها هزار قطعه) تهیه متا-دیتای مستحکم و قابل اعتماد کار بسیار دشواری است، چرا که برای هرکدام از آثار مختلف بروشورهای متفاوتی ارائه شده و تفاوت‌های موجود در آنها کار جستجو را بسیار دشوار و غیر قابل اعتماد می‌کند. علاوه بر این، از آنجایی که بروشورهای مختلف بیانگر عقاید و طرز تفکرهای گوناگون می‌باشند یک نوع نظارت بر بروشورها لازم و ضروری است.

علاوه بر سیستمهای مبتنی بر متا-دیتا، اطلاعات مربوط به مفهوم موسیقی نیز می‌تواند در یافتن موسیقی مورد نظر به کاربران کمک کند. در سیستمهای مبتنی بر شرح مفهومی موسیقی، آنچه که کاربر

<sup>24</sup> Music Information Retrieval (MIR)

<sup>25</sup> International society on Music Information Retrieval (ISMIR), <http://www.ismir.net/>



به دنبال آن است تشخیص داده می‌شود، هرچند ممکن است خود کاربر دقیقا نداند به دنبال چیست. به عنوان مثال اگر شما در یک مجلس عمومی یا مراسمی هستید که در آنجا موسیقی در حال پخش شدن است و بخواهید اطلاعاتی در مورد موسیقی در حال پخش کسب کنید یا حتی بخواهید آنرا به طور آن-لاین از اینترنت خریداری کنید کافی است با ارائه یک نمونه صوتی توسط تلفن همراهتان به سیستم شازم<sup>۲۶</sup> به خواسته خود برسید. یا گاهی ممکن است تنها اطلاعاتی که از قطعه مورد نظرتان دارید ملودی آن قطعه باشد؛ نه شرکت توزیع کننده آنرا می‌شناسید، نه نوازندگان یا خواننده آن را، و نه حتی اسم آهنگ را می‌دانید، تنها ملودی آن قطعه در ذهن شما باقی مانده است، در این گونه موارد می‌توانید به پایگاه اطلاع رسانی نایو<sup>۲۷</sup> مراجعه کرده، ملودی قطعه مورد نظر را زمزمه کنید و سیستم آن را برای شما تشخیص می‌دهد.

در بین رویکردها و کاربردهای مختلفی که برای بازیابی اطلاعات موسیقی وجود دارد، آنالیز مفهومی سیگنال های صوتی<sup>۲۸</sup> از اهمیت ویژه ای برخوردار است. برای یک فایل موسیقی نمایش های مختلفی وجود دارد اما آنالیز مفهومی موسیقی تنها با نوع نمایش صوتی آن سروکار دارد.

قبل از هرچیز لازم است با معانی واژه های مختلفی که در حوزه بازیابی اطلاعات موسیقی با آن سروکار داریم آشنا شویم. بنابراین در ادامه به تعریف مفاهیم مختلف در این حوزه می‌پردازیم.

---

<sup>26</sup> <http://www.shazem.com/>

<sup>27</sup> <http://www.nayio.com/>

<sup>28</sup> Audio

## ۲-۲ تعاریف:

### ۲-۲-۱ بازیابی:

در روانشناسی "بازیابی" به فرآیند فراخوانی اطلاعات ذخیره شده بر روی یک حافظه جانبی اطلاق می‌شود، و یا در کاربردهای مربوط به شبکه و رایانه منظور از "بازیابی" یافتن اطلاعات و ارسال آن به عنوان خروجی برای کاربر مورد نظر است.

### ۲-۲-۲ بازیابی اطلاعات:

"بازیابی اطلاعات" دانش جستجو برای اسناد، اطلاعات موجود در اسناد و یا متا-دیتای مربوط به اسناد می‌باشد. در عمل اگرچه بین "بازیابی دیتا"، "بازیابی اسناد"، "بازیابی اطلاعات" و "بازیابی متن" همپوشانی و اشتراکاتی وجود دارد اما هر یک دارای تئوری، ساختار و فناوری خاص خود می‌باشد. "بازیابی اطلاعات" یک گرایش میان‌رشته‌ای است که بر پایه‌های علمی همچون علم رایانه، ریاضیات، علم کتابداری، علوم اطلاعاتی، معماری اطلاعات، روانشناسی شناختی، زبان‌شناسی و آمار بنا شده است.

از سیستم‌های خودکار بازیابی اطلاعات برای کاهش آنچه که آنرا بار اضافی اطلاعات<sup>۲۹</sup> می‌نامند استفاده می‌شود. بسیاری از دانشگاه‌ها و کتابخانه‌های عمومی از سیستم‌های بازیابی اطلاعات برای دستیابی به کتابها، مجلات و دیگر اسناد استفاده می‌کنند. موتورهای جستجو گر وب از قبیل موتور جستجو گر گوگل و یاهو از جمله موارد کاربرد بازیابی اطلاعات می‌باشند.

---

<sup>29</sup> [information overload](#)

فرآیند بازیابی اطلاعات از جایی شروع می‌شود که یک کاربر پرسشی<sup>۳۰</sup> را به سیستم اعمال می‌کند. پرسش‌ها<sup>۳۱</sup> عباراتی تفصیلی در مورد اطلاعات مورد نیاز می‌باشند، به عنوان مثال رشته کلماتی که در هنگام جستجوی یک مطلب در جایگاه مورد نظر در موتور جستجو گر وب وارد می‌شود. پرسشهای کاربران، بسته به کاربرد اطلاعات مورد نظر(متن، تصویر و یا ویدئو) با اطلاعات موجود در دیتا بیس تطبیق داده می‌شود.

## ۲-۲-۳ بازیابی مفهومی اطلاعات موسیقی:

"بازیابی مفهومی اطلاعات موسیقی"<sup>۳۲</sup> یا به طور خلاصه "بازیابی اطلاعات موسیقی"<sup>۳۳</sup> یک علم میان‌رشته ایست که به بازیابی اطلاعات از موسیقی می‌پردازد و شامل مباحث زیر می‌شود:

- روشهای محاسباتی برای کلاسه بندی، خوشه بندی و مدل کردن
- استخراج ویژگی موسیقایی از موسیقی تک‌صدایی و چندصدایی<sup>۳۴</sup>
- تطبیق الگو و مباحث تشابه<sup>۳۵</sup>
- جداسازی<sup>۳۶</sup>
- و ...

---

<sup>30</sup> query

<sup>31</sup> Queries

<sup>32</sup> Content-Based Music Information Retrieval

<sup>33</sup> MIR

<sup>34</sup> mono and polyphonic music

<sup>35</sup> Similarity and pattern matching

<sup>36</sup> segmentation

بازیابی اطلاعات موسیقی جزء علوم نوین و بسیار فعال در عرصه فعالیت‌های تحقیقاتی به حساب می‌آید. در بازیابی اطلاعات موسیقی، رشته های علمی مختلفی از قبیل موسیقی‌شناسی، پردازش سیگنال، سایکواکوستیک<sup>۳۷</sup>، علوم ارتباطات، علوم رایانه و آمار به هم همگرا شده و روشهایی برای آنالیز و تحلیل موسیقی در فرم دیجیتالی آن ارائه می‌دهند.

## ۲-۲-۳-۱ موارد استفاده:

مشکلی که در آنالیز سیگنال موسیقی در مقایسه با سیگنال گفتار وجود دارد این است که به سختی می‌توان سیگنال موسیقی را با کلمات توصیف کرد. به همین دلیل، بازیابی اطلاعات موسیقی بر پایه آنالیز هوشمند و خودکار موسیقی، با هدف دستیابی آسان به موسیقی و یا اطلاعات مربوط به آن بنا شده است. برای نائل شدن به این هدف، بیشتر تحقیقات مربوط به بازیابی اطلاعات، بر شرح و ارزیابی روشهای موجود آنالیز خودکار موسیقی، متمرکز می‌شوند. در جدول ۱ برخی از موارد کاربرد بازیابی اطلاعات موسیقی به طور خلاصه آورده می‌شود. در صنعت ضبط و پخش موسیقی، شرکتها از سیستم های مبتنی بر آنالیز مفاهیم موسیقی مثل ملودی، ریتم، شعر و آکورد برای بررسی اثرات بالقوه یک کار در بازار موسیقی استفاده می‌کنند. از آنجا که جنبه ها و کاربردهای مختلفی برای بازیابی اطلاعات موسیقی وجود دارد خاطر نشان می‌کنیم که مورد نظر ما تنها آنالیز و بازیابی مفهومی موسیقی، به خصوص روش هایی است که بواسطه آنها اطلاعات مهمی درباره موسیقی از درون سیگنال موسیقی، به طور خودکار<sup>۳۸</sup> استخراج و جهت استفاده در سیستمهای بازیابی اطلاعات پردازش می‌شوند. جزئیات بیشتر در منابع [۸و۷] در دسترس می‌باشد.

<sup>37</sup> Psychoacoustics

<sup>38</sup> automatic

در مواجهه با موارد استفاده بازیابی اطلاعات موسیقی سه رویکرد کلی و متفاوت وجود دارد که هر کدام در جایگاه خاص خودش کاربرد دارد. رویکرد اول بر مبنای متا-دیتا پایه ریزی شده است؛ اطلاعاتی که به صورت کدگذاری شده موجود بوده و مانند متن قابل جستجو می‌باشد. رویکرد دوم بر پایه بازیابی مفاهیم موسیقی از قبیل ملودی، ریتم و... می‌باشد و به آنالیز ساختار و مفهوم قطعات موسیقی می‌پردازد. رویکرد سوم بر اساس استفاده از اطلاعات موجود در سیگنال صوتی یا به بیان دیگر استفاده از ویژگی‌های سطح پایین صوتی استوار است. در ادامه جنبه‌های مختلف این سه رویکرد به طور خلاصه و به اجمال معرفی می‌گردد.

جدول ۱-۲: (کاربردهای بازیابی اطلاعات موسیقی - برگرفته از [۶])

موارد استفاده	توضیحات
هویت موسیقی	تشخیص یک دیسک فشرده، تهیه متا-دیتا برای یک قطعه موسیقی نامعلوم، بازیابی اطلاعات دریافتی از روی یک تلفن همراه <sup>۳۹</sup>
کپی رایت	نظارت بر رعایت حقوق کپی رایت برای آثار منتشر شده.
ملودی (لحن)	یافتن کارهایی که شامل یک قطعه ملودیک (لحنی) خاص می‌شوند.
عناوین یا کارهای یکسان	بازیابی اجراهای مختلفی که از یک قطعه می‌تواند وجود داشته باشد.
شبیه بودن	یافتن آثاری که به یک اثر هنری شبیه‌اند.
انطباق اجرا	تطبیق یک اثر با اثر دیگر و بررسی میزان شباهت آنها بدون در نظر گرفتن سرعت یا تکرار جملات.
آهنگساز	یافتن کارهای مختلفی که از یک آهنگساز موجود است.
حالت	یافتن موسیقی بر مبنای حالت قطعه: شاد، غمگین، ترسناک، انرژی دهنده و ...

سبک یافتن موسیقی بر اساس سبک های موجود موسیقی از قبیل: جاز، راک، سنتی، پاپ و ...  
ساز یافتن قطعاتی که توسط یک ساز خاص نواخته شده است، به عنوان مثال: تار، سنتور، ویولن،  
گیتار، پیانو و ...

## ۲-۲-۳-۲ رویکردها:

## ۲-۲-۳-۱ متا-دیتا:

در یک سیستم بازیابی اطلاعات، متا-دیتا حکم درایور و راه انداز را برای ورودی سیستم داراست. از این روی، اقدامات گسترده‌ای در جهت فراهم کردن متا-دیتا برای مجموعه‌های موسیقی، هم برای کاربران شخصی<sup>۴۰</sup> و هم برای مراکز تجاری، صورت می‌پذیرد. بسیاری از علاقه مندان و شنوندگان در هنگام گوش دادن به موسیقی از متا-دیتا استفاده می‌کنند. در این زمینه از کاربرد متا-دیتا، می‌توان از آن به عنوان کاتالوگ نیز یاد کرد؛ چنانکه در تعریف متا-دیتا هم اشاره خواهیم کرد، زمینه کاربرد متا-دیتا است که آن را به طور دقیق تعریف می‌کند.

عموما متا-دیتا را دیتا درباره دیتا یا اطلاعات درباره اطلاعات می‌نامند. اگر چه این تعریف، تعریف دقیقی نمی‌باشد اما به خوبی بیانگر این مطلب است که متا-دیتا خود یک نوع دیتا است. نکته‌ای که حائز اهمیت است اینکه صرفاً با یک نگاه کلی نمی‌توان متا-دیتا را تشخیص داد و فهمید که دیتای مورد نظر دیتاست یا متا-دیتا. متا-دیتا دیتایست که برای توصیف دیتای دیگر مورد استفاده قرار می‌گیرد، به عبارتی، کاربرد یک دیتا است که آن را به متا-دیتا تغییر هویت می‌دهد. متا-دیتا به کاربران این امکان را می‌دهد که معنی و کیفیت دیتا را درک کنند، چرا که اطلاعاتی را درمورد محتویات یک موضوع مشخص

<sup>40</sup> end users

بدست می‌دهد. به عنوان مثال برای یک تصویر متا-دیتا است که اندازه تصویر، میزان رنگ، وضوح و اطلاعات مربوط به زمان ساخت تصویر را بدست می‌دهد، یا به عنوان مثالی دیگر، متا-دیتای یک فایل متنی حاوی اطلاعاتی در مورد طول متن، نویسنده متن، زمان نوشته شدن آن و ... می‌باشد.

با توجه به موارد استفاده متا-دیتا، الگوهای مختلفی برای آن ارائه شده و استفاده می‌شود. از معتبرترین استانداردهای موجود در زمینه مفاهیم موسیقی MPEG-7 و Dublin Core می‌باشند. MPEG-7 توسط گروه MPEG (Moving Picture Experts Group) ارائه شده و یک فرمت خاص برای شرح و توصیف منابع صوتی تصویری را معرفی می‌کند. در جدول زیر یک الگوی متا-دیتا برای فایل های موسیقی به عنوان نمونه آورده شده است.

جدول ۲-۲: برگرفته از [۹]

Name	Definition
Data Default Language	The language used for the descriptive metadata.
Title	This is the full title of the piece of material used as shown on the recording or sheet music (if used). Where only a part of a long work is used, the overall title of the work, followed by the details of the movement(s) used must be given.
Collection Name	Where appropriate, this is the name of the published collection (e. g. CD name, LP name, set name) from which the material used was taken.
Side Count	Where appropriate, this is the side number of the recording on which the track used appears.
Track Count	Where appropriate, this is the track number of the item used.
Label Name	This is the name of the publication label from which the material used was taken.
Prefix Number	This is the alphanumeric prefix to the catalogue number for the recording from which the material used was taken.
Number	This is the main catalogue number for the recording from which the material used was taken.
Signature Tune Indicator	This indicates that the music was used as the signature tune of the item/program in which it is used.
External Scheme Exchange Set	An element to allow the communication of any additional information using external references such as BBC's music origin codes for individual pieces of

TV Background Indicator	<p>music.</p> <p>Music used in a television program is categorized as either BACKGROUND or VISUAL. Background music is music added solely for the benefit of the television audience and, thus, forms no intrinsic part of the program. VISUAL music forms part of the program in as much as it appears to be, or is, audible to those taking part in the program. Examples of visual music include: music sung or played in vision; music which characters in a play can hear, even if it comes from a performer or player not in vision. All actuality music is visual. A value of TRUE would indicate that the music is BACKGROUND music; a value of FALSE indicates VISUAL music.</p>
Identifier	The Identifier is for reporting the ISRC of music (or other) sourced from video.
Contribution	<p>Contributions required are Composer, Publisher, Arranger, and Performer(s).</p> <p>Composer: if not known then enter NFI as surname (No Further Information); if the item is a traditional piece with no known composer(s) enter TRAD as surname; composer includes lyricist.</p> <p>Publisher: if an item is unpublished or played from manuscript enter MS as surname; if there is no publisher enters NA as surname.</p> <p>Arranger: if there is no arranger enter NA as surname.</p> <p>Performer: if there are many performers list them in order of importance.</p>
Musical Instrument Type	Identifies a musical instrument by its name or code
Music Piece Duration	Indicates the duration of the music (or other) item. If the same piece of music is used more than once then each use of the piece should be reported as a separate duration.

متا-دیتا می تواند به طرق زیر پیکر بندی شود:

- در صفحه وب تعبیه شود
- به عنوان یک فایل HTML جداگانه به فایل منبع لینک شود
- به صورت یک فایل دیتابیس جداگانه به منبع لینک شود
- و یا اینکه به طور مستقیم در دیتا بیس ذخیره شود.



یکی از کارهای رایج بازیابی اطلاعات موسیقی یافتن متا-دیتا در اینترنت برای موسیقی‌های دیجیتال است که از روی دیسک‌های فشرده به روی رایانه ذخیره یا اصطلاحاً ریپ<sup>۴۱</sup> شده‌اند. این کار اساس فعالیت وب سایت‌هایی از قبیل [musicbrainz.org](http://musicbrainz.org) و [gracenote.com](http://gracenote.com) را تشکیل می‌دهد. در متا-دیتایی که اینگونه سرویسها ارائه می‌دهند هم مشخصات عینی درمورد قطعه موسیقی مورد نظر، مثل نام هنرپیشه و آلبوم، سال انتشار، عنوان قطعه و طول آن، و هم مفاهیم موضوعی از قبیل حالت قطعه، میزان هیجانی بودن آن، کلاس و سبک کار، موجود است. به خاطر پیچیدگی های موجود در تهیه متا-دیتا، به خصوص برای مفاهیم موضوعی، متا-دیتا قادر به پاسخ گویی به تمام نیازهای موجود در زمینه بازیابی اطلاعات موسیقی نیست. سیستم های تجاری رایج به صورت قابل توجهی به متا-دیتا وابسته‌اند، اما قادر نیستند که ابزار کافی را برای جستجو در اختیار کاربرانی قرار دهند که دانش کافی درمورد موضوع یا نحوه جستجو ندارند. این شکاف فرصت مناسبی است برای استفاده از روش های مبتنی بر مفاهیم موسیقی و این نوید را می‌دهد که با کامل کردن روشهای مبتنی بر استفاده از متا-دیتا، این امکان برای کاربران فراهم شود که طی یک فرایند هوشمند، بدون نیاز به اطلاعات زیاد و به شکل کاملاً خودکار، به موسیقی مورد علاقه خود دست یابند.

اینجاست که بحث اصلی ما که اساس تعریف این پایان نامه نیز می‌باشد مطرح می شود؛ بازیابی اطلاعات موسیقی بر مبنای مفاهیم موجود در آن<sup>۴۲</sup>.

در ادامه به بحث تشریح مفهومی موسیقی پرداخته و حوزه های مربوط به آن را معرفی می‌کنیم. هرچند ذکر این نکته لازم است که استفاده از روشهای آنالیز مفهومی موسیقی در بازیابی اطلاعات موسیقی تنها برای تقویت روشهای مبتنی بر متا-دیتا است و نه به عنوان یک جایگزین برای آنها.

---

<sup>41</sup> rip

<sup>42</sup> Content-Bases Music Information Retrieval

## ۲-۲-۳-۲-۲ استخراج ویژگی های سطح بالای موسیقی<sup>۴۳</sup>:

اطلاعات سطح بالای موسیقی شامل آن دانش و درکی می شود که یک شنونده ماهر و آشنا به موسیقی، نسبت به یک قطعه موسیقی، از آن برخوردار است. یکی از روش های موجود در بازیابی اطلاعات موسیقی برای توصیف محتویات یک قطعه، مستقیماً استفاده از مفاهیم موسیقایی از قبیل ملودی (لحن)، هارمونی (هماهنگی) و یا ریتم است. به عنوان مثال سیستم هایی وجود دارند که با گرفتن اطلاعاتی که به منزله ملودی قطعه مورد سؤال است (به صورت زمزمه یا نواختن ملودی با دهان) به جستجوی اصل آهنگ می پردازند. آنچه که کار استخراج ویژگی های سطح بالای موسیقی را بسیار دشوار می کند پیچیدگی های ناشی از مفاهیم انتزاعی و غیر قابل لمس موسیقی است. بنابراین استخراج ویژگی های سطح بالای موسیقی، یک هدف غایی در بازیابی اطلاعات موسیقی و موضوع بسیاری از مقالات در این زمینه است. برخی از ویژگی های رایج سطح بالای موسیقی در جدول ۲-۳ آورده شده اند. هدف از این چنین کاربردهایی تبدیل موسیقی به فرمی است که بتواند مفاهیم موسیقی را به شکلی قابل درک برای ماشین ارائه کرده و در سیستم های جستجو و بازیابی اطلاعات مورد استفاده قرار گیرد.

با توجه به در دست بودن منابع قابل توجهی از قطعات موسیقی ضبط شده که می تواند به عنوان ورودی برای سیستم های سطح بالای تشریح موسیقی مورد استفاده قرار گیرد، عموماً استخراج ویژگی سطح بالا از روی محتویات ویژگی های سطح پایین صورت می پذیرد که در بخش بعد به توضیح آنها می پردازیم.

---

<sup>43</sup> High\_Level Features

جدول ۲-۳: ویژگی های سطح بالای موسیقی و توضیح آنها (برگرفته از [۶])

<b>High-level Description</b>	<b>Data Source</b>	<b>Task Description</b>
Timbre	Audio	Instrument Recognition Percussive, Pitched, Ensemble Recognition
Melody / Bass	Audio / Symbolic	Melody-line extraction Bass-line extraction
Rhythm	Audio	Onset detection Meter identification Meter alignment (bars) Beat (tactus) tracking Tempo tracking Average tempo
Pitch	Audio	Single fundamental freq. Multiple fundamental freq.
Harmony	Audio / Symbolic	Chord label extraction Bass-line extraction
Key	Audio / Symbolic	Modulation tracking Pitch spelling
Structure	Audio / Symbolic	Verse / chorus extraction Repeat extraction
Lyrics	Audio	Singing detection, lyrics-identification, word recognition
Non-Western music	Audio	Micro-tonal tuning systems Non-Western canon of concepts

## ۲-۲-۳-۲-۳ استخراج ویژگی های سطح پایین<sup>۴۴</sup>:

یکی دیگر از تکنیک های رایج برای آنالیز مفهومی موسیقی، استفاده از اطلاعات موجود در سیگنال صوتی است. ویژگی های سطح پایین، مقیاس هایی برای سیگنال صوتی اند که شامل اطلاعاتی در مورد یک قطعه موسیقی یا اجرای خاص می باشند. این ویژگی ها، حاوی اطلاعاتی از سیگنال صوتی اند که در ظاهر ارتباط خاصی با موسیقی و جنبه های مورد بحث ندارند و به خودی خود نمی توانند چیز زیادی در مورد آن بگویند. به بیان دیگر بر خلاف ویژگی های سطح بالا، این ویژگی ها آنچنان برای کاربران قابل فهم نیستند، هر چند که استخراج آنها از سیگنال صوتی به مراتب آسان تر از ویژگی های سطح بالاست. بنابراین، از آنجا که اندازه گیری دقیق و مستقیم یک جنبه خاص از موسیقی (ویژگی های سطح بالا) بسیار دشوار است، شاهد یک نوع همزیستی یا هم گونی<sup>۴۵</sup> در استفاده از ویژگی های سطح بالا و ویژگی های سطح پایین برای بدست آوردن جنبه های مختلف موسیقی خواهیم بود.

در زیر برخی از ویژگی های سطح پایین در آنالیز مفاهیم موسیقی آورده می شوند:

- (۱) طیف دامنه
- (۲) طیف Constant-Q
- (۳) پروفایل پیچ (کروما گرام)
- (۴) تشخیص نقطه آغاز<sup>۴۶</sup>
- (۵) ضرائب مل یا کپستروم
- (۶) تغییرات پی در پی طیفی<sup>۴۷</sup>

---

<sup>44</sup> Low-Level Audio Features

<sup>45</sup> tradeoff

<sup>46</sup> Onset detection

<sup>47</sup> Spectral Flux

(۷) سرعت سنج

(۸) متر سنج

همانطور که قبلا نیز اشاره کردیم ویژگی های سطح پایین به خودی خود قادر به دادن اطلاعات زیادی در مورد موسیقی نیستند، چرا که در آنها اطلاعات در مقیاس های زمانی بسیار جزئی کد شده و حالت کلی و جنبه عمومی خود را از دست می دهد. یک کار معمول برای استفاده از این ویژگی ها در بازیابی اطلاعات موسیقی این است که فریم های صوتی آنها به یکی از صورت های نمایش دسته جمعی درآیند. از جمله این تکنیک ها استفاده از مدل مخفی مارکوف<sup>۴۸</sup> و یا Gaussian Mixture Model است که قادرند در اندازه گیری میزان تشابه مورد استفاده قرار گیرند.

برای جزئیات بیشتر در مورد بازیابی مفهومی اطلاعات موسیقی و انواع رویکردهای آن به منابع [۶، ۸ و ۱۰] مراجعه شود.

## ۲-۳ بازیابی اطلاعات موسیقی، موسیقی غیر غربی و چالشهای روبرو:

در حال حاضر بیشتر کاربردهای بازیابی اطلاعات موسیقی به موسیقی پاپ و کلاسیک غربی<sup>۴۹</sup> محدود می شود. در حالی که با وجود تنوع فرهنگ ها و انواع گوناگون و غنی موسیقی، ارائه روشهایی برای بازیابی اطلاعات انواع دیگر موسیقی (موسیقی غیر غربی) و برآورده کردن نیازهای کاربران امری ضروری و غیرقابل انکار است. اما صرفا با اعمال روشهای مورد استفاده در بازیابی اطلاعات موسیقی غربی به موسیقی غیر غربی نمی توان به نتایج مطلوب جهت ارائه خدمات به کاربران دست یافت. تفاوت های اساسی که در

---

<sup>48</sup> HMM

<sup>49</sup> Western Music

ساختارهای موسیقایی موسیقی غیر غربی و غربی وجود دارد به خوبی گویای این مسئله است. به عنوان مثال در موسیقی غربی، پیچ (نغمه) از اهمیت بسیار بالایی برخوردار است، در حالی که در موسیقی ایرانی (به عنوان نمونه ای از موسیقی غیر غربی و شرقی آن) آنچه که اهمیت دارد نه نغمه‌ها، بلکه فواصل تشکیل‌دهنده موسیقی ایرانی است. ضمن اینکه اساساً موسیقی ایرانی و نیز بسیاری دیگر از انواع موسیقی غیرغربی، از جمله موسیقی هندی و یا آفریقایی، تک‌صدایی<sup>۵۰</sup> بوده و به کل دارای ساختار و توپولوژی متفاوتی در مقایسه با موسیقی غربی که یک موسیقی چندصدایی<sup>۵۱</sup> است می‌باشند. از جمله مشکلات دیگری که در بازیابی اطلاعات موسیقی غیر غربی با آن روبرو هستیم کمبود متا-دیتا و اختلاف بیش از حد متا-دیتاهای موجود است. مشکل اساسی دیگر نبود یک دیتا بیس کامل و البته دیجیتال شده برای استخراج ویژگی و اطلاعات مربوط به متا-دیتا است. تنوع در نوع ضبط آرشیه‌های موجود، و دشواری‌هایی که جهت دستیابی به آنها وجود دارد همگی لزوم نیاز به یک دیتا بیس واحد و دیجیتال شده را بیان می‌دارد. همچنین کمبود یک تئوری منسجم و قابل اطمینان جهت پایه ریزی روشهای آنالیز سیگنال از مهمترین مسائلی است که در آنالیز موسیقی ایرانی با آن روبرو هستیم [۵].

بنابراین موسیقی غیر غربی الگوریتم‌های خاص خود را در بازیابی اطلاعات موسیقی می‌طلبد چنانکه بایستی در مواجهه با آن، ملاحظات خاص این نوع موسیقی را باید در نظر گرفت.

در بررسی خواص تشابه در کلاسه بندی موسیقی ایرانی، همگی این ملاحظات در نظر گرفته شده و راه کارهای مناسب جهت رفع آنها پیشنهاد شده است. در فصل مربوط به تئوری موسیقی جنبه‌های مختلف موسیقی ایرانی بررسی شده و زمینه لازم جهت اعمال روشهای بازیابی اطلاعات فراهم آمده است.

---

<sup>50</sup> monophonic

<sup>51</sup> polyphonic

## فصل سوم:

### تبدیل با Q ثابت<sup>۵۲</sup>

#### ۳-۱ مقدمه ای بر آنالیز سیگنالهای موسیقی در حوزه فرکانس:

تبدیل CQT یکی از پرکاربردترین ابزار لازم جهت آنالیز سیگنالهای موسیقایی است که می‌تواند در سنجش فرکانس اصلی نغمه‌های موسیقی و آنالیز آنها به خوبی استفاده شود. این تبدیل حتی نسبت به تبدیل فوریه نیز دارای مزایایی است. در واقع از آنجایی که تبدیل فوریه اجزای فرکانسی را تحت یک الگوی خطی با یک رزولوشن فرکانسی خاص (پنجره‌های با عرض ثابت) محاسبه می‌کند بنابراین طیف حاصل دارای رزولوشن فرکانسی پایین در فرکانسهای موسیقایی کم و رزولوشن فرکانسی زیادتر از حد معمول در فرکانسهای موسیقایی بالا خواهد شد.

به عنوان مثال اگر فرض کنیم که فرکانس نمونه‌برداری ۲۲۰۵۰ هرتز و اندازه پنجره ۵۱۲ نمونه باشد آنگاه رزولوشن فرکانسی  $43\text{Hz} = 22050/512$  می‌شود. برای کمترین فرکانس ساز ویولن (۱۹۶ هرتز) و نغمه بعد از آن تنها ۱۲ هرتز فاصله فرکانسی وجود دارد بنابراین، با رزولوشن فرکانسی ۴۳ هرتز تمامی اطلاعات مربوط به ۳ تا ۴ نغمه مجاور در یک بین فرکانسی گنجانده می‌شود. به طور مشابه برای بالاترین نغمه موجود در ساز پیانو (۴۱۸۶ هرتز) و نغمه قبل از آن (۳۹۵۱ هرتز) ۲۳۵ هرتز فاصله موسیقایی وجود

---

<sup>52</sup> Constant-Q-Transform

دارد، یعنی ۶ بین فرکانسی اطلاعات تنها یک نغمه موسیقایی را دربر دارند. بنابراین واضح است که تبدیل فوریه از کارایی چندانی در کاربردهای موسیقایی برخوردار نیست.

در آنالیز سیگنالهای موسیقایی آنچه که لازم است اطلاعات مربوط به اجزای طیفی است که می‌تواند توسط یک ساز موسیقایی با وسعت فرکانسی کامل تولید شود. برای رسیدن به چنین اطلاعاتی به بررسی روابط نغمه‌ها در ساز پیانو می‌پردازیم.

در ساز پیانو نغمه‌های تعدیل شده از رابطه ۱-۳ بدست می‌آیند:

$$f_k = f_{\min} \cdot (2^{(k/12)}) \quad (1-3)$$

که در آن  $f_{\min}$  فرکانس کمینه و  $f_k$  فرکانس نغمه  $k$ ام است. اگر برای هر نغمه در موسیقی غربی ۲ بین فرکانسی در نظر بگیریم (با توجه به ۱۲ نغمه موجود در موسیقی غربی) رابطه بالا به صورت زیر تغییر می‌کند:

$$f_k = f_{\min} \cdot (2^{(k/24)}) \quad (2-3)$$

از آنجا که رزولوشن فرکانسی معادل با اختلاف فرکانسی نغمه‌های مجاور است بنابراین:

$$\Delta f_k = f_k - f_{k-1} = f_k \cdot (2^{(1/24)} - 1) \quad (3-3)$$

$$Q = f_k / \Delta f_k = \frac{1}{(2^{(1/24)} - 1)} \cong 34 \quad (4-3)$$

بنابراین نسبت فرکانس به رزولوشن،  $Q$ ، برای هر نغمه حدود ۳۴ می‌شود. به عبارت دیگر تبدیل مورد نظر معادل می‌شود با یک فیلتر بانک با  $Q$  ثابت ۱/۲۴ اکتاوی.

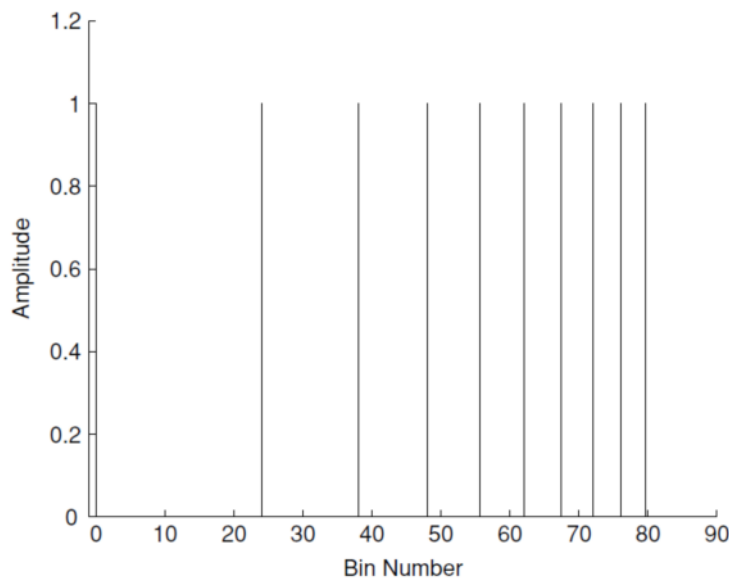
اگر از طرفین رابطه ۱-۳ لگاریتم بگیریم خواهیم داشت:



$$\log(f_k) = \frac{\log 2}{12} k + \log(f_{\min}) \quad (5-3)$$

که در آن فرکانس لگاریتمی نغمه‌ها با شماره نغمه رابطه خطی و مستقیم دارد. نکته‌ی بسیار مهم در این رابطه این است که برای اصواتی که از اجزای هارمونیکی تشکیل می‌شوند، فاصله گذاری غیر یکسان اجزاء، که به صورت تابعی از شماره بین نشان داده می‌شوند، مستقل از فرکانس اصلی  $f_0$  است. این الگوی فاصله گذاری در شکل ۱ نشان داده شده است.

موقعیت بینها بر روی محور افقی (که به صورت لگاریتمی تعیین می‌شود و بیانگر مفهوم constant Q است) برای هر سری از اجزای فرکانسی که دارای روابط هارمونیکی هستند یکسان است. به عبارت دیگر برای اصوات دارای اجزای هارمونیکی، الگوی فوق یکسان است. به دیگر سخن، اگر از سیگنال صوتی که رابطه بین اجزای فرکانسی آن ثابت است<sup>۵۳</sup> تبدیل فوریه بگیریم، الگوی شکل ۳-۱ را خواهیم داشت.



شکل ۳-۱: الگوی تبدیل فوریه برای ۱۰ جزء فرکانسی هارمونیک برحسب فرکانس لگاریتمی (شماره بینهای ثابت Q) با فاصله گذاری ۲۴ بین در هر اکتاو.

<sup>53</sup> constant Q

### ۲-۳ محاسبه تبدیل CQT :

برای محاسبه تبدیل constant Q به صورت زیر عمل می‌کنیم. همانگونه که در روابط ۳-۳ و ۴-۳ عنوان شد رزولوشن یا پهنای باند به صورت  $\Delta f_k$  و نسبت فرکانس به پهنای باند به صورت  $Q$  تعریف می‌شود که مجدداً در زیر می‌آید:

$$\Delta f_k = f_k - f_{k-1} = f_k \cdot (2^{(1/24)} - 1)$$

$$Q = f_k / \Delta f_k = \frac{1}{(2^{(1/24)} - 1)} \cong 34$$

در تبدیل فوریه پهنای باند معادل با تعداد نمونه‌هایی است که توسط عرض پنجره جدا می‌شود. بنابراین عرض پنجره برابر با تعداد نمونه‌هایی است که توسط پهنای باند جدا می‌شوند:

$$N[k] = \frac{f_s}{\Delta f_k} \quad (۶-۳)$$

پس اگر نسبت فرکانس به پهنای باند (constant-Q) ثابت باشد، عرض پنجره به طور معکوس با فرکانس بین تغییر می‌کند:

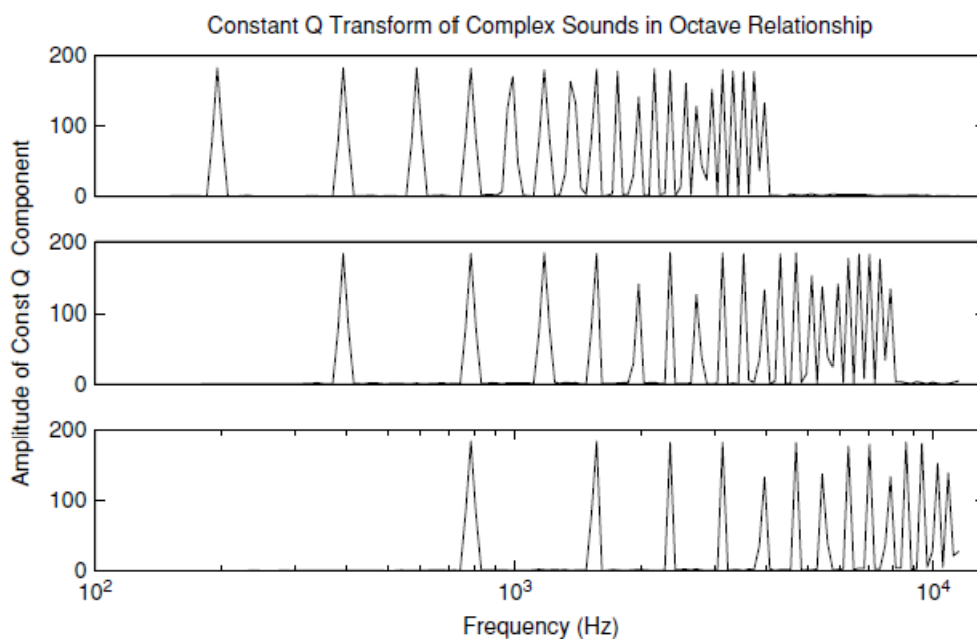
$$N[k] = \frac{f_s}{\Delta f_k} = \frac{f_s}{(f_k / Q)} = \frac{f_s \cdot Q}{f_k} \quad (۷-۳)$$

با جایگزینی پنجره متغیر در فرمول تبدیل فوریه و نرمالیزه کردن آن رابطه تبدیل constant-Q به صورت زیر بدست می‌آید:

$$X^{cq}[k] = \frac{1}{N[k]} = \sum_{n=0}^{N[k]-1} w[n, k] x[n] e^{-j2\pi kn/N[k]} \quad (۸-۳)$$

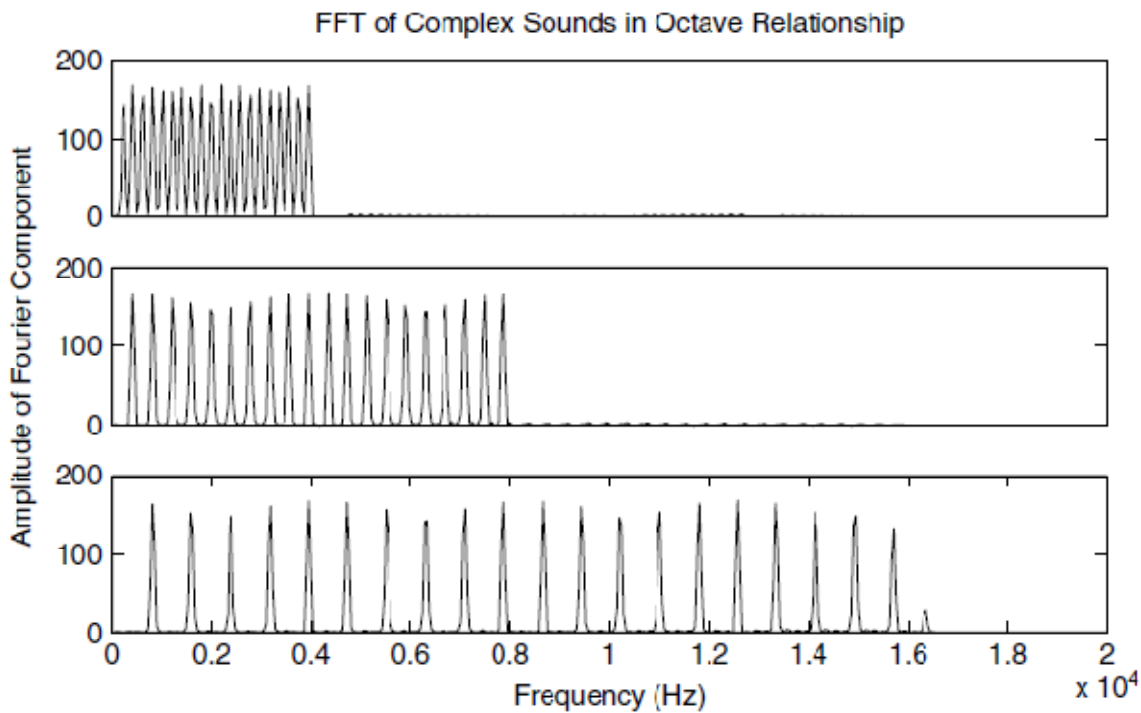
### ۳-۳ مقایسه تبدیل فوریه و تبدیل CQT :

همانگونه که در شکل ۲-۳ مشاهده می‌کنید برای سه نغمه مختلف الگوی فرکانسی ثابت است و تنها موقعیت نغمه‌ها تغییر کرده است. با توجه به اینکه هر سه الگو برای نغمه G بدست آمده‌اند قابل مشاهده است که تغییر فرکانس اصلی تغییری در الگوی فرکانسی نغمه‌ها نداشته است.



شکل ۲-۳: تبدیل Constant-Q برای سه نغمه یکسان (G) که در سه اکتاو مختلف نواخته شده‌اند.

در تبدیل فوریه طیف سیگنال بر حسب تغییرات خطی فرکانس نمایش داده شده و برای اجزای فرکانسی هارمونیک یک جداسازی ثابت ارائه می‌کند. این ویژگی برجسته اجزای فرکانسی تولید شده است که در آن هم نحوه جداسازی اجزای فرکانسی و هم موقعیت کلی الگوها با تغییر فرکانس اصلی تغییر می‌کند که در شکل ۳-۳ به وضوح قابل مشاهده است.



شکل ۳-۳: تبدیل فوریه برای سه نغمه (G) با نسبت های فرکانسی ۱، ۲ و ۳.

از طرف دیگر، نمایش بر حسب فرکانس لگاریتمی یک الگوی منحصر به فرد از فاصله گذاری اجزای طیفی هارمونیک به دست می‌دهد که به واسطه آن مشکل یافتن فرکانس اصلی در تشخیص این الگو خلاصه می‌شود.

در جدول ۱-۳ مقایسه بین پارامترهای مختلف تبدیل فوریه و تبدیل CQT آورده شده است.

Parameter	DFT	Constant-Q transform
Frequency $f_k$	$k \Delta f$ (Linear in $k$ )	$(2^{1/24})^k \cdot f_{\min}$ (Exponential in $k$ )
Window size	Constant = $N$	Variable = $N[k] = f_s Q / f_k$
Resolution $\Delta f$	Constant = $f_s / N$	Variable = $f_k / Q$
$\frac{f_k}{\Delta f}$	Variable = $k$	Constant = $Q$
Cycles in window	Variable = $k$	Constant = $Q$

جدول ۱-۳: مقایسه تبدیل فوریه با تبدیل CQT

همانطور که از جدول قابل مشاهده است رابطه اجزای فرکانسی در تبدیل فوریه خطی است در حالی که در تبدیل CQT اجزای فرکانسی با هم رابطه لگاریتمی دارند، که دقیقاً با موقعیت هندسی نغمات موسیقایی همخوانی دارد.

عرض پنجره در تبدیل فوریه ثابت است در حالی که در تبدیل CQT عرض متغیر پنجره ها با توجه به فرکانسی که تبدیل در آن محاسبه می‌شود موجب رزولوشن فرکانسی ثابت در گستره فرکانسی نغمات موسیقی می‌شود.

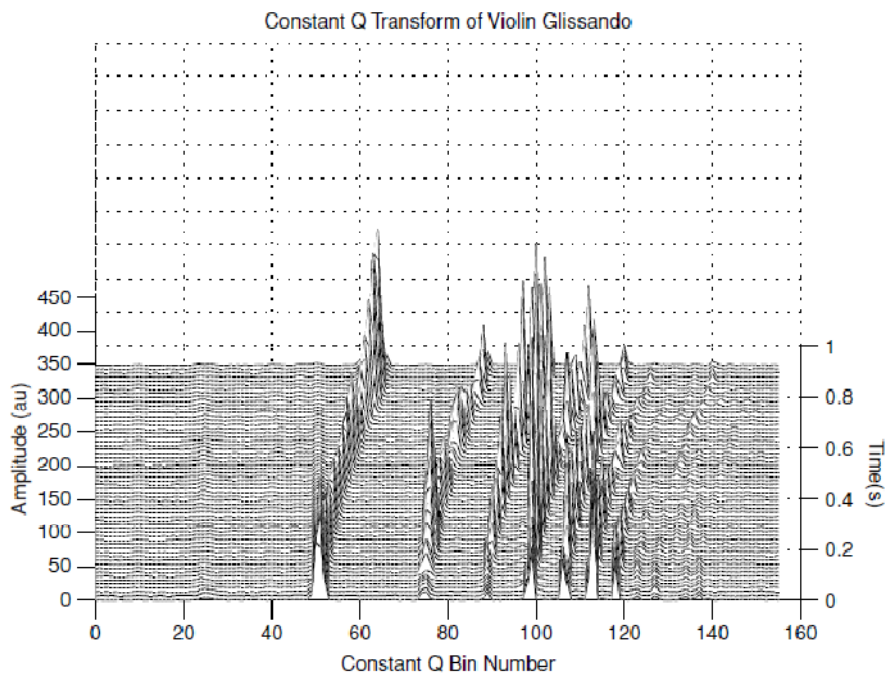
در شکل ۳-۴ نمای سه بعدی از تبدیل CQT برای نغمات موسیقی قابل مشاهده است. نظم خاصی که در جهت محور افقی (هم برای شکل الف و هم برای شکل ب) وجود دارد و در طول زمان نیز می‌توان آن را مشاهده کرد گویای یکسان بودن الگوی فرکانسی برای اصوات با اجزای هارمونیک است.

### ۳-۴ کاربردها:

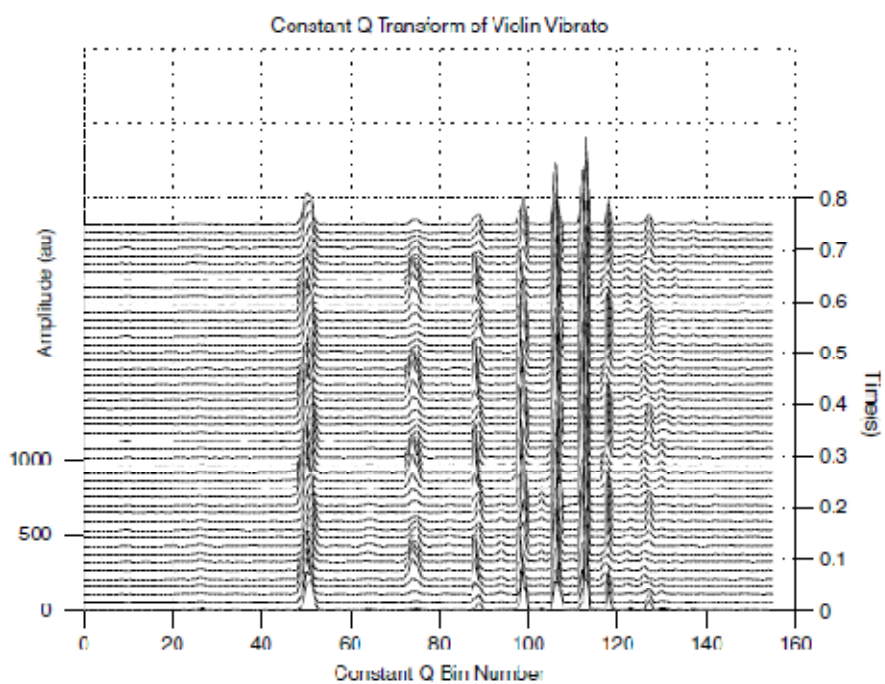
در مقالات مختلف استفاده‌های گوناگونی از این تبدیل شده که یکی از ساده‌ترین کاربردهای این تبدیل در بدست آوردن فرکانس اصلی نغمه‌هاست که در [۳۱] توسط خانم J. C. Brown ارائه شده است.

اساس این روش بر این استوار است که برای اصوات دارای اجزای فرکانسی هارمونیک، تبدیل فوریه با فاصله گذاری در حوزه فرکانس لگاریتمی، مستقل از فرکانس اصلی است.

در اینجا فرض می‌شود که فرکانسهای مجهول مضارب صحیحی از فرکانس کمینه ( $f_{\min}$ ) هستند.



الف



ب

شکل ۳-۴ تبدیل Constant-Q برای نغمات نواخته شده توسط ساز ویولن

بنابراین شماره بینهای تبدیل CQT به کمک رابطه ۲-۳ به صورت زیر محاسبه می‌شوند:

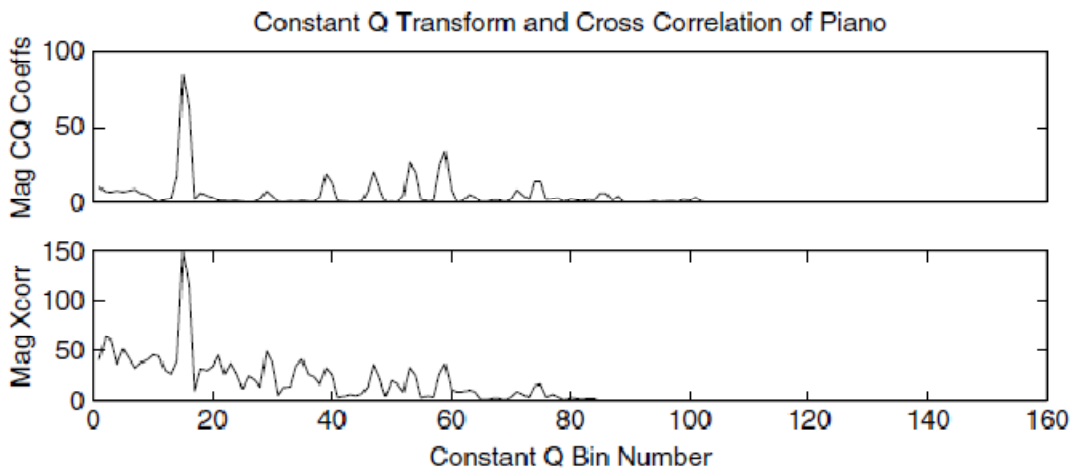
$$k = 24 \log_2 \left( \frac{f_k}{f_{\min}} \right) \cong 79.73 \log_{10} \left( \frac{f_k}{f_{\min}} \right) \quad (9-3)$$

که ۱۰ بین اول آن عبارتند از: ۰، ۲۴، ۳۸، ۴۸، ۵۵.۷، ۶۲، ۶۷.۴، ۷۲، ۷۶ و ۷۹.۷ البته در رزولوشن‌های بالاتر، این بینها روند می‌شوند. از آنجایی که این الگوی متشکل از ۱ ها برای اجزای فرکانسی هارمونیک یکسان است می‌توان همبستگی متقابل آن را با تبدیل هر سیگنال موسیقایی بدست آورد، ماکزیمم در فرکانس اصلی سیگنال اتفاق می‌افتد. پس از تعیین آن بین فرکانسی که موقعیت فرکانس اصلی را نشان می‌دهد، فرکانس اصلی از رابطه زیر محاسبه می‌شود:

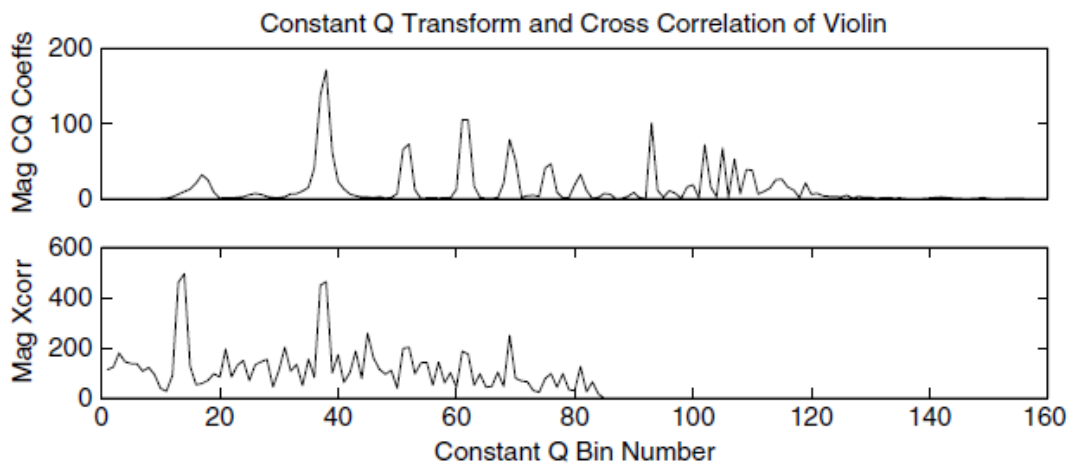
$$f_0 = (2^{1/24})^k f_{\min} \quad (10-3)$$

در شکل ۳-۵ طیف سیگنال و تابع همبستگی متقابل آن نشان داده شده است. در شکل الف ۳-۵ دامنه فرکانس اصلی ماکزیمم است در حالی که در شکل ب ۳-۵ ماکزیمم در فرکانس اصلی اتفاق نمی‌افتد، اما تابع همبستگی متقابل در هر دو حالت فرکانس اصلی را درست تشخیص داده است. که آن به دلیل عدم وابستگی این روش به فرکانس اصلی نغمه مورد نظر است.

کاربرد دیگری که برای این تبدیل میتوان مثال زد مربوط می‌شود به استفاده از این تبدیل در مراحل اولیه در یک الگوریتم تشخیص کلید که در فصل بعد به طور مفصل توضیح داده می‌شود. اما آنچه که انکار ناپذیر است کارایی بسیار بالای این CQT در کاربردهای مربوط به بازیابی اطلاعات موسیقایی است.



الف



ب

شکل ۳-۵: طیف سیگنال برای ۲ نغمه متفاوت به همراه تابع همبستگی متقابل آنها. الف: ماکزیمم در فرکانس اصلی ب:

ماکزیمم در فرکانسی غیر از فرکانس اصلی



## فصل چهارم:

### مروری بر برخی کارهای انجام شده در حوزه بازیابی اطلاعات

### موسیقایی جهت کلاسه بندی موسیقی.

#### ۴-۱ مقدمه:

یکی از ویژگی‌های پر کاربرد در کلاسه بندی موسیقی پیچ است. در کنار روشهای متعددی که برای بدست آوردن پیچ در سیگنالهای صوتی، اعم از سیگنال گفتار و سیگنال موسیقی وجود دارد [۱۲-۱۷]، روشهای بسیاری نیز جهت بدست آوردن پروفایل پیچ برای سیگنالهای موسیقی ارائه شده است [۱۸-۲۲]. در بدست آوردن پیچ نغمه ها آنچه که کار را دشوار می کند حضور همزمان نغمه‌هایی است که توسط مجموعه متنوعی از سازهای موسیقی نواخته می شوند. در کاربردهایی نظیر تشخیص مد و مقام (در موسیقی ایرانی و ترکی) یا تشخیص کلید (در موسیقی غربی) استفاده از پروفایل پیچ به جای پیچ، در عین حالی که نتیجه مطلوب را بدست می دهد، می تواند تا حدود زیادی از پیچیدگی های کار بکاهد. تفاوت اصلی این دو ویژگی در این است که پیچ تنها حاوی اطلاعات مربوط به فرکانس اصلی نغمه هاست، در حالی که پروفایل پیچ علاوه بر اطلاعات مربوط به فرکانس اصلی نغمه ها، اطلاعات هارمونیکهای آنها را نیز به دست می دهد.

در این فصل به طور موردی سه روش مختلف را که به کمک مفهوم پیچ به کلاسه بندی موسیقی در حوزه بازیابی اطلاعات موسیقایی می‌پردازند بررسی می‌کنیم.

در روش اول الگوریتمی برای استخراج ویژگی پروفایل پیچ ارائه می‌گردد که از آن در تشخیص کلید برای موسیقی غربی استفاده می‌شود.

روش دوم یکی از جدیدترین کارهایی است که در حوزه موسیقی غیر غربی منتشر شده و به تشخیص مقام در موسیقی ترکی می‌پردازد.

در بخش سوم روش ارائه شده توسط پیمان حیدریان برای آنالیز سیستم tuning در موسیقی ایرانی مورد بررسی قرار می‌گیرد.

## ۲-۴ روش اول:

### تشخیص کلید با استفاده از پروفایل پیچ دقیق

در [۱۸] روشی برای استخراج ویژگی پروفایل پیچ دقیق معرفی می‌شود که نسبت به کوک‌دررفتگی<sup>۵۴</sup> و اصوات نویزی ناشی از سازهای کوبه‌ای مقاوم بوده و برای کاربردهایی نظیر تشخیص کلید در موسیقی غربی مورد استفاده قرار می‌گیرد.

در تئوری موسیقی غربی کلید به یک الگوی<sup>۵۵</sup> فرکانسی اطلاق می‌شود که در آن سلسله نغمه‌های متوالی که اساس ساخت قطعه موسیقی را تشکیل می‌دهند به صورت پی‌درپی آورده می‌شوند. برای

---

<sup>54</sup> mistuning

<sup>55</sup> scale

موسیقی غربی ۲۴ الگوی فرکانسی (کلید) مختلف وجود دارد که همگی آنها زیر مجموعه دو الگوی اصلی ماژور و مینور<sup>۵۶</sup> می‌باشند. به عبارتی برای هر کدام از الگوهای اصلی مینور و ماژور ۱۲ کلیشه مختلف وجود دارد.

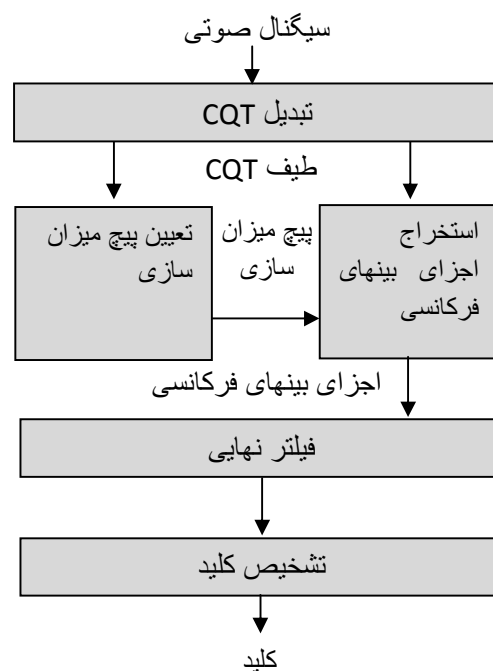
#### ۴-۲-۱ الگوریتم کلی:

الگوریتم مورد بحث از دو بخش اصلی تشکیل می‌شود:

بخش اول : استخراج پیچ

بخش دوم : تشخیص کلید

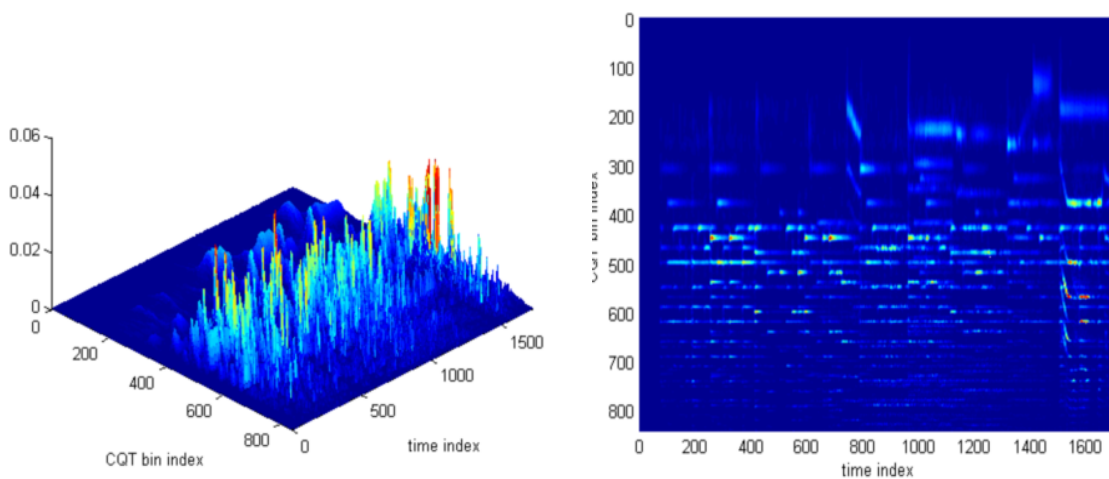
الگوریتم کلی روش در شکل ۴-۱ نشان داده شده است.



شکل ۴-۱: الگوریتم کلی تشخیص کلید برای موسیقی غربی، [۱۸].

<sup>56</sup> major and minor Scales

در ابتدا با اعمال تبدیل CQT بر سیگنال صوت طیف فرکانسی سیگنال موسیقی مورد نظر بدست می-آید. سیگنال صوتی با فرکانس ۴۴.۱KHz نمونه برداری شده و استپسایز برای تمامی اجزای فرکانسی ۵۱۲ نمونه است. محدوده فرکانسی در نظر گرفته شده در تبدیل CQT ۷ اکتاو است که از فرکانس کمینه ۲۷.۵HZ تا فرکانس بیشینه ۳۵۲۰ HZ (که همان محدوده صوتی ساز پیانو است) وسعت دارد. رزولوشن فرکانسی معادل  $b=120$  انتخاب شده که به منزله  $7 \times 120 = 840$  بین فرکانسی برای ۷ اکتاو موجود می‌باشد. طیف فرکانسی حاصل در شکل ۴-۲ بر حسب زمان نشان داده شده است.



شکل ۴-۲: اسپکتروگرام (طیف فرکانسی) سیگنال موسیقی از دو زاویه مختلف

در مرحله بعد برای استخراج اجزای فرکانسی نغمه‌ها عمل تعیین پیچ میزان سازی<sup>۵۷</sup> انجام می‌شود. در یک گروه موسیقی غربی همگی سازها بر اساس یک ساز مرجع (معمولاً ابوا) کوک می‌شوند. حال با توجه به اینکه آیا ساز مرجع بر اساس استانداردهای موجود برای نغمات موسیقی کوک شده یا نه، میزان

<sup>57</sup> tuning pitch

از کوک در رفتگی اتفاق می‌افتد. استاندارد<sup>58</sup> ISO 16 برای نت A فرکانس 440 Hz را تخصیص داده که به آن پیچ کنسرت<sup>59</sup> می‌گویند. پیچ میزان‌سازی بر اساس تخمین مقدار انحراف از پیچ استاندارد<sup>60</sup> تعیین می‌شود. از آنجایی که کمینه فاصله بین نغمه‌های موسیقی غربی یک نیم‌نت (100 سنت) است، میزان انحراف می‌تواند بین 50- تا 50+ سنت باشد. با انتخاب دقتی معادل با 10 سنت برای انحراف از پیچ استاندارد، پیچ میزان‌سازی می‌تواند عددی بین 1 تا 10 را اختیار کند.

برای بدست آوردن پیچ میزان‌سازی ابتدا پیکهای محلی در طیف فرکانسی جدا می‌شوند. اگر  $x^{cq}[t, k]$  اسپکتروگرام سیگنال مورد نظر باشد آنگاه:

$$\hat{x}^{cq}[t, k] = x^{cq}[t, k] \text{ if } x^{cq}[t, k] > x^{cq}[t, k-1] \& x^{cq}[t, k+1]$$

پیکهای محلی اسپکتروگرام را نشان می‌دهد. در قدم بعدی پیکها بر اساس اندازه فاصله نسبت به پیچ استاندارد دسته بندی می‌شوند.

$$\hat{E}[t, n] = \sum_{k=n+10m} \hat{x}^{cq}[t, k] \quad (1-4)$$

که  $n=[1,10]$  و  $m=[1,12*7]=[1,84]$ . برای  $n=1$  میزان انحراف از پیچ استاندارد صفر است. در قدم بعد موقعیت پیچ میزان‌سازی  $n$  با بیشینه انرژی تراکمی در هر پنجره زمانی به صورت زیر بدست می‌آید.

$$p[t] = \arg \max(\hat{E}[t, n]), n = [1,10] \& p = [1,10] \quad (2-4)$$

پس از اینکه  $p$  برای تمامی پنجره های زمانی تعیین شد، یک هیستوگرام پیچ میزان‌سازی  $H_p[p[t]]$  برای کل اسپکتروگرام تشکیل می‌شود که شامل 10 بین می‌باشد. آن بینی از  $H$  که بیشینه می‌شود بیانگر  $p$  است.

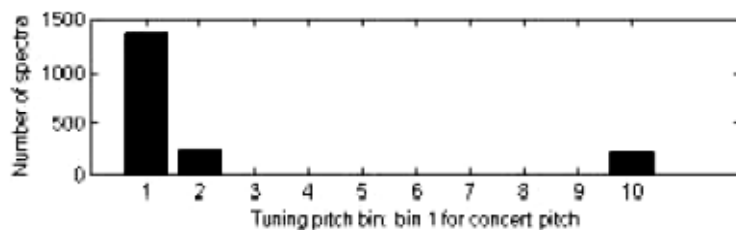
$$p_{tuning} = \arg \max(Hp[p]) \quad (3-4)$$

<sup>58</sup> International Organization for Standardization

<sup>59</sup> concert pitch

<sup>60</sup> concert pitch

که در آن  $P$  عددی بین ۱ تا ۱۰ است که میزان انحراف از پیچ استاندارد را تعیین می‌کند. به عنوان مثال  $p=10$  به منزله ۱۰ سنت پایین تر از پیچ استاندارد و  $p=2$  به منزله ۱۰ سنت بالاتر از پیچ استاندارد است. در مرحله بعدی با استفاده از پیچ میزان‌سازی اجزای فرکانسی نغمه‌ها از اسپکتروگرام سیگنال موسیقی استخراج می‌شود. در شکل ۳-۴ نمونه‌ای از هیستوگرام پیچ میزان‌سازی برای یک قطعه موسیقی ۲۰ ثانیه‌ای آورده شده است.

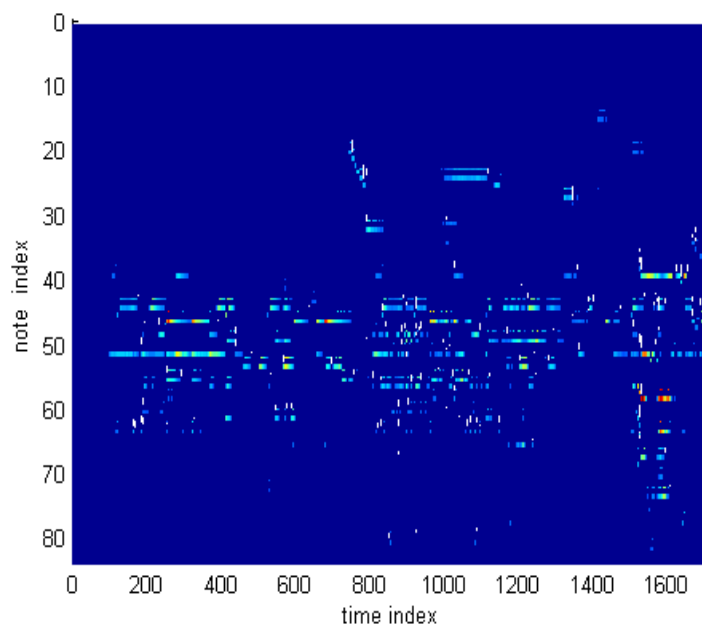


شکل ۳-۴: از هیستوگرام پیچ میزان‌سازی برای یک قطعه موسیقی ۲۰ ثانیه‌ای

حال به کمک پیچ میزان‌سازی بدست آمده می‌توان عمل جداسازی اجزای فرکانسی را انجام داد. با این فرض که تنها آن اجزای فرکانسی که با پیچ میزان‌سازی هم‌کوک هستند بایستی در پروفایل نهایی ظاهر شوند، اجزای فرکانسی اسپکتروگرام با توجه به میزان انحرافشان از پیچ میزان‌سازی به صورت نشان داده شده در زیر فیلتر می‌شوند.

$$X[t, m] = \max_{10(m-1)+p-2 \leq k \leq 10(m-1)+p+2} [x^{cq}[t, k]] \quad (4-4)$$

که در آن  $p \in [1, 10]$ ،  $t \in [1, T]$  و نهایتاً  $m \in [1, 84]$  می‌باشند. اجزای فرکانسی استخراج شده با فاصله نیم‌نت از هم قرار می‌گیرند. به عبارت دیگر زولوشن فرکانسی از ۱۰ بین برای هر نیم‌نت به یک بین کاهش یافته است. طیف حاصل در شکل ۴-۴ نشان داده شده است.



شکل ۴-۴: اجزای فرکانسی استخراج شده برای قطعه موسیقی قبلی.

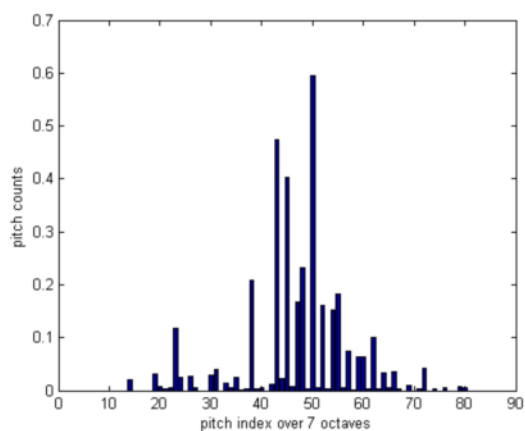
مرحله آخر شامل یک مرحله فیلترینگ<sup>۶۱</sup> و نهایتاً تولید پروفایل پیچ می‌شود. فیلتر هماهنگ در [۴۲] به طور جداگانه توضیح داده شده است. پروفایل نهایی از رابطه زیر محاسبه می‌شود.

$$p_a[n] = \sum_{t=1}^T \tilde{X}[t, n] \text{ for } n \in [1, N], \text{ where } N = 84, \text{ and } T \text{ is the time window length} \quad (۵-۴)$$

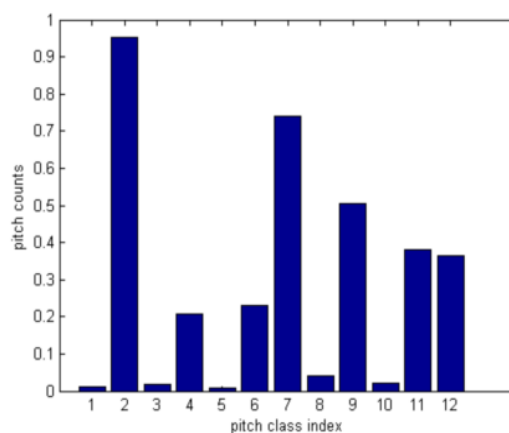
$$p_b[m] = \sum_{k=0}^{N/12-1} p_a[m + 12k], \text{ where } m \in [1, 12] \quad (۶-۴)$$

$p_a$  شامل مجموع بینهای فرکانسی برای هر پیچ  $n$  در تمامی پنجره های زمانی است. در  $p_b$  بینهای فرکانسی که در  $p_a$  در ۷ اکتاو گسترده شده بودند به یک اکتاو نگاشته می‌شوند. نتیجه در شکل ۴-۵ نشان داده شده است.

<sup>61</sup> consonance filtering



الف



ب

شکل ۴-۵: الف)  $P_a$ ، ب) پروفایل نهایی  $p_b$

شکل ب ۴-۵ پروفایل نهایی را نشان می‌دهد که همانطور که در ابتدای بحث نیز عنوان شد در کاربردهایی نظیر تشخیص کلید در موسیقی غربی استفاده می‌شود.

نتیجه استفاده از این ویژگی در تشخیص کلید موسیقی غربی در جدول ۴-۱ به شرح زیر آورده شده است:

	Classical pieces		Popular pieces	
	Exact	Exact and related	Exact	Exact and related
Pitch profile feature	53%	74%	83%	89%
Chroma feature	49%	68%	73%	78%

جدول ۴-۱: نتایج تشخیص کلید با استفاده از ویژگی پروفایل پیچ، [۱۸].

همانطور که در جدول فوق قابل مشاهده است استفاده از ویژگی پروفایل پیچ نتایج بهتری را نسبت به ویژگی chroma که در مقالات دیگر پیشنهاد شده به دست می‌دهد.



## بازیابی اطلاعات موسیقی ترکی با استفاده از هیستوگرام پیچ

در [۲۲] استفاده از هیستوگرام پیچ در آنالیز موسیقی غربی و غیرغربی بحث شده و مشکلات اعمال تکنیکهای غربی مبتنی بر پروفایل پیچ به موسیقی ترکی (به عنوان مثالی از موسیقی غیر غربی) بررسی شده است. برای دو کاربرد مهم تشخیص نغمه تونیک و تشخیص مقام نشان داده شده است که پروفایل پیچ های با رزولوشن بالا نتایج بهتری را برای موسیقی ترکی به دست می‌دهند.

در حوزی بازیابی اطلاعات موسیقایی برای اعمال روشهای موجود برای موسیقی غربی<sup>۶۲</sup> به موسیقی غیرغربی<sup>۶۳</sup> به دو مشکل اساسی بر می‌خوریم:

۱- تفاوت فواصل موسیقایی در موسیقی غربی و موسیقی غیرغربی

۲- خلا حاصل از نبود یک تئوری منسجم برای موسیقی غیرغربی

بنابر این یک پیشنهاد برای کار در حوزه بازیابی اطلاعات موسیقایی برای موسیقی غیرغربی نمایش صحیحی از فواصل در موسیقی مورد نظر است. در این راستا مولفین در این مقاله به نمایش فواصل<sup>۶۴</sup> برای موسیقی ترکی پرداخته‌اند و آنالیز خود را بر تشخیص نغمه تونیک و تشخیص مقام در موسیقی ترکی متمرکز کرده‌اند.

در [۲۲] از پروفایل پیچ<sup>۶۵</sup> بدین‌گونه استفاده شده که بر خلاف موسیقی غربی در اینجا پیچها به فضای ۱۲ بعدی فواصل نگاشته نمی‌شوند بلکه پیچها در یک فضای ۱۲۰ بعدی از فواصل موسیقایی نمایش داده

---

<sup>62</sup> western music

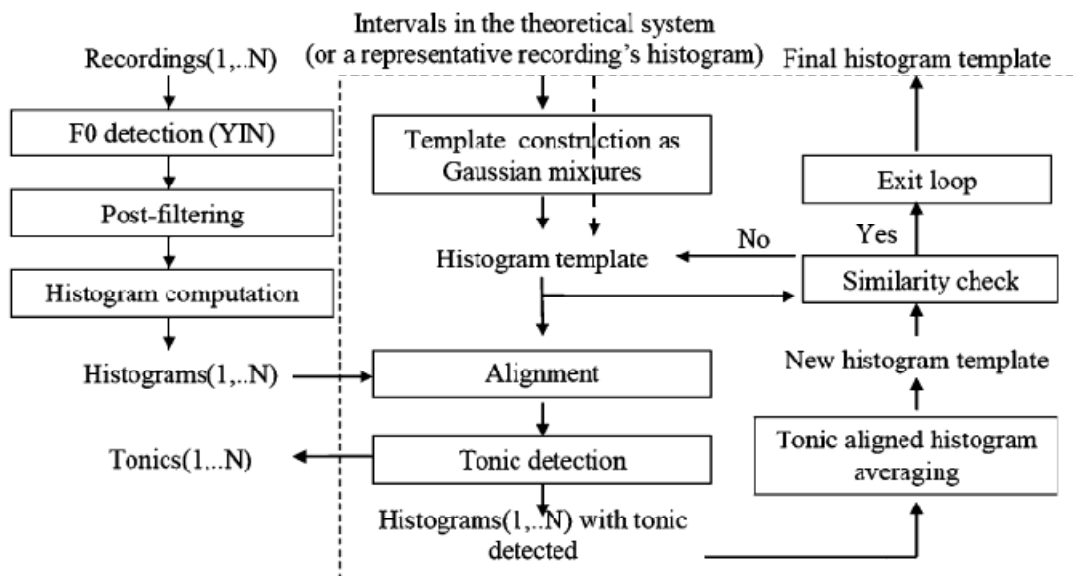
<sup>63</sup> non-western music

<sup>64</sup> pitch space

<sup>65</sup> HPCP-reference[۲۳]

می‌شوند، هیچ پیچ استاندارد یا سیستم کوکی استفاده نشده است و نمونه‌ها به جای اینکه به یک اکتاو نگاشته شوند در اکتاو خودشان گسترده شده‌اند.

الگوریتم کلی شامل دو بخش می‌شود: محاسبه هیستوگرام پیچ و تشخیص نغمه تونیک که در شکل ۶ نشان داده شده است.



شکل ۶ الگوریتم استخراج هیستوگرام پیچ و تشخیص کلید، [۲۲].

### ۴-۳-۱ محاسبه هیستوگرام پیچ

در ابتدا فرکانس اصلی نغمه‌ها از طریق الگوریتم ارائه شده در [۱۱] محاسبه شده است. با محاسبه فرکانس اصلی نغمه‌ها یک هیستوگرام پیچ ( $Hf_0[n]$ ) طبق رابطه زیر تولید می‌شود.

$$Hf_0[n] = \sum_{k=1}^K m_k, \begin{cases} m_k = 1 & \text{for } f_n \leq f_0[k] \leq f_{n+1} \\ m_k = 0 & \text{otherwise} \end{cases} \quad (7-4)$$

که در آن  $(f_n, f_{n+1})$  مقادیر مرزی برای تعیین فرکانس اصلی ( $f_0$ ) در بین  $n$  ام می‌باشند.

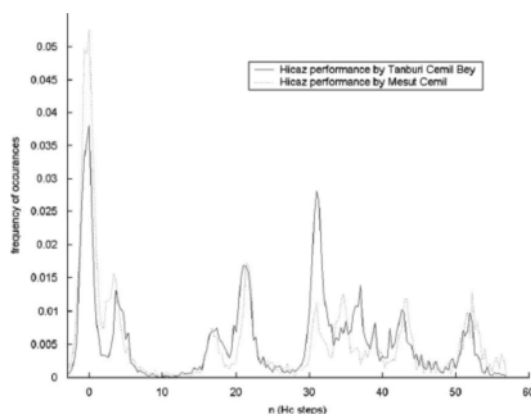
با توجه به تعداد بینها (N) و محدوده فرکانسی ( $f_{0\min}, f_{0\max}$ )، پهنای بین ( $W_b$ ) و لبه‌های هیستوگرام)  $f_n$  به صورت زیر محاسبه می‌شوند:

$$W_b = \frac{\log_2(f_{0\max}) - \log_2(f_{0\min})}{N} \quad (8-4)$$

$$f_n = 2^{f_{0\min} + (n-1)W_b} \quad (9-4)$$

در آنالیز فرکانس اصلی واحدهای لگاریتمی مختلفی از قبیل سنت یا کوما مورد استفاده قرار می‌گیرند. در استفاده از واحد سنت یک اکتاو به ۱۲۰۰ سنت تقسیم می‌شود، در حالی که برای واحد کوما هر اکتاو به ۵۳ واحد لگاریتمی تقسیم می‌شود. در این مقاله برای آسان‌تر شدن مقایسه بین نتایج و تئوری موسیقی ترکی از واحد کوما استفاده شده است. رزولوشن فرکانسی با دقت ۳ بین به ازای هر کوما در نظر گرفته شده است. نمونه‌ای از هیستوگرام مورد نظر در شکل ۴-۷ آورده شده است. محور افقی بینهای فرکانسی را بر حسب واحد کوما<sup>۶۶</sup> و محور عمودی نرخ رخداد فرکانس nام در قطعه موردنظر را نشان می‌دهد.

با محاسبه هیستوگرام پیچ در مرحله بعد نغمه تونیک برای موسیقی ترکی استخراج می‌شود.



شکل ۴-۷: هیستوگرام پیچ برای یک قطعه موسیقی ترکی، [۲۲].

<sup>66</sup> Holderian comma

## ۴-۳-۲ تشخیص نغمه تونیک:

در آنالیز دیتابیس موسیقی ترکی مقایسه فایل‌های متنوع از پیچیدگی زیادی برخوردار است. به دلیل میزان انحرافی که در کوک سازها نسبت به کوک استاندارد وجود دارد به خط کردن دیتاهای آنالیز شده بدون داشتن یک نقطه مرجع امکان‌پذیر نیست.

در تئوری موسیقی ترکی هر قطعه موسیقی، در یک مقام خاص، در نغمه تونیک به عنوان نغمه آخر پایان می‌یابد. در اینجا از روش ارائه شده در [۳۲] برای تشخیص نغمه تونیک استفاده شده است.

در ادامه با تنظیم هیستوگرام بدست آمده برای قطعه موسیقی مورد نظر با الگوهای هیستوگرام پیچ مقام، نغمه تونیک قطعه مورد نظر بدست می‌آید. الگوهای هیستوگرام پیچ برای مقامهای موسیقی ترکی به کمک Gaussian mixture models از قبل محاسبه می‌شوند. جزئیات بیشتر در باب چگونگی محاسبه مدلها در مقاله فوق ارائه شده است.

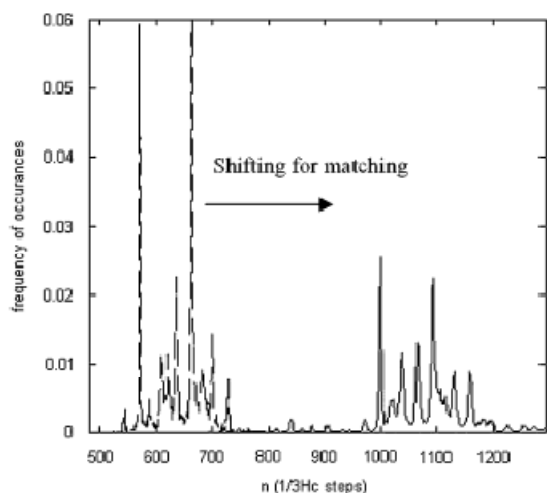
با داشتن مدل‌های مورد نیاز برای مقامها تشخیص مقام برای یک قطعه موسیقی ترکی به ترتیب زیر صورت می‌پذیرد:

۱- حرکت دادن مدل بر روی هیستوگرام پیچ (شکل الف ۴-۸)

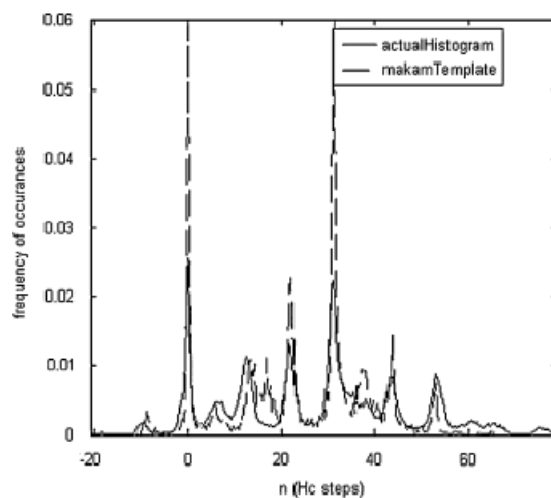
۲- محاسبه میزان انحرافی که حذف آن منجر به بیشینه همبستگی با یک مدل خاص می‌شود. (شکل ۸ ب)

۳- انتخاب آن پیک از هیستوگرام که با تونیک مدل مطابقت دارد، برای تعیین تونیک قطعه.

خلاصه موارد فوق در شکل ۴-۶ نشان داده شده است.



ب



الف

شکل ۴-۸: الف) تنظیم هیستوگرام قطعه مورد نظر با هیستوگرام مدل ب) تعیین میزان انحراف از پیچ استاندارد، [۲۲].

برای انتخاب نغمه تونیک، بهترین نقطه تطابق بین هیستوگرام ها با یافتن بیشینه تابع همبستگی

متقابل،  $c[n]$ ، که از رابطه زیر محاسبه می شود بدست می آید:

$$c[n] = \frac{1}{k} \sum_{k=0}^{k-1} h_r[k] h_t[k-1] \quad (10-4)$$

که در آن  $h_r[n]$  هیستوگرام پیچ قطعات و  $h_t[n]$  الگوی هیستوگرام پیچ برای مقامهای مربوطه می باشد.

به جز تابع همبستگی متقابل که در این قسمت استفاده شد ۴ متریک دیگر:

(City Block, Euclidean Distance, Intersection و Bhattacharya Distance) نیز استفاده شدند که

در جدول ۲ نتایج حاصل از هر کدام آورده شده است.

	CrossCorr.	City Block	Euclidean	Intersection	Bhattacharyya
<i>RAST (24 songs)</i>					
Mean tonic f0 (Hz)	110.05	109.59	109.51	109.59	109.61
Std (Hz)	0.42	0.16	0.19	0.16	0.17
MaxDist (Hz)	1.12	0.26	0.34	0.26	0.27
Std (Cents)	6.52	2.54	2.97	2.54	2.62
MaxDist (Cents)	17.49	4.03	5.29	4.03	4.32
<i>SEGAH (22 songs)</i>					
Mean tonic f0(Hz)	276.62	274.71	274.44	274.71	274.16
Std (Hz)	0.48	0.6	0.5	0.6	0.43
MaxDist (Hz)	1.18	1.77	0.95	1.77	1.23
Std (Cents)	3.01	3.8	3.12	3.8	2.72
MaxDist (Cents)	7.38	7.96	6	7.96	7.72
<i>HUZZAM (16 songs)</i>					
Mean tonic f0 (Hz)	138.17	136.48	136.37	136.48	136.37
Std (Hz)	0.69	0.19	0.21	0.19	0.2
MaxDist (Hz)	1.17	0.38	0.34	0.38	0.32
Std (Cents)	8.64	2.37	2.61	2.37	2.5
MaxDist (Cents)	14.65	4.78	4.28	4.78	4.12
<i>SABA (23 songs)</i>					
Mean tonic f0 (Hz)	123.72	123.32	123.39	123.32	123.37
Std (Hz)	0.49	0.22	0.17	0.22	0.17
MaxDist (Hz)	1.08	0.39	0.34	0.39	0.32
Std (Cents)	6.84	3.08	2.33	3.08	2.42
MaxDist (Cents)	15.05	5.41	4.78	5.41	4.46
<i>HICAZ (19 songs)</i>					
Mean tonic f0 (Hz)	132.16	123.19	131.93	123.19	123.1
Std (Hz)	20.81	0.21	20.83	0.21	0.15
MaxDist (Hz)	52.59	0.44	52.82	0.44	0.29
Std (Cents)	253.14	2.93	253.75	2.93	2.15
MaxDist (Cents)	579.9	6.22	583.01	6.22	4.13
<i>HUSEYNI (23 songs)</i>					
Mean tonic f0 (Hz)	123.54	123.54	123.52	123.54	123.56
Std (Hz)	0.32	0.12	0.13	0.12	0.14
MaxDist (Hz)	1.19	0.21	0.33	0.21	0.33
Std (Cents)	4.49	1.71	1.89	1.71	1.89
MaxDist (Cents)	16.54	3	4.56	3	4.57
<i>USSAK (23 songs)</i>					
Mean tonic f0 (Hz)	125.06	123.35	125.1	123.35	123.28
Std (Hz)	8.37	0.2	8.64	0.2	0.19
MaxDist (Hz)	38.25	0.38	39.04	0.38	0.36
Std (Cents)	112.1	2.85	115.69	2.85	2.64
MaxDist (Cents)	461.95	5.3	476.55	5.3	5.05
Mean-Stds (Cents)	56.39	2.75	54.62	2.75	2.42
Max-MaxDist (Cents)	579.9	7.96	583.01	7.96	7.72
# of false peak detections	4	0	4	0	0

جدول ۲ مقایسه عملکرد متریک های مختلف در تشخیص نغمه تونیک، [۲۲].

### ۳-۳-۴ تشخیص مقام:

در مرحله آخر با استفاده از هیستوگرام پیچ تشخیص مقام انجام می‌شود. برای این کار ابتدا هیستوگرام

پیچ به صورت زیر نرمالیزه می‌شود:

$$Hf_0 N = \frac{Hf_0}{\sum Hf_0} \quad (11-4)$$

الگوهای مورد نظر برای هر مقام از فرمول زیر محاسبه شده و سپس نرمالیزه می‌شود:

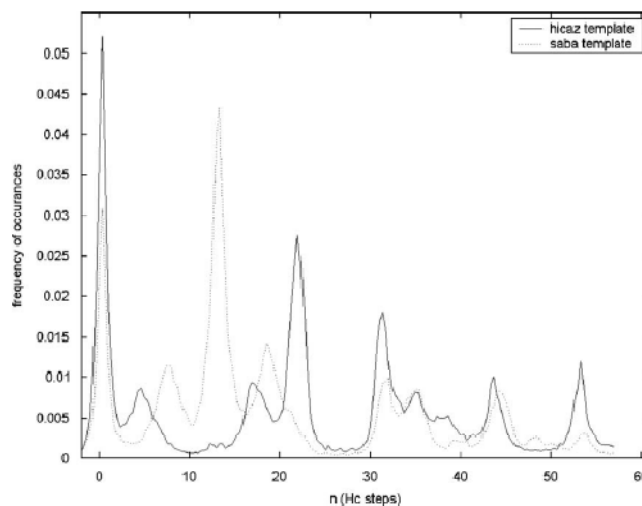
$$T_k = \sum_{i=1}^N Hf_0 N_k(i), \quad TN_k = \frac{T_k}{\sum T_k} \quad (12-4)$$

که  $Hf_0 N_k(i)$  بیانگر هیستوگرام پیچ نرمالیزه شده قطعه  $i$ ام برای مقام نوع  $k$ ام است.  $N$  تعداد قطعات

با مقام  $k$ ام است.  $T_k$  الگوی بدست آمده برای مقام نوع  $k$ ام است و  $T_k N$  الگوی نرمالیزه شده است.

البته در مقاله ذکر نشده که برای هر مقام چه تعداد الگو استخراج می‌شود.

دو الگو برای دو مقام مختلف در شکل ۴-۹ آورده شده است.



شکل ۴-۹: الگوی هیستوگرام پیچ برای دو مقام مختلف موسیقی ترکی، [۲۲].

پس از بدست آوردن تمامی الگوها شباهت بین هر قطعه و الگوهای ممکن بررسی می‌شود.

به کمک پارامترهای زیر سیستم تشخیص مقام معرفی شده ارزیابی می‌شود:

$$recall = \frac{TP}{TP + FN}, \quad precision = \frac{TP}{TP + FP} \quad (13-4)$$

$$F - measure = \frac{2 \cdot recall \cdot precision}{(recall + precision)} \quad (14-4)$$

TP: true positive, TN: true negative, FP: false positive, FN: false negative (15-4)

نتایج در جدول ۳-۴ نشان داده شده است.

Makam type	TP	TN	FP	FN	R	P	F-measure
<i>hicaz</i>	14	150	2	6	70	88	78
<i>rast</i>	14	151	2	5	73	88	79
<i>segah</i>	17	149	3	3	85	85	85
<i>kürdili h.</i>	10	145	11	6	63	48	55
<i>huzzam</i>	10	152	6	4	71	63	67
<i>rihavend</i>	14	143	11	4	78	56	65
<i>hüseyni</i>	10	146	6	10	50	63	56
<i>uşşak</i>	15	138	10	9	63	60	62
<i>saba</i>	16	150	1	5	76	94	84
Mean	13	147	6	6	68	68	68

جدول ۳-۴: نتایج ارزیابی برای سیستم تشخیص مقام توضیح داده شده، [۲۲].

همانطور که از جدول فوق پیداست برای مقامهایی که پارامتر F-measure در آنها کمتر از ۷۰ است

کلاسه بندی به خوبی انجام نشده است.



## ۴-۴ روش سوم:

### بررسی سیستم<sup>۶۷</sup> Tuning برای سیگنال های صوتی

منظور از سیستم Tuning، سیستمهای متنوعی از پیچ‌هاست که برای کوک کردن یک ساز مورد استفاده قرار گرفته و اساس تئوریک آنهاست. به طور واضح تر، منظور از سیستم Tuning انواع کوک‌هاست که الگوی فواصل را تعیین کرده و نقش اساسی در اثربخشی قطعه مورد نظر دارد. در موسیقی های مدال، بین سیستم کوک و مد قطعه رابطه ای مستقیم وجود دارد که موجب ایجاد تنوع در کلاسه بندی قطعات مختلف می‌شود.

در [۳۳]، سیستم Tuning در موسیقی ایرانی بررسی شده است. در این مقاله الگوها (ویژگی‌ها) بر پایه میانگین طیفی و ویژگی کروما از قطعه موسیقی استخراج می‌شود که در ادامه به کمک متریک‌هایی از قبیل manhattan یا همبستگی متقابل تشابه بین آنها با هر کدام از کلاس های کوک ممکن بررسی می‌شود.

البته محدودیتی که در این مقاله برای الگوریتم ارائه شده در نظر گرفته شده است این است که دیتا بیس به قطعات نواخته شده توسط ساز سنتور منحصر می‌شود و آزمایش‌ها نیز بر روی همین دیتابیس انجام گرفته است.

## ۴-۴-۱ ویژگی های استفاده شده:

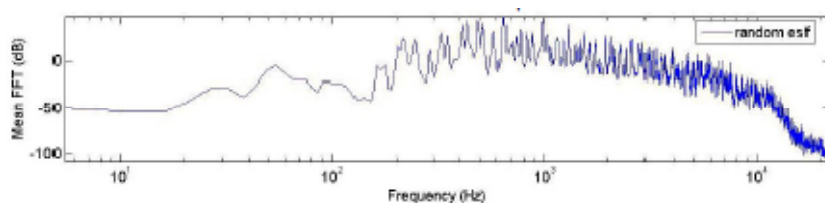
## ۴-۴-۱-۱ میانگین طیفی:

---

<sup>67</sup> the Tuning Systems

پس از پنجره‌بندی که با استفاده از پنجره همینگ با طولهای مختلف صورت می‌گیرد FFT و طیف توان سیگنال محاسبه می‌شود. میانگین طیفی همه فریمها محاسبه شده و سپس نرمالیزه می‌شود. به این ترتیب طی مرحله آموزش<sup>۶۸</sup> الگوهای مرجع برای کلاسهای مختلف بدست می‌آیند. مراحل فوق برای نمونه های تست نیز انجام می‌شود. در پایان یک معیار تشابه برای سنجش میزان تشابه بین نمونه های تست و مدل استفاده می‌شود.

در شکل ۴-۱۰ میانگین طیفی استخراج شده برای یک فایل صوتی موسیقی ایرانی آورده شده است.



شکل ۴-۱۰: میانگین طیفی یک سری نغمه های نواخته شده در مد بیات اصفهان، [۳۳]

با استفاده از نمونه های آموزش، این الگو می‌تواند برای هر کدام از کلاسهای مختلف کوک<sup>۶۹</sup> به کار رود. سپس میانگین طیفی نمونه های تست با آنها مقایسه می‌شود. از معایب این ویژگی این است که بین ملودیهای یکسان که در اکتاوهای مختلف نواخته شده‌اند فرق می‌گذارد. همچنین به نوع ساز و محتوای هارمونیک قطعه نیز وابسته است.

## ۴-۱-۴-۲ کروما:

ویژگی کروما که در این مقاله استفاده شده است ورژن بسیار ساده‌ای از همان پروفایل پیچ ۲۴ تایی است که در قسمت اول نمونه پیشرفته ای از آن معرفی شد.

<sup>68</sup> Training

<sup>69</sup> Tuning

در ابتدا تبدیل CQT با رزولوشن فرکانسی  $\beta = 72$  از رابطه زیر محاسبه می‌شود.

$$X(k) = \sum_{n=0}^{N(k)-1} W(n,k)x(n)e^{-2j\pi f_k n} \quad (16-4)$$

که در آن :

$$f_k = 2^{k/\beta} \cdot f_{\min} \quad (17-4)$$

پس از محاسبه تبدیل فوق ویژگی کروما با نگاشتن طیف فرکانسی فوق به یک اکتاو از رابطه زیر بدست می‌آید:

$$Chroma(k) = \sum_{n=1}^{noc} X(k + n\beta) \quad (18-4)$$

در رابطه فوق  $noc$  تعداد اکتاو است. نهایتاً از رابطه فوق یک پروفایل ۲۴ تائی استخراج می‌شود که از آن در تشخیص مد برای موسیقی ایرانی استفاده می‌شود.

مقایسه نسبتاً کاملی بین این پروفایل، پروفایل ارائه شده در روش اول، [۱۸] و نیز پروفایل جدید دیگری که مولفین ارائه کرده اند در [۵] آورده شده است.

#### ۴-۲-۴ نتایج ارائه شده:

دیتابیزی که در این مقاله استفاده شده از دو جنبه قابل بحث است. یکی اینکه قطعات تنها توسط ساز سنتور نواخته شده‌اند و دیگر اینکه تنها چند کلاس محدود از موسیقی ایرانی (بیات اصفهان، حجاز، چهارگاه و شور) مد نظر قرار گرفته‌اند.

برای دیتا بیس فوق با استفاده از ویژگی اول (میانگین طیفی) ۹۲.۴۸٪ از داده‌ها درست کلاسه بندی شدند. ضمن اینکه همانگونه که در مقاله ذکر شده با تغییر نمونه‌های آموزشی و تست نتیجه فوق نیز تغییر خواهد کرد.

ویژگی دوم (کروما) نتیجه ای معادل با ۸۰٪ برای دیتا بیس مورد استفاده در بر داشته است. علی‌رغم اینکه نتیجه نسبت به میانگین طیفی کاهش داشته است اما از مزایای این ویژگی مقاوم بودن آن نسبت به نوع ساز است.

در اینجا لازم است به پایان نامه فوق لیسانس آقای حیدریان [43] نیز اشاره مختصری داشته باشیم که در آن با استفاده از روش های تشخیص الگو به تشخیص فرکانس اصلی نغمه های موسیقی ایرانی برای ساز سنتور پرداخته شده است.

به طور خلاصه ایشان در این پایان نامه با نمونه‌گیری از دیتا بیس موجود برای ساز سنتور برای هر نغمه موسیقی ۱۰ نمونه تهیه می‌کنند. در ادامه اثرات افزایش دما بر تغییرات بسامد نیز بررسی شده‌اند و پس از تهیه نمونه ها با استفاده از تابع همبستگی متقابل و مقایسه هرکدام از نمونه های تست با نمونه مدل فرکانس اصلی هرکدام از نغمه ها تعیین شده است. در نهایت درصد صحت ۹۰.۸٪ برای تشخیص نغمه های ساز سنتور بدست آمده است.

## فصل پنجم:

### کلاسه بندی دستگاه‌های موسیقی ایرانی با استفاده از روش

#### پیشنهادی

##### ۵-۱ مقدمه:

در این بخش روش پیشنهادی جهت کلاسه بندی دستگاه‌های موسیقی ایرانی توضیح داده می‌شود. ابتدا ویژگی مورد نظر جهت بررسی خواص مشابه در کلاس‌های مختلف موسیقی ایرانی (پروفایل پیچ) بدست می‌آید. دو چالش اصلی در رویارویی با آنالیز موسیقی ایرانی تبیین تئوری مورد استفاده و دیتابیس می‌باشد که در دو بخش جداگانه بررسی و ابزار لازم جهت آماده سازی آنها تشریح می‌شود. و در مرحله آخر نتایج استفاده از ویژگی پروفایل پیچ در کلاسه‌بندی دستگاه‌های موسیقی ایرانی ارائه می‌گردد.

##### ۵-۲ روش پیشنهادی:

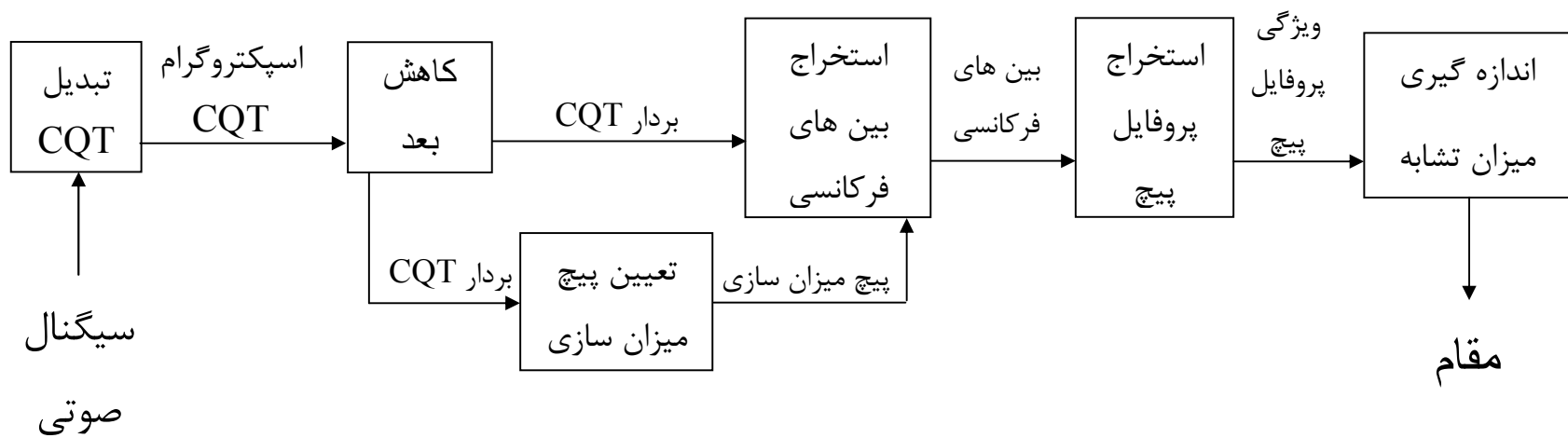
در حوزه پردازش مفاهیم سیگنال‌های صوتی موسیقی، ویژگی‌های مهمی از قبیل ملودی، آکورد، هارمونی، مقام و ویژگی‌های مشابه همگی به منزله ترکیبات مختلفی از پیچ به حساب می‌آیند. در واقع از بین ویژگی‌های مختلفی که در آنالیز ملودیک صوت نقش دارند، پیچ از اهمیت ویژه‌ای برخوردار است [۲۴]. به خاطر اهمیت ویژه پیچ به عنوان یک ویژگی سطح پایین در آنالیز مفهومی موسیقی، تکنیک‌های بسیاری برای تشخیص پیچ ارائه شده است [۳۱-۲۵]. همزمانی هماهنگ‌ها و ترکیبات آنها به خصوص در اجراهای گروهی و نیز در حضور سازهای غیر ملودیکِ ضربی، کار تشخیص پیچ و فرکانس اصلی نغمات را

بسیار دشوار می‌کند. بنابراین، پروفایل پیچ، با ارائه اطلاعاتی آماری در مورد کلیت هماهنگ‌های موجود در قطعه، بسیار مفید تر و بهینه تر از الگوریتم هایبست که به استخراج تنها فرکانس اصلی نغمه ها می-پردازند. در اینجا هدف ما استخراج ویژگی پروفایل پیچ از قطعات موسیقی است که حاوی اطلاعاتی در مورد نغمات نواخته شده در قطعه و هماهنگ های آنها می‌باشد.

نمودار کلی الگوریتم پیشنهادی برای تشخیص دستگاه های موسیقی ایرانی در شکل ۵-۱ نشان داده شده است. در مرحله اول تبدیل CQT برای سیگنال صوتی، با توجه به رزولوشن لازم برای موسیقی ایرانی محاسبه می‌شود. در قدم دوم، طی یک مرحله کاهش بُعد، ماتریس اسپکتروگرام بدست آمده در مرحله قبل به یک بردار تبدیل می‌شود. در واقع، ازین مرحله به بعد به جای آنالیز بر روی ماتریس اسپکتروگرام، محاسبات بر روی یک بردار که حاوی همان ویژگی‌های ماتریس اولیه است انجام می‌گیرد. در مرحله سوم، پیچ میزان‌سازی تعیین و بینهای فرکانسی که احتمال بیشتری در تشکیل مقام قطعه دارند استخراج می-شوند. در قدم چهارم، با استفاده از بین‌های فرکانسی استخراج شده در مرحله قبل، پروفایل پیچ تشکیل می‌شود. نهایتاً در مرحله آخر، محاسبات مربوط به تعیین میزان تشابه پروفایل استخراج شده با هریک از مدل‌های موجود برای چهار گام موسیقی ایرانی انجام شده و مقام قطعه موسیقی تعیین می‌شود. در ادامه هریک از این مراحل به تفصیل بیان می‌شود. روش پیشنهادی و نتایج آن در مقاله [۵] آمده است.

## ۵-۲-۱ محاسبه اسپکتروگرام (تبدیل CQT):

همانطور که در فصل چهارم به طور مفصل توضیح داده شد تبدیل CQT از تبدیل‌های مهم در آنالیز سیگنال‌های موسیقی به شمار می‌رود و نسبت به تبدیل فوریه دارای مزایایی است که در روش



شکل ۵-۱: الگوریتم کلی تشخیص مقام برای موسیقی ایرانی.

پیشنهادی نیز از آن استفاده می‌شود. برای محاسبه تبدیل CQT می‌توان از روابط ۵-۱ تا ۵-۵ استفاده می‌شود:

$$X^{cqt}[k] = 1/N_k \sum_{n=0}^{N_k-1} W_{N_k}[n] x[n] e^{(-j2\pi kn/N_k)} \quad (1-5)$$

$$f_k = (2^b)^k f_{\min} \quad (2-5)$$

$$Q = f_k / \Delta f_k = f_k / [(2^{\frac{1}{b}} - 1) f_k] = (2^{\frac{1}{b}} - 1)^{-1} \quad (3-5)$$

$$N_k = f_s / \Delta f_k = f_s / (f_k Q) = f_s Q / f_k \quad (4-5)$$

$$W[k, n] = \alpha + (1 - \alpha) \cos(2\pi n / N[k]) \quad (5-5)$$

جهت آنالیز سیگنال‌های موسیقی ایرانی پارامترهای موجود در روابط ۵-۱ تا ۵-۵ بدین صورت معین می‌گردند:

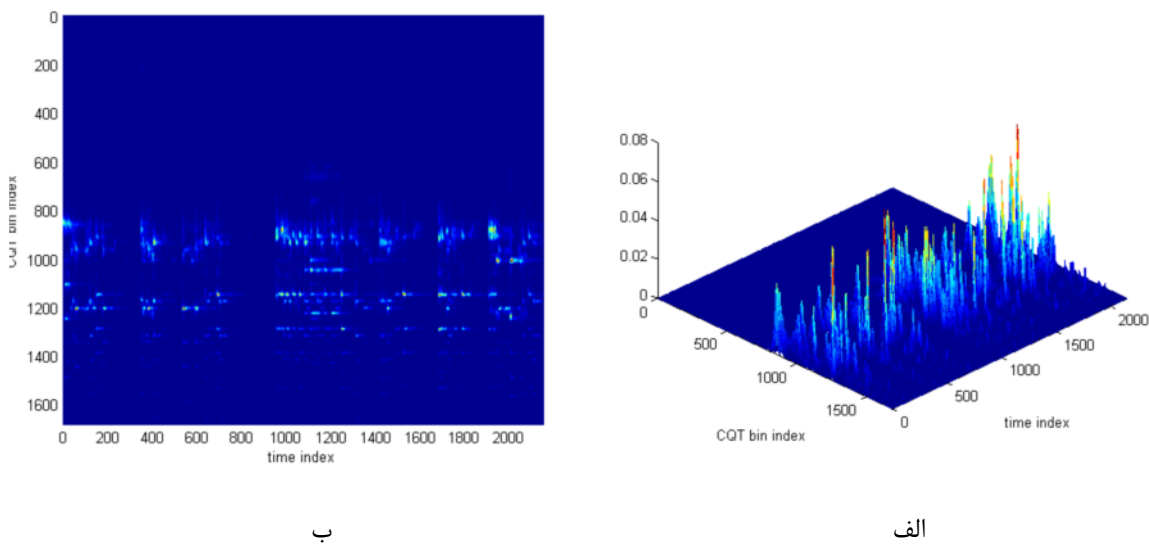
در روابط فوق رزولوشن فرکانسی با  $b$  نشان داده می‌شود. از آنجا که کمترین فاصله بین نغمه‌های مجاور در موسیقی ایرانی ربع پرده است و با توجه به اینکه در یک گام موسیقایی ۲۴ ربع پرده وجود دارد بنابراین برای دستیابی به رزولوشن فرکانسی ۱۰ بین به ازای هر نغمه بایستی  $b = 240$  انتخاب گردد. اگر گستره تغییرات نغمات موسیقی بر اساس ساز پیانو در نظر گرفته شود با توجه به بهترین و زیرترین نغمات ساز پیانو ( $A_1=27.5$  Hz and  $A_7=3520$  Hz)، تغییرات فرکانس در محدوده ۷ هنگام اتفاق خواهد افتاد. فرمول شماره (۱) تبدیل CQT را برای سیگنال صوتی گسسته در زمان، برای  $k$  امین بین فرکانسی نشان می‌دهد. بین فرکانسی  $k$  ام از رابطه ۲ و با توجه به فرکانس مرجع  $f_{\min} = 27.5$  Hz محاسبه می‌-



شود. با انتخاب  $b = 240$  ،  $Q = (2^{(1/b)} - 1)^{-1} = 345.747$  می‌شود. روابط ۴ و ۵ نوع پنجره و اندازه آن را

برای

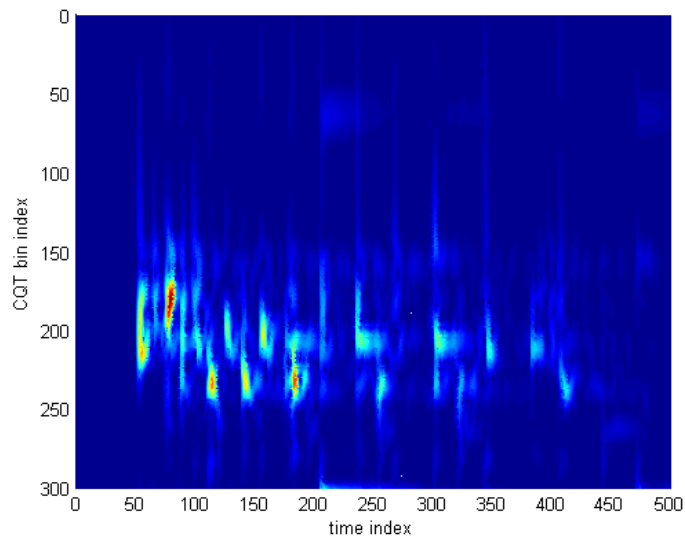
کامین بین فرکانسی نشان می‌دهد. پنجره مورد استفاده در این تبدیل پنجره همینگ با  $\alpha = 25/46$  بوده و فرکانس نمونه برداری  $f_s = 44.1KHz$  بوده و استپ سایز برای تمامی اجزای فرکانسی ۵۱۲ می‌باشد. در شکل ۲-۵ یک نمونه از اسپکتروگرام بدست آمده برای یک قطعه موسیقی ایرانی که قطعاتی از نواخته های استاد محمد رضا لطفی می‌باشد نشان داده شده است.



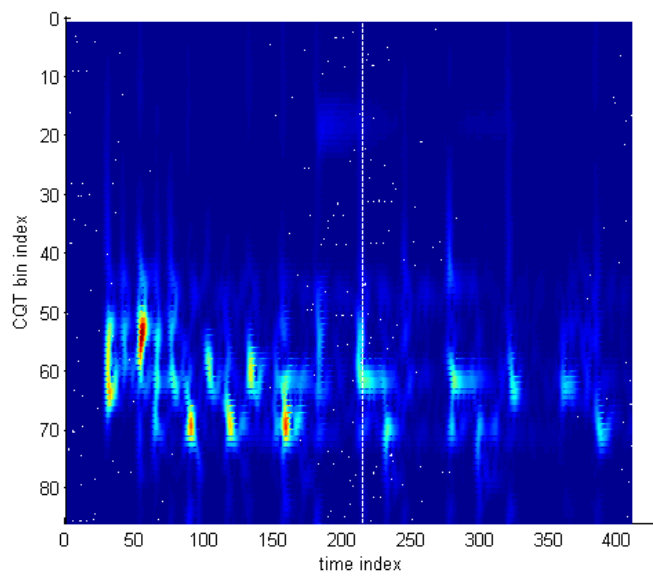
شکل ۲-۵: اسپکتروگرام یک قطعه موسیقی ایرانی از دو زاویه مختلف

همان طور که از شکل ۲-۵ پیداست و در ابتدای بحث نیز اشاره کردیم در محدوده ۷ هنگام و با رزولوشن فرکانسی ۱۰ بین برای هر نغمه موسیقی نهایتاً ۱۶۸۰ بین فرکانسی خواهیم داشت که بر روی محور عمودی نشان داده شده‌اند. حال اگر رزولوشن فرکانسی را تا ۳ بین برای هر نغمه پایین بیاوریم مشاهده خواهیم کرد که چه وضوحی اتفاق خواهد افتاد. شکل ۳-۵ اسپکتروگرام سیگنال قبلی را برای دو رزولوشن مختلف نشان می‌دهد. کاهش وضوح اسپکتروگرام در مراحل بعدی موجب تداخل اطلاعات

موجود در بین های فرکانسی شده و تولید پروفایل های نویزی می کند که در بخش های بعد توضیح داده می شود.



الف



ب

شکل ۵-۳: مقایسه اسپکتروگرام های با رزولوشن مختلف.

## ۵-۲-۲ کاهش بُعد:

ویژگی‌هایی که در ساختار یک مقام نقش دارند را می‌توان به دو دسته کلی ویژگی‌های زمانی و ویژگی‌های فرکانسی تقسیم نمود. به عنوان مثال پروفایل پیچ یک ویژگی فرکانسی است. یعنی تغییرات فرکانسی صوت را بررسی می‌کند و اهمیتی به تغییرات زمان نمی‌دهد. از طرفی ویژگی سیر ملودیک مربوط به جنبه‌های زمانی صوت می‌شود و تغییرات زمانی نقش به‌سزایی در تشکیل و تغییر این ویژگی دارد. بنابراین، در بدست آوردن ویژگی‌های فرکانسی دو رویکرد مختلف را می‌توان اتخاذ کرد. یکی اینکه تغییرات فرکانسی را در پنجره‌های زمانی مختلف بررسی کرده و یک نوع آنالیز دینامیک انجام دهیم. دیگر اینکه با توجه به بی‌اهمیت بودن زمان در ویژگی فرکانسی، در اینجا پروفایل پیچ، آنالیز را به صورت استاتیک و بدون در نظر گرفتن تغییرات زمانی انجام دهیم. بنابراین، در مواردی که آنالیز استاتیک می‌تواند کارایی داشته باشد، وارد کردن محاسبات مربوط به زمان برای استخراج یک ویژگی فرکانسی زمان بر بوده، بدون اینکه تاثیر آنچنان زیادی در نتیجه نهایی داشته باشد. به بیان دیگر، در استخراج پروفایل پیچ، از آنجا که هیچ نوع ویژگی زمانی استخراج نمی‌شود پس، نیازی به استفاده از هیچ نوع ویژگی زمانی احساس نمی‌شود [۵]. در روش ارائه شده در [۱۸] که در فصل پنجم به تفصیل معرفی شد تغییرات زمانی صوت در محاسبات وارد شده و جنبه‌های دینامیک آن مورد بررسی قرار گرفته است. البته شاید برای موسیقی غربی به خصوص موسیقی پاپ که سازهای کوبه‌ای حضور چشمگیری در قطعات موسیقی دارند استفاده از جنبه‌های دینامیک صوت نتایج بهتری در بر داشته باشد. ما در این بخش نشان خواهیم داد که استفاده از جنبه‌های استاتیک صوت در آنالیز موسیقی ایرانی نتایج قابل قبولی داشته و از طرفی سرعت محاسبات را نیز به شکل چشم‌گیری تا ۱۰ برابر افزایش می‌دهد.

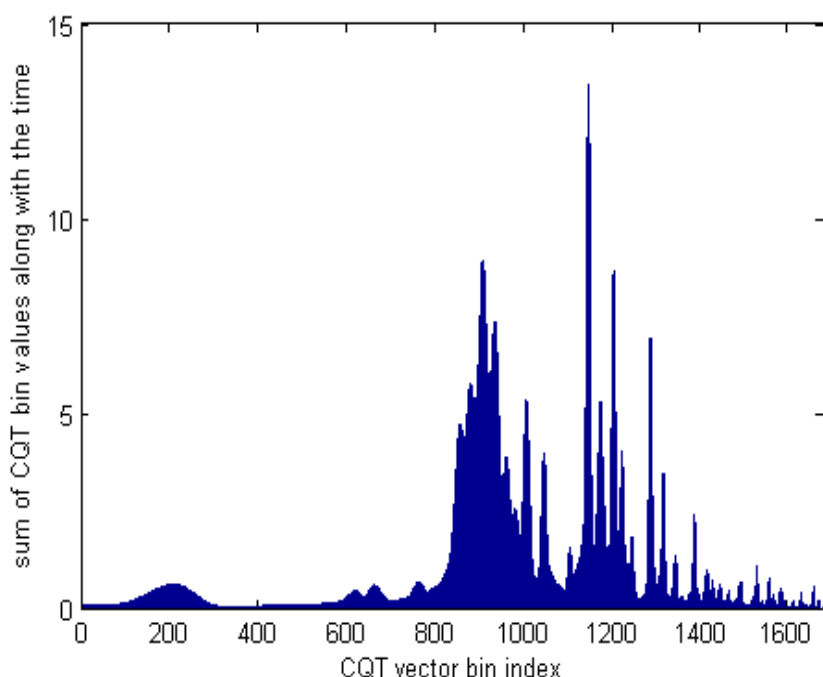
در این بخش متغیر زمان از روابط حذف شده و محاسبات بدون دخالت پارامتر زمان انجام می‌شود. با این کار یک بردار بدست می‌آید که تنها بین‌های فرکانسی و درصد حضور هریک از آنها در اسپکتروگرام را نشان می‌دهد بدون اینکه هیچ‌گونه وابستگی به زمان و تغییرات آن داشته باشد.

اگر اسپکتروگرام سیگنال را با  $X^{cq}[t,k]$  نشان دهیم، آنگاه  $V^{cq}[k]$  بردار  $CQT$  است که در آن  $k$  طول بردار می‌باشد. ارزش  $k$  امین جزء این بردار مجموع آرایه‌های اسپکتروگرام  $X^{cq}[t,k]$  در طول زمان است که در رابطه ۵-۶ نشان داده شده است.

$$v^{cq}[k] = \sum_{t=1}^T x^{cq}[t,k] \quad (۶-۵)$$

در این رابطه  $T$  تعداد پنجره‌های زمانی اسپکتروگرام و واحدی برای تغییرات زمان است.

نمایشی از بردار  $CQT$  برای سیگنال قبلی که اسپکتروگرام آن محاسبه شد در شکل ۵-۴ نشان داده شده است. این بردار تمامی اطلاعات اسپکتروگرام را در بر دارد با این تفاوت که ماتریس اسپکتروگرام به بردار تبدیل شده است. واضح است که آنالیز این بردار بسیار ساده و سریع تر از ماتریس اسپکتروگرام  $CQT$  خواهد بود. چنان که از شکل پیداست، این بردار یک مدل فشرده شده از سیگنال است که چگالی بین‌های فرکانسی در آن کاملاً با اسپکتروگرام سیگنال هم‌خوانی دارد. این بردار در مرحله بعد جهت بدست آوردن پیچ میزان سازی استفاده شده و تفاوت اصلی روش ارائه شده در اینجا با روش ارائه شده در [۱۸] در استفاده از همین بردار در محاسبات است.



شکل ۴-۵: بردار  $CQT$  برای سیگنال موسیقی ایرانی، استخراج شده از اسپکتروگرام سیگنال. برگرفته از [۵].

### ۳-۲-۵ تعیین پیچ میزان سازی:

عموما نوازندگان، به خصوص در موسیقی شرقی، تحت تاثیر توانایی‌های ساز یا حس و حال حاکم بر زمان اجرا، سازهایشان را متفاوت با استاندارد های موجود برای نغمه‌های موسیقایی کوک می‌کنند. بنابراین معیار کوک کردن سازها به جای یک وسیله استاندارد و مرجع، مثل دیپازون، گوش نوازندگان و حال و هوای آنها در لحظه می‌شود. به این ترتیب، به خصوص برای سازهایی مثل تار که موقعیت نغمه‌ها بر روی دسته تار متغیر است، به هم‌ریختگی کوک<sup>۷۰</sup> زیادی اتفاق می‌افتد که بایستی اصلاح شود. ما از روش ارائه شده در [۱۸] برای اصلاح از کوک در رفتگی استفاده می‌کنیم. اصلاح از کوک در رفتگی در واقع تنظیم بین های فرکانسی در موقعیت‌های مشخص شده در فواصل تعدیل شده است. در روش ارائه

<sup>70</sup> mistuning

شده در [۱۸] تنها مزیت تعیین پیچ مرجع اصلاح از کوک در رفتگی موجود در گروه نوازی هاست. در حالی که از مزایای دیگر تعیین پیچ مرجع، تنظیم بین های فرکانسی است که از سازهای غیر دقیق [۵] و یا از فواصل تعدیل نشده که در موسیقی شرقی به خصوص ایرانی فراوان اجرا می شوند، ناشی می شوند. به بیان ساده تر با تعیین پیچ مرجع، بین های فرکانسی را در موقعیت های تعریف شده برای فواصل تعدیل شده قرار می دهیم. در جمله فوق منظور از سازهای غیر دقیق، همان سازهای با پرده بندی متغیر مثل تار و سه تار است که امکان اجرای فواصل تعدیل نشده را برای نوازنده فراهم می کند.

در [۱۸] عمل تعیین پیچ مرجع برای هر کدام از پنجره های زمانی اسپکتروگرام به طور جداگانه انجام شده و در پایان پیچ مرجع را از بین  $T$  پیچ مرجعی که بدست آورده انتخاب می کند. ولی ما در اینجا آنالیز را به طور استاتیک و تنها یک بار، آن هم برای بردار  $CQT$  انجام داده و آن را به ورودی مرحله بعد اعمال می کنیم. نتایج نشان دادند که میزان انحراف از پیچ مرجع در هر دو روش یکسان است.

برای محاسبه پیچ مرجع از روی بردار  $CQT$  ابتدا پیک های بردار را استخراج می کنیم.

$$\text{برای: } v^{cq}[k] > v^{cq}[k-1] \& v^{cq}[k+1] \quad \hat{v}^{cq}[k] = v^{cq}[k]$$

از آنجا که کمینه فاصله نغمه های مجاور در موسیقی ایرانی یک ربع پرده یا ۵۰ سنت است، انحراف از پیچ مرجع می تواند بین ۲۵- تا ۲۵+ اتفاق بیافتد. دقت بردار  $CQT$  ۵ سنت بود. بنابراین برای انحراف از پیچ مرجع ۱۰ مقدار می تواند وجود داشته باشد. انرژی موجود برای هر یک از ۱۰ مقدار ممکن برای انحراف از پیچ مرجع به صورت زیر محاسبه می شود:

$$\hat{E}[n] = \sum_{k=n+10m} \hat{v}^{cq}[k] \quad (7-5)$$

که در آن  $n$  بین ۱ تا ۱۰ و  $m$  از ۱ تا ۱۶۸\*۲۴ تغییر می کند. در ادامه موقعیت پیچ مرجع با توجه به میزان انرژی ماکزیمم در موقعیت  $n$ م بدست می آید:

$$p = \arg \max(\hat{E}[n]) \quad (8-5)$$

در رابطه ۸،  $p$  میزان انحراف از پیچ مرجع را نشان می‌دهد که عددی بین ۱ تا ۱۰ است. برای عدد  $p=1$  هیچ گونه انحرافی از پیچ استاندارد صورت نگرفته است. عدد ۲ نشانگر انحراف ۵ سنتی به سمت بالای پیچ استاندارد است و به طور مشابه عدد ۱۰ بیانگر انحراف ۵ سنتی به سمت پایین پیچ استاندارد می‌باشد. پس از تعیین میزان انحراف از پیچ استاندارد همگی بین‌های فرکانسی به میزان مشخص شده تغییر موقعیت می‌دهند. در مرحله بعد بین‌های فرکانسی تغییر موقعیت داده و آنهایی که احتمال بیشتری به شرکتشان در ساخت مقام قطعه است استخراج می‌شوند.

### ۵-۲-۴ استخراج بین‌های فرکانسی:

در این بخش، بین‌های فرکانسی نغمات که احتمال حضور آن‌ها در تشکیل مقام قطعه بیشتر است از بردار  $CQT$  استخراج می‌شوند. بین‌های فرکانسی برای بردار  $CQT$  به صورت زیر محاسبه می‌شود:

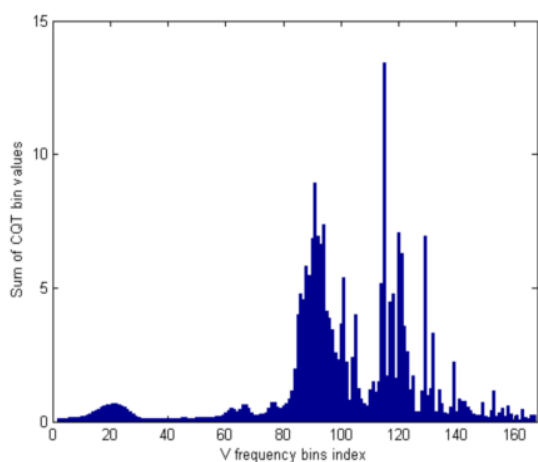
$$V[k] = \max_{k=10(m-1)+p-2}^{10(m-1)+p+2} [v^{cq}[k]] \quad (9-5)$$

در رابطه فوق  $m$  از ۱ تا  $168 = 7 * 24$  و  $p$  از ۱ تا ۱۰ تغییر می‌کنند. در رابطه ۵-۹ بردار  $V$  حاوی آن دسته از بین‌های فرکانسی است که با پیچ مرجع هم‌کوئند. فاصله بین‌های فرکانسی از هم یک ربع پرده بوده و در محدوده ۷ هنگام گسترده شده‌اند. در این قسمت یک مرحله فیلترینگ ساده بر روی بردار  $V$  صورت گرفته و برای استفاده در مرحله تشکیل پروفایل پیچ آماده می‌شود.

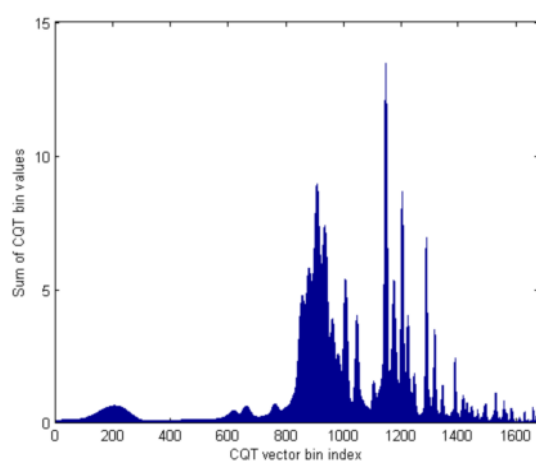
$$\hat{V}[k] = V[k] \quad \text{if} \quad V[k] > V[k-1] \& V[k+1] \quad (10-5)$$

در شکل ۵-۵ نمایشی از بردار  $CQT$  و بین‌های فرکانسی استخراج شده در این مرحله قبل و بعد از فیلترینگ مشاهده می‌شود. همانطور که از شکل پیداست، بین‌های فرکانسی از ۱۶۸۰ بین به ۱۶۸ بین فشرده شده‌اند. ضمن اینکه بین‌های اضافی در شکل نهایی، ج ۵، حذف شده‌اند.

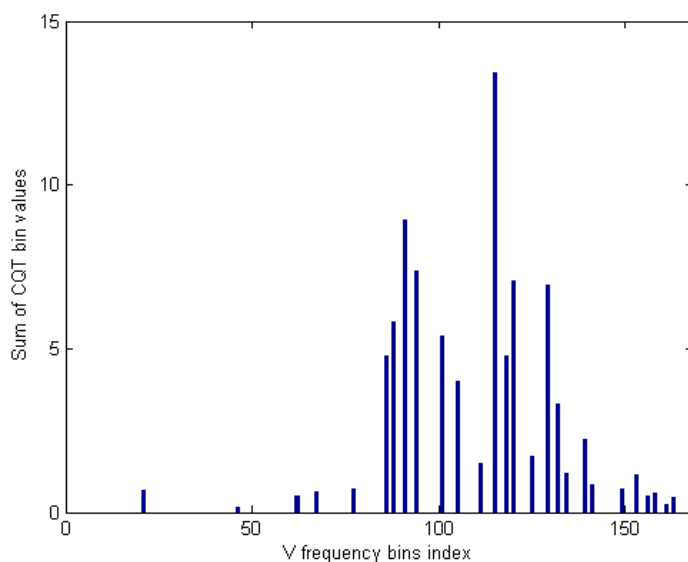




ب



الف



ج

شکل ۵-۵: الف: بردار  $CQT$ ، ب: بردار  $V$  قبل از فیلتر شدن، ج: بردار  $V$  بعد از فیلتر شدن.

## ۵-۲-۵ استخراج پروفایل پیچ:

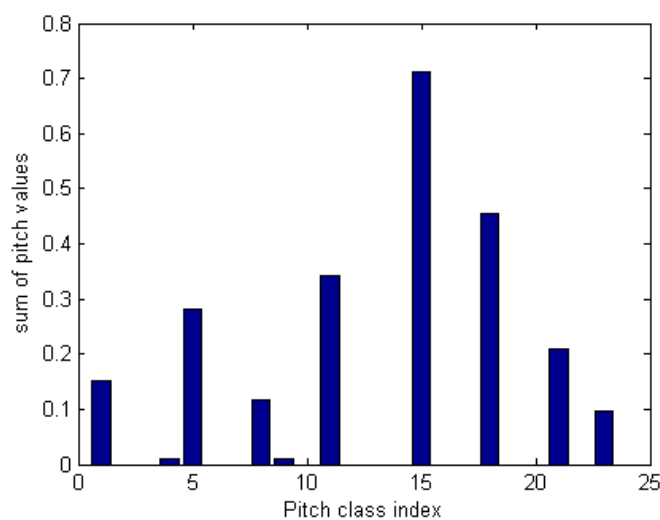
پروفایل پیچ ویژگی ایست که توزیع اجزای فرکانسی نغمات حاضر در سیگنال موسیقی را در یک پنجره زمانی نشان می‌دهد. مفهوم فواصل و موقعیت نغمه‌های شاهد و ایست از توانایی‌های بالقوه پروفایل پیچ است. اگر چه در موسیقی ایرانی مقام می‌تواند در فاصله کمتر از چند ثانیه تغییر کند، استخراج پروفایل

پیچ از قطعاتی با طول ۲۰ تا ۲۵ ثانیه می‌تواند به خوبی تغییرات لحظه ای مقام را نشان دهند. در ادامه، از بردار  $\hat{V}$ ، جهت استخراج پروفایل پیچ، بدین صورت استفاده می‌شود:

$$P_b(n) = \sum_{k=0}^6 \hat{V}(n + 24k) \quad (11-5)$$

که در آن  $n$  از ۱ تا ۲۴ (ربع پرده موسیقی ایرانی) است و تغییرات  $k$ ، ۷ هنگامی که نغمه موسیقی ممکن است در یکی از آنها نواخته شده باشد را نشان می‌دهد.

نمایی از پروفایل پیچ برای سیگنال موسیقی قبلی در شکل ۵-۶ نشان داده شده است. قطعه مورد استفاده که از نواخته های استاد محمد رضا لطفی است در مقام شور نواخته شده است. بین پانزدهم نغمه شاهد را معرفی می‌کند. اگر از نغمه شاهد (بین پانزدهم) شروع کنیم می‌بینیم که الگوی " ۳ ۴ ۴ ۳ ۳ ۴ "، که الگوی مقام شور است دقیقاً در آن رعایت شده است. هریک از شماره های موجود در پروفایل پیچ نشان داده شده معرف یکی از ۲۴ نغمه موجود در موسیقی ایرانی است که در جدول ۵-۱ معرفی شده‌اند.



شکل ۵-۶: پروفایل پیچ برای یک قطعه موسیقی ایرانی نواخته شده در مقام شور.

جدول ۵-۱: نغمه های موجود در یک هنگام برای موسیقی ایرانی					
شماره بین	نغمه معادل	شماره بین	نغمه معادل	شماره بین	نغمه معادل
1	C	9	E <sup>#</sup>	17	A
2	C <sup>sori</sup>	10	F <sup>coron</sup>	18	A <sup>sori</sup>
3	C <sup>#</sup>	11	F	19	A <sup>#</sup>
4	D <sup>coron</sup>	12	F <sup>sori</sup>	20	B <sup>coron</sup>
5	D	13	F <sup>#</sup>	21	B
6	D <sup>sori</sup>	14	G <sup>coron</sup>	22	B <sup>sori</sup>
7	E	15	G	23	B <sup>#</sup>
8	E <sup>sori</sup>	16	G <sup>sori</sup>	24	C <sup>coron</sup>

### ۵-۲-۶ اندازه گیری میزان تشابه:

چنان که در شکل ۵-۲ نیز نشان داده شده است آخزین مرحله در یک سیستم تشخیص مقام، اندازه گیری میزان تشابه است. به بیان دیگر، خروجی یک سیستم تشخیص مُد یک مقیاس تشابه یا عدم تشابه بین نمونه‌های تست و مدل به صورت دوه‌دو است. برخی از مقیاس های مورد استفاده برای تعیین میزان تشابه عبارتند از: همبستگی متقابل، انواع متریک ها و یا حتی ضرب نقطه ای، چنان که در [۲۲] از استفاده شده است. در قسمت نتایج ما از فاصله مینکفسکی برای تعیین میزان تشابه نمونه ها استفاده کردیم که جزئیات آن در بخش بعد به تفصیل بیان می‌شود.

### ۵-۳ نتایج:

در مطالعه فواید پروفایل پیچ و اثرات آن جهت بررسی خواص تشابه یک نوع موسیقی، دو پارامتر اساسی لازم است؛ اول یک تئوری مناسب که بتوان پایه نظری تحلیل ها را بر آن استوار کرد و دیگر یک دیتابیس کامل برای مرحله تست و آزمایش. متأسفانه در موسیقی های شرقی همچون موسیقی ایرانی کمبود یک تئوری منسجم و نیز یک دیتا بیس کامل و جامع از مشکلات چالش برانگیز برای بازیابی

اطلاعات آن هاست. ازین رو، ما در دو بخش جداگانه به توضیح و تشریح تئوری و دیتابیس مورد استفاده در این مجموعه می‌پردازیم و آنالیزمان را بر مبنای اصول ارائه شده در این بخش قرار می‌دهیم.

در ادامه میزان کارایی روش پیشنهادی مورد بررسی قرار گرفته و با دو روش ارائه شده در [۱۸] و [۳۳] که در فصل ۵ نیز توضیح داده شدند مورد مقایسه قرار می‌گیرد.

همگی نمونه‌ها سیگنال‌های موسیقی 16 bits single channel Waveform، 44100 Hz می‌باشند. برای محاسبات از یک لپ‌تاپ وایو با پردازنده ۲.۴ گیگا هرتز دو هسته‌ای و سیستم عامل ویندوز ویستا استفاده شده است.

## ۵-۳-۱ تئوری موسیقی مبنا:

اگر چه حضور دانشمندان و موسیقی دانان برجسته‌ای مثل ابو نصر فارابی، ابو علی سینا، صفی‌الدین ارموی، قطب‌الدین شیرازی و عبد القادر مراغه‌ای گواهی‌دهنده یک دوره درخشان برای موسیقی قدیم ایران است اما، متأسفانه، در دوره معاصر، یک تئوری منسجم و قابل اعتماد در بین تئوری‌های مختلفی که توسط موسیقی دانان یکصد سال گذشته ایران ارائه شده یافت نمی‌شود به طوری که مسائل بحث برانگیز بسیاری در بین آنان وجود دارد.

اولین تئوری ارائه شده در دوران معاصر مربوط می‌شود به تئوری گام ۲۴ ربع پرده‌ای استاد علی نقی وزیری که به عنوان پایه موسیقی ایرانی معرفی شد [۳۴]. او تقسیم بندی ردیف دستگاهی موسیقی ایران را نپذیرفت و نوا و راست پنجگاه را از مشتقات شور و ماهور معرفی کرد. طبق نظر او موسیقی ایرانی تنها دارای پنج دستگاه بوده و هفت آواز دیگر به عنوان مشتقات این پنج دستگاه اصلی به شمار می‌آیند.

تئوری دیگر مربوط می‌شود به گام ۲۲ پرده ای توسط مهدی برکِشلی ارائه شد [۳۵] و البته در حد تئوری باقی ماند و هیچگاه جنبه عملی پیدا نکرد. به عنوان مثالی دیگر می‌توانیم تئوری فرهنگ در باب فواصل متغیر را یادآور شویم [۳۶]. او با معرفی پنج فاصله اساسی که همگی مقام های موسیقی ایرانی توسط آن پنج فاصله تعریف می‌شوند تئوری گام برای موسیقی ایرانی را به کل مردود دانست. تئوری آخر مربوط می‌شود به تئوری دانگی که توسط داریوش طلایی مطرح شد [۳۷] و در ادامه حسین مهرانی آن را بسط داد [۳۸].

ما برای ساختن مدل‌ها (الگوها)ی مقام های موسیقی ایرانی، با آنالیز اجراهای مختلفی که از انواع مختلف موسیقی ایرانی در دست بود، تئوری خودمان را که جنبه عملی بیشتری در مقایسه با تئوری‌های ذکر شده در بالا دارد مورد استفاده قرار دادیم. از نظر ما، موسیقی ایرانی از چهار گام اصلی تشکیل شده که مقام های موسیقی ایرانی همگی در بستره این چهار گام اصلی تعریف می‌شوند. برای مقام های شور، ابوعطا، بیات زند، افشاری و نوا گام "۳ ۳ ۴ ۴ ۳ ۳ ۴"، برای ماهور، راست پنجگاه و دشتی گام "۴ ۴ ۴ ۲"، برای سه گاه گام "۳ ۴ ۳ ۴ ۳ ۴ ۳" و نهایتاً برای چهارگاه، همایون و بیات اصفهان گام "۳ ۵ ۱" خواهد بود. نکته جالب اینکه، اگر چه برای برخی از کلاس‌ها الگوی فرکانسی (گام) مشترک است اما کلاس‌ها با یکدیگر فرق دارند. این تفاوت به خاطر نغمه های شاهد و ایست و جنبه های زمانی است که در هر یک از کلاس‌ها فرق دارد.

## ۵-۳-۲ دیتابیس:

برای دیتابیس، مجموعه‌ای از انواع نمونه‌های موسیقی ایرانی جمع‌آوری شده است. نمونه‌ها مجموعه‌ای از قطعات ضبط شده از نوازندگان چیره‌دست موسیقی ایرانی است که آثارشان در شرکت های معتبر نشر و تکثیر سی‌دی و کاست از قبیل: موسسه فرهنگی و هنری ماهور، موسسه آوای شیدا، موسسه چهارباغ و

برخی موسسات دیگر ازین دست منتشر شده است [۳۹-۴۱]. یک موسیقی‌دان آگاه به همگی قطعات گوش داده و آن‌ها را بر اساس مدل‌های پیشنهادی کلاسه‌بندی کرده است. نمونه‌ها دارای طول مختلفی بودند اما تنها ۲۰ تا ۳۰ ثانیه از آن‌ها در آنالیز مورد استفاده قرار گرفتند. جدول شماره ۲ آماری از دیتابیزی که در ارزیابی صحت روش تشخیص مقام ارائه شده و مقایسه آن با دو روش دیگر مورد استفاده قرار گرفته است را ارائه می‌کند. ارقام نشان داده شده در جدول بیانگر تعداد نمونه‌هایی است که برای هر کلاس، با توجه به نوع سازی که آن را نواخته، در این مجموعه گنجانده شده است. دیتابیس مجموعاً از ۱۷۹ نمونه تشکیل شده است. تعداد نمونه‌های موجود برای گروه نوازی ۲۶ عدد، برای ساز سه تار ۳۱ عدد، برای ساز کمانچه ۱۵ عدد، برای ساز نی ۲۰ عدد، برای ساز سنتور ۲۱ عدد، برای ساز تار ۵۵ عدد و ۱۱ قطعه باقی مانده برای ساز ویلن ایرانی می‌باشد.

جدول ۵-۲: گزارش آماری دیتابیس.

کلاس‌ها	سازها							جمع کل
	سه تار	گروه نوازی	کمانچه	نی	سنتور	تار	ویلن	
شور	7	2	3	2	2	6	6	28
سه گاه	2	3	0	1	2	10	2	20
همایون	2	2	1	3	3	3	0	14
ماهور	3	2	0	2	2	8	1	18
دشتی	2	2	2	3	2	3	1	15
بیات اصفهان	2	2	1	1	2	6	0	14
بیات زند	2	3	1	2	2	2	1	13
ابوعطا	3	2	2	1	2	2	0	12
افشاری	2	2	0	1	2	5	0	12
نوا	2	3	1	2	0	5	0	13
چهارگاه	2	3	3	2	2	3	0	15
راست پنجگاه	2	0	1	0	0	2	0	5
جمع کل	31	26	15	20	21	55	11	179

۵-۳-۳ الگوریتم‌های مورد مقایسه:

برای ارزیابی نتایج بدست آمده در این بخش، مقایسه ای بین روش پیشنهادی با دو الگوریتم پیشین که در فصل قبل توضیح داده شد، [۱۸ و ۳۳] انجام گرفته است. این دو الگوریتم دو روش کاملا متفاوت می باشند؛ یکی برای تشخیص کلید در موسیقی غربی و دیگری برای بررسی سیستم Tuning در موسیقی ایرانی. ایده اولیه روش پیشنهادی ما از روش Zhu گرفته شده است. علاوه بر تطبیق پارامترهای استفاده شده در موسیقی غربی با مفاهیم موسیقی ایرانی، یک مرحله کاهش بعد نیز به آن اضافه شده که توانسته است سرعت محاسبات را به شکل چشمگیری افزایش دهد. در [۳۳] میانگین طیفی و یک روش ساده مبتنی بر ویژگی کروما برای کلاسه بندی موسیقی ایرانی استفاده شده است. هرچند که الگوریتم اساسا برای قطعات نواخته شده با ساز سنتور طراحی شده است. برای تشخیص مقام های موسیقی ایرانی از روی سیگنال های صوتی، یک الگوریتم بهینه و جامع ارائه شد که نه تنها الگوریتم استخراج پروفایل پیچ را بهبود می بخشد، بلکه بدون وابستگی به سازی که قطعه را نواخته و بدون هیچ گونه محدودیتی در مقایسه با روش های پیشین کار می کند.

#### ۵-۴ مقایسه نتایج:

در قسمت آزمایش ها، برای ارزیابی روش پیشنهادی سه بخش جداگانه ارائه می شود. در بخش اول نگاهی کلی و اجمالی به سه پروفایل بدست آمده در سه روش مورد بحث می اندازیم و میزان شباهت هرکدام به پروفایل ایده آل را بررسی می کنیم. در بخش دوم سرعت الگوریتم پیشنهادی با روش Zhu بررسی می شود و نهایتا در بخش سوم نتایج تشخیص مقام برای موسیقی ایرانی توسط سه الگوریتم ارائه می گردد.

#### ۵-۴-۱ مقایسه اول:

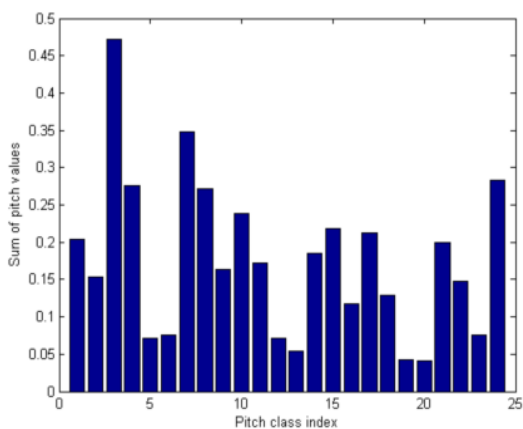
نمایی از پروفایل پیچ های بدست آمده از هر سه روش مورد بحث در شکل ۵-۷ نشان داده شده است. در شکل الف ۵-۷ پروفایلی که توسط یک سیستم ایده آل برای قطعه مورد نظر می‌تواند استخراج شود نشان داده شده است. قطعه در مقام سه گاه نواخته شده است که الگوی آن "۳ ۴ ۳ ۴ ۳ ۴ ۳" است. در هر سه روش مورد بحث، تلاش در جهت دستیابی به پروفایلی است که بیشترین شباهت را به پروفایل واقعی و ایده آل داشته باشد. در شکل ب ۵-۷ پروفایل بدست آمده از روش آقای حیدریان نشان داده شده است. اگر چه ویژگی کروما که در این الگوریتم استفاده شده دارای رزولوشن ۳ بین به ازای هر نغمه است، اما به دلیل عدم استفاده از تکنیک های فیلترینگ، یک پروفایل بسیار نویزی که نغمه های مجاور در یکدیگر تداخل دارند حاصل شده است. در شکل ج ۵-۷ پروفایلی را مشاهده می‌کنید که از روش آقای Zhu بدست آمده است. در مقایسه با پروفایل ایده آل، برخی بین های فرکانسی اضافی در آن دیده می‌شود، اما به هر حال شباهت آن به پروفایل واقعی قابل قبول است. پروفایل پیچی که از روش پیشنهادی ما بدست آمده است در شکل ۵-۷ نشان داده شده. همان طور که از شکل مشخص است، در مقایسه با دو پروفایل دیگر شباهت بسیار زیادی به پروفایل واقعی دارد. شباهت این پروفایل به پروفایل ایده آل بسیار واضح بوده و تنها با یک نگاه قابل تشخیص است، که خود مزیتی برای آن به حساب می‌آید.

#### ۵-۴-۲ مقایسه دوم:

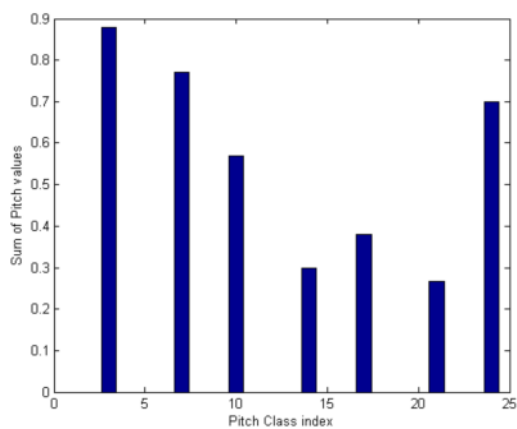
روش حیدریان از نظر ساختار کاملا با دو روش دیگر متفاوت است، به همین دلیل این بخش تنها به مقایسه سرعت دو روش دیگر که الگوریتم آنها شباهت زیادی به یکدیگر دارند و یکی بهینه شده دیگری است می‌پردازد. نتایج این مقایسه در جدول ۳ نشان داده شده است. برای مقایسه سرعت آن ها، دیتابیس به سه دسته با طول های متفاوت تقسیم شد: دسته اول قطعات با طول ۲۰ ثانیه، دسته دوم قطعات با



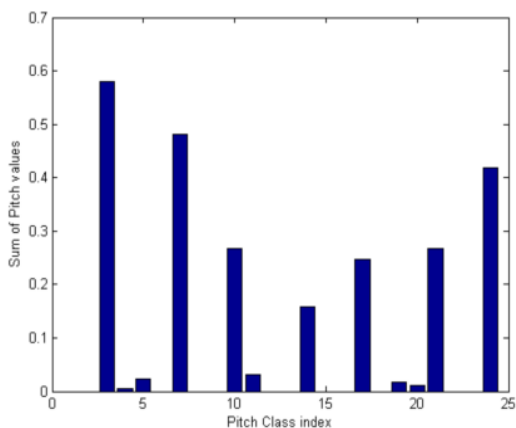
طول ۴۰ ثانیه و دسته سوم قطعات با طول ۶۰ ثانیه. مدت زمان انجام محاسبات برای هر قطعه توسط هر دو الگوریتم اندازه گیری و ثبت شد که در جدول ۳ نشان داده شده است. برای دسته اول (۲۰ ثانیه) میانگین زمان انجام محاسبات برای روش ما ۰.۶۳۹۶ ثانیه بود در حالی که برای روش Zhu این زمان به ۶.۵۰۵۲ ثانیه رسید. برای دسته دوم میان زمان انجام محاسبات برای روش ما ۰.۶۸۶۴ ثانیه و برای روش Zhu به ۱۷.۷۹۶۱ ثانیه رسید. برای دسته سوم این اندازه گیری ها به ترتیب ۰.۷۱۷۶ ثانیه برای روش ما و ۳۳.۶۸۰۶ ثانیه برای روش Zhu بود. همانطور که در جدول نیز قابل مشاهده است روش پیشنهادی سرعت محاسبات را با نسبت  $۱۰.۱۷ = ۰.۶۳۹۶ / ۶.۵۰۵۲$  ارتقا بخشید. برای دسته دوم و سوم این نسبت ها به ترتیب ۲۵.۹۲ و ۴۶.۹۳ بدست می آیند. به علاوه اینکه با مقایسه زمان انجام محاسبات برای هر کدام از سه دسته فوق با استفاده از روش Zhu مشاهده می شود که اگر چه طول قطعات با نسبت ثابتی افزایش یافته اند، مدت زمان انجام محاسبات به صورت تصاعدی زیاد می شوند. در حالی که در روش پیشنهادی افزایش طول قطعه هیچ گونه اثری در مدت زمان انجام محاسبات نداشته و زمان بسیار ناچیزی است. بنابر این می توان استنباط کرد که برای قطعات با طول زیاد روش پیشنهادی نتایج قابل توجهی خواهد داشت.



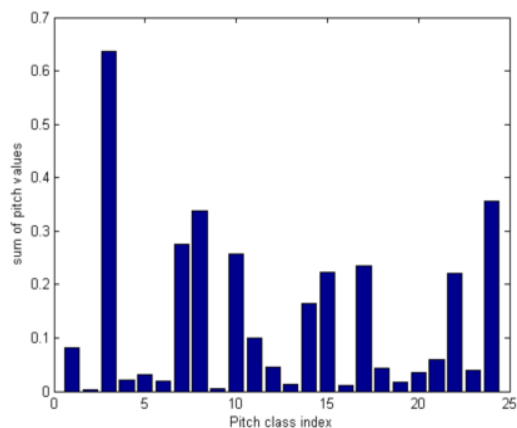
ب



الف



د



ج

شکل ۵-۷: پروفایل های بدست آمده در هر سه روش به همراه پروفایل ایده آل برای یک قطعه موسیقی ایرانی نواخته شده در مقام سه گاه. الف: پروفایل ایده آل، ب: پروفایل روش حیدریان، ج: پروفایل روش Zhu، د: پروفایل روش پیشنهادی.

۲۰ ثانیه	۴۰ ثانیه	۶۰ ثانیه	
۰.۶۳۹۶	۰.۶۸۶۴	۰.۷۱۷۶	روش پیشنهادی
۶۵.۰۵۲	۱۷.۷۹۶۱	۳۳.۶۸۰۶	روش Zhu
۱۰.۱۷	۲۵.۹۲	۴۶.۹۳	نسبت روش دوم به اول

### ۵-۴-۳ مقایسه سوم:

برای ارزیابی کارایی روش تشخیص مقام پیشنهادی، فاصله مینکوفسکی بین نمونه تست و تمام مدل های موجود محاسبه می شود. در نهایت آن مقامی که کمترین مقدار فاصله برای آن بدست آمده به عنوان مقام قطعه در نظر گرفته می شود. فاصله مینکوفسکی از رابطه ۵-۱۲ به صورت زیر محاسبه می شود:

$$D(a, b) = \left[ \sum_{k=1}^d (a_k - b_k)^2 \right]^{1/m} \quad (۵-۱۲)$$

که در آن  $D(a, b)$  فاصله مینکوفسکی از مرتبه  $m$  بین بردارهای  $d$  بعدی  $a$  و  $b$  است. انتخاب  $m = 2$  فاصله اقلیدسی را خواهد داد. ما در اینجا از فاصله مینکوفسکی مرتبه ۴ استفاده کرده ایم. اگر پیچ های یک پروفایل را شیفته دهیم برای هر کدام از پروفایل ها ۲۴ حالت مختلف خواهیم داشت که هر کدام بیانگر یک مقام جداگانه اند. با توجه به اینکه ۴ گام مختلف برای موسیقی ایرانی وجود دارد جمعا  $4 * 24 = 96$  فاصله جداگانه برای بررسی هر نمونه تست بایستی محاسبه شود. نتایج تشخیص مقام با استفاده از هر سه روش مورد بحث در جدول ۵-۴ آورده شده است. برای روش پیشنهادی و روش Zhu نتایج تقریباً یکسانند. به طور متوسط، ۷۱٪ از مقام قطعه ها به طور صحیح شناسایی شده اند. در واقع می توان نتیجه گرفت که الگوریتم پیشنهادی علاوه بر اینکه صحت و دقت روش Zhu را داراست توانسته است الگوریتم را بهینه کرده و سرعت آن را نیز تا حد چشمگیری افزایش دهد. مقایسه روش پیشنهادی با روش حیدریان افزایش ۱۳.۶٪ در روش پیشنهادی را نشان می دهد.

از آنجا که روش حیدریان بر اساس دیتابیس برای ساز سنتور بود، ما قطعات مربوط به ساز سنتور را از دیتابیس خودمان جدا کردیم تا دو الگوریتم را برای تنها ساز سنتور نیز تست کنیم. چنان که در جدول ۵-۵ نشان داده شده است روش پیشنهادی ما برای ساز سنتور نیز نتایج بهتری دارد.

جدول ۴-۵: مقایسه سه الگوریتم مورد بحث از لحاظ درصد دقت	
نتایج نهایی	
٪۷۱.۳۵	روش پیشنهادی
٪۷۱.۲۴	روش Zhu
٪۵۷.۷۵	روش حیدریان

جدول ۵-۵: مقایسه الگوریتم پیشنهادی با الگوریتم حیدریان برای ساز سنتور	
ساز سنتور	
٪۸۱.۸۱	روش پیشنهادی
٪۷۴.۴۱	روش حیدریان

در آزمایشی دیگر صحت الگوریتم روش پیشنهادی برای انواع سازها به طور جداگانه نیز بررسی شد که نتایج آن در جدول ۵-۶ آورده شده است. بهترین نتیجه مربوط می‌شود به ساز نی با ٪۹۰ صحت و بدترین آن برای ساز ویلن با ٪۵۴.۵۴. البته لازم به ذکر است که عدم در دسترس بودن دیتای کافی برای ساز ویلن و کمانچه موجب پایین آمدن صحت الگوریتم شده است.

جدول ۵-۶: نتایج الگوریتم پیشنهادی برای ساز های مختلف							
سه تار	گروه نوازی	کمانچه	نی	سنتور	تار	ویلن	
۶۴.۵۱	۶۶.۶۷	۶۴.۷	۹۰	۸۱.۸۱	۷۲.۸۸	۵۴.۵۴	درصد صحت

## ۵-۵ نتیجه گیری:

در این فصل روشی ارائه شد که قادر بود سیگنال صوتی یک قطعه موسیقی ایرانی را دریافت کرده و مقام آن را به درستی تشخیص دهد. در ادامه تئوری موسیقی، متفاوت با آنچه که تاکنون برای موسیقی ایرانی گفته شده است، برای کاهش فاصله بین جنبه های عملی و نظری کار پیشنهاد شد. علاوه بر همه اینها یک دیتابیس، شامل اجراهای مختلف موسیقی ایرانی، انواع مقام های موجود در موسیقی ایرانی که توسط سازهای مختلفی نواخته شده اند تدوین و ارائه شد. همچنین تعداد سازها و نمونه هایی که برای کلاس های مختلف انتخاب شده بودند نیز مشخص گردید.

این الگوریتم که بر پایه تبدیل CQT می باشد، روش ارائه شده در [۱۸] را بسط داده، بهینه کرده و نهایتا پارامترهای لازم را بر مفاهیم موسیقی ایرانی منطبق کرده است. به علاوه اینکه، در بخش استخراج ویژگی، با اضافه کردن مرحله کاهش بعد، محاسبات را کاهش داده و موجب افزایش قابل توجه سرعت الگوریتم گردید.

مقایسه پروفایل های پیچ، شباهت قابل توجه پروفایل حاصل از روش پیشنهادی به پروفایل واقعی را در مقایسه با دو روش دیگر به خوبی نشان داد. نتایج نشان دادند که علاوه بر اینکه استفاده از جنبه استاتیک صوت و حذف متغیر زمان از محاسبات، موجب کاهش محاسبات و افزایش سرعت گردید، همچنان الگوریتم این قابلیت خود را داشت که کماکان دقت الگوریتم پایه، که آنالیز را به صورت دینامیک انجام می داد، حفظ کند. در مقایسه با روش حیدریان، نتایج به طور واضح بهبود یافتند، به علاوه اینکه روش پیشنهادی موفق شد دیتابیس کامل تری را که شامل سازهای گوناگون و اجراهای مختلف است اثناء کند.

## پیشنهاد برای کارهای آینده:

با توجه به کارهای انجام شده و تجربیات برست آمده پیشنهادهای برای ادامه کار به شرح زیر خلاصه می‌شود:

۱- ارائه روش‌های متفاوتی جهت بهینه‌سازی، ارتقاء بخش استخراج ویژگی و تعیین پیچ میزان‌سازی

۲- جمع‌آوری دیتابیس کامل و جامع تر برای دستیابی به یک مجموعه سیستماتیک و هماهنگ قابل استفاده در آنالیزهای بازیابی اطلاعات موسیقی ایرانی.

۳- اعمال بخش کاهش بعد به الگوریتم‌های موجود برای موسیقی غربی و مقایسه نتایج آن برای دیتابیس موسیقی کلاسیک و پاپ غربی.

## مراجع

- [1] "Musimathic, The Mathematical Foundations of Music", Gareth Loy, The MIT Press, 2006.
- [۲] "مبانی نظری موسیقی ایرانی"، حسین علیزاده، هومان اسعدی، مینا افتاده و ... انتشارات ماهور، چاپ اول ۸۸.
- [3] Schouten, J. F., R. J. Ritsma, and B. Lopes Cardozo. 1962. "Pitch of the Residue." *Journal of the Acoustical Society of America* 34: 1418–1424.
- [۴] "نظری به موسیقی"، روح الله خالقی، رهروان پویش، ۱۳۸۶.
- [5] E. Gavahian and H. Marvi, "Mode Detection of Persian Music from Musical Audio", submitted for publication, for Digital Signal Processing, Elsevier.
- [6] M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney, Content-based music information retrieval: current directions and future challenges, *Proc. IEEE* 96 (4) (2008) 668–696.
- [7] J. Downie. Music information retrieval. *Annual Review of Information Science and Technology*, 37:295–340, 2003a.
- [8] N. Orio. Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval*, 1(1):1–90, 2006.
- [9] Music Reporting Metadata Set, EBU – TECH 3332, Geneva July 2009.
- [10] P. Polotti and D. Rocchesso, "**Sound to Sense, Sense to Sound**", Copyright Logos Verlag Berlin GmbH 2008
- [11] A. de Cheveigne, H. Kawahara, Yin: a fundamental frequency estimator for speech and music, *Journal of the Acoustic Society of America* 111 (2002) 1917–1930.
- [12] T. Tolonen and M. Karjalainen, "Acomputationally efficient multipitch analysis model," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 708–716, Nov. 2000.
- [13] M. Davy and S. J. Godsill, "Bayesian harmonic models for musical signal analysis," in *Bayesian Statistics VII*. Oxford, U.K.: Oxford Univ. Press, 2003.
- [14] M. Goto, "A predominant-F0 estimation method for real-world musical audio signals: MAP estimation for incorporating prior knowledge about F0s and tone models," in *Proc. Workshop on Consistent and Reliable Acoustic Cues for Sound Analysis*, Aalborg, Denmark, Sep. 2001.
- [15] A. Klapuri, T. Virtanen, and J.-M. Holm, "Robust multipitch estimation for the analysis and manipulation of polyphonic musical signals," in *Proc. COST-G6 Conf. on Digital Audio Effects*, Verona, Italy, 2001, 2000.

- [16] B. Bozkurt, An automatic pitch analysis method for Turkish maqam music, *Journal of New Music Research* 37 (1) (2008) 1–13.
- [17] A. Krishnaswamy, Application of pitch tracking to South Indian classical music, in: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 19–22 October 2003, pp. 49–.
- [18] Y. Zhu, M.S. Kankanhalli, Precise pitch profile feature extraction from musical audio for key detection, *IEEE Transactions on Multi- media* 8 (3) (2006) 575–584.
- [19] B.S. Ong, E. Gómez, S. Streich, Automatic extraction of musical structure single pitch class distribution features, in: *Proceedings of the workshop on Learning the Semantics of Audio Signals (LSAS)*, Athens, Greece, 6 December 2006, pp. 53–65.
- [20] D. Temperley, Pitch-class distribution and the identification of key, *Music Perception* 25 (3) (2008) 193–212.
- [21] B. Bozkurt, An automatic pitch analysis method for Turkish maqam music, *Journal of New Music Research* 37 (1) (2008) 1–13.
- [22] A. C. Gedik and B. Bozkurt, “Pitch-frequency histogram-based music information retrieval for Turkish music,” *Elsevier, Digital Signal Processing*, 90 (2010) 1049-1063
- [23] E. Gomez, P. Herrera, Comparative analysis of music recordings from western and non-western traditions by automatic tonal feature extraction, *Empirical Musicology Review* 3 (3) (2008) 140–156.
- [24] E. Gómez, A. Klapuri and B. Meudic, “Melody Description and Extraction in the Context of Music Content Processing,” accepted for publication, 2002
- [25] W. Hess. *Pitch Determination of Speech Signals. Algorithms and Devices*. Springer Series in Information Sciences. Springer-Verlag, Berlin, New York, Tokyo, 1983.
- [26] P. Cano. Fundamental frequency estimation in the SMS analysis. In *Proceedings Digital Audio Effects Conference*, 1998.
- [27] A. Klapuri. *Signal Processing Methods for the Automatic Transcription of Music*. Unpublished PhD thesis, Tampere University of Technology, Tampere, Finland, 2004.
- [28] R. Maher and J. Beauchamp. Fundamental frequency estimation of musical signals using a two-way mismatch procedure. *Journal of the Acoustical Society of America*, 95:2254–2263, 1993.
- [29] R. Meddis and M. Hewitt. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America*, 89(6):2866–2882, 1991.



- [30] B. Gold and L. Rabiner. Parallel processing techniques for estimating pitch periods of speech in the time domain. *Journal of the Acoustical Society of America*, 46:442–448, 1969.
- [31] J. C. Brown, “Musical fundamental frequency tracking using a pattern recognition method” *J. Acoust. Soc. Am.* 92(3), September 1992.
- [32] B. Bozkurt, “An automatic pitch analysis method for Turkish maqam music”, *Journal of New Music Research* 37 (1) (2008) 1–13.
- [33] P. Heydarian, L. Jones and A. Seago “The Analysis and Determination of the Tuning System in Audio Musical Signal” AES 123<sup>rd</sup> Convention, New York, NY, USA, 2007 October 5-8.
- [۳۴] موسیقی نظری، علی نقی وزیری، تهران، ۱۳۱۹.
- [۳۵] مبانی آکوستیک، مترجم: مهدی برکشلی، امیر کبیر ۱۳۸۲.
- [36] H. Farhat, “The Dastgah Concept in Persian Music,” Cambridge University Press, 1990.
- [۳۷] “نگرشی نو به تئوری موسیقی ایران”، داریوش طلایی، ماهور.
- [۳۸] “کتاب سرایش: آموزش سلفژ ایرانی”، حسین مهرانی، سوره مهر، ۱۳۸۴.
- [39] <http://www.mahoor.com/Default.aspx>
- [40] <http://www.shayda.info/>
- [41] <http://www.chaharbaghbang.com/>
- [42] Y. Zhu and M. Kankanhalli , “Music scale modeling for melody matching”, *Proc. ACM Multimedia03*, Nov. 2003
- [۴۳] پیمان حیدریان، تشخیص نت موسیقی برای ساز سنتور، پایان نامه، دانشگاه تربیت مدرس، استاد راهنما دکتر کبیر، ۱۳۷۹

## Abstract:

The concept of intervals in Persian music and its effects on music classification and configuration is one of the most important issues considered by musicians. It has also been taken into account in signal processing applications to achieve automatic analytic methods. Mode (scale) detection from music data directly relates to the concept of intervals. One of the most applicable features, used in mode detection systems, is pitch. Such a melodic feature, besides other melodic and rhythmic features, is used for classification and different musical similarity related analysis in the area of Music Information Retrieval (MIR).

In this thesis, a novel algorithm for recognizing different modes of Persian music is proposed. For mode detection, a 24-pitch profile feature, which contains essential information about the mode of the piece of music, has been extracted from the musical audio. Moreover, for feature extraction, a dimension reduction step was proposed, which obviously reduced the calculations and improved the speed of the algorithm. Finally the experiments on the collected database showed promising results for Persian music mode detection.

Keywords: Persian music, Maqam, Music Information Retrieval, Intervals, Pitch profile.



**Shahrood University of Technology**  
**Faculty of Electrical and Robotic Engineering**

## **Persian Music Classification using Pitch Profile Feature**

By  
**Ebrahim Gavahian**

Under supervision of  
**Dr. Hossein Marvi**

Consultant  
**Dr. Ali Soleimani**

July 2010

