



دانشگاه صنعتی شاهرود
دانشکده برق و ریاضیات

پایان نامه کارشناسی ارشد

گرایش الکترونیک

استخراج ویژگی برای تشخیص آواها به کمک تبدیل میلین

نگارش:

مهدی جماعتی

استاد راهنما:

دکتر حسین مروی

استاد مشاور:

دکتر علیرضا احمدی فرد

تابستان ۸۷

به نام خدا

دانشگاه صنعتی شاهرود

دانشکده برق و رباتیک

رساله کارشناسی ارشد

عنوان : استخراج ویژگی برای تشخیص آواها به کمک تبدیل میلین

نگارش: مهدی جماعتی

کمیته ممتحنین:

استاد راهنما : دکتر حسین مروی

امضا :

استاد مشاور: دکتر علیرضا احمدی فرد

امضا :

استاد مدعو: دکتر حسن زاده و دکتر سلیمانی

امضا :

تاریخ :

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

چکیده

سیستم شنوایی انسان قادر است، آواهای یکسان تولید شده توسط افراد مختلف با جنسیت و سن متفاوت را شناسایی نماید. اما انجام اینکار برای یک ماشین با شرط داشتن دقت بالا کاری دشوار است. تبدیلات بسیار زیادی با خصوصیات متفاوت پیشنهاد گردیده‌اند که در آنها هدف ساده‌تر نمودن محاسبات و ارائه ویژگیهای قدرتمندتر می‌باشد. در بین آنها، تبدیل میلین، و نوع محدود شده آن که آنرا تبدیل مقیاس می‌نامند، می‌تواند یک سیگنال را بر حسب مقیاس بیان نمایند. ویژگی کلیدی تبدیل میلین ثابت ماندن اندازه تبدیل نسبت به تغییرات مقیاس است.

در این رساله الگوریتم استخراج ویژگی جدیدی برای مدلسازی و تشخیص آواها با استفاده از تلفیق تبدیل مقیاس و روشهای معروف استخراج ویژگی معرفی شده است. نتایج تجربی نشان می‌دهد که ویژگیهای جدید دارای دقت و کارایی بالاتری در مقایسه با ویژگیهای استخراج شده از تبدیل میلین و دیگر روشها برای شناسایی آواها است.

کلمات کلیدی: تبدیل مقیاس، استخراج ویژگی، پوش طیفی، آواها

تشکر و قدردانی

در ابتدا لازم می‌دانم از زحمات استاد راهنمایم، دکتر حسین مروی، کمال تشکر و قدردانی را داشته باشم. تجربیات، ایده‌ها و پیشنهادات ایشان نقش مهمی در آزمایشات و بررسیهای من داشته است. در انتها، این رساله را تقدیم به پدر و مادر عزیزم می‌کنم. امیدوارم انجام این رساله بتواند گوشه‌ای از زحمات تمامی اعضای خانواده‌ام را جبران نماید.

فهرست مطالب

ا	چکیده
ب	تشکر و قدردانی
ت	فهرست مطالب
ح	فهرست اشکال
ر	فهرست جداول

۱- بیان مسئله

- ۱-۱- مقدمه ۱
- ۲-۱- نگاهی اجمالی به فصل‌ها ۳
- ۳-۱- مقالات استخراج شده از رساله ۴

۲- معرفی تبدیل میلین

- ۱-۲- مقدمه ۶
- ۲-۲- تعریف تبدیل میلین ۷
- ۳-۲- محدوده تعریف تبدیل میلین ۸
- ۴-۲- رابطه تبدیل میلین با تبدیلات دیگر ۹
- ۵-۲- فرمول پرسوال و ویژگیهای مربوط به آن ۱۱
- ۶-۲- تابع گاما ۱۲
- ۷-۲- تابع PSI ۱۴
- ۸-۲- نماد POCHHAMMER ۱۵
- ۹-۲- انتگرالهای میلین - بارنز ۱۵
- ۱۰-۲- خواص تبدیل میلین ۱۶
- ۱۱-۲- مراجع ۲۱

۳- تبدیل میلین سریع

- ۱-۳- مقدمه ۲۳
- ۲-۳- محاسبه تبدیل میلین توابع پیوسته ۲۳
- ۳-۳- محاسبه تبدیل میلین توابع گسسته ۲۴
- ۱-۳-۳- تئوری نمونه برداری ۲۶
- ۲-۳-۳- نمونه برداری نمایی از تابع گسسته ۳۰
- ۳-۳-۳- درون‌یابی ۳۳
- ۴-۳-۳- اجرا و نتایج ۳۴
- ۴-۳-۳- مراجع ۳۶

۴- کاربردهای تبدیل میلین

- ۱-۴- مقدمه ۳۸
- ۲-۴- کاربرد تبدیل میلین در سیگنالهای صوتی ۳۹
- ۱-۲-۴- تبدیل میلین زمان کوتاه ۴۰

۴۱ ۲-۲-۴- فیلتر کردن در حوزه میلین
۴۳ ۳-۲-۴- بازیابی سیگنال با استفاده از دامنه یا فاز
۴۴ ۴-۲-۴- کاهش نویز در سیگنالهای صوتی با استفاده از تبدیل میلین
۴۶ ۳-۴- تبدیل میلین دو بعدی و کاربردهای آن
۴۹ ۱-۳-۴- فیلتر نمودن تصاویر در حوزه میلین
۴۹ ۲-۳-۴- کاهش نویز در تصاویر
۵۱ ۳-۳-۴- تشخیص الگوهای نامتغیر با چرخش، مقیاس و انتقال
۵۴ ۴-۴- مراجع

۵- سیگنالهای گفتار و سیستمهای تشخیص اتوماتیک گفتار

۵۶ ۱-۵- مقدمه
۵۹ ۲-۵- اصوات گفتار و ویژگیهای آنها
۶۱ ۱-۲-۵- آواها
۶۴ ۲-۲-۵- صداهای ترکیبی
۶۵ ۳-۲-۵- شبه آواها
۶۷ ۴-۲-۵- حروف بیصدای تو دماغی
۶۸ ۵-۲-۵- اصوات سایشی بیصدا
۷۰ ۶-۲-۵- اصوات سایشی صدا دار
۷۱ ۷-۲-۵- مکشهای بیصدا و صدا دار
۷۲ ۳-۵- روشهای تشخیص اتوماتیک گفتار با استفاده از ماشین
۷۵ ۱-۳-۵- روش واج شناسی شنیداری تشخیص گفتار
۷۷ ۱-۱-۳-۵- دسته کننده آوایی واجهای صوتی
۷۹ ۲-۱-۳-۵- دسته بندی اصوات گفتار
۸۰ ۳-۱-۳-۵- مثالهای برجسب زنی واج شناسی صوتی
۸۱ ۴-۱-۳-۵- مشکلات روش واج شناسی شنیداری
۸۳ ۲-۳-۵- روش تشخیص الگوی آماری برای تشخیص گفتار
۸۵ ۳-۳-۵- روشهای هوش مصنوعی برای تشخیص گفتار
۸۷ ۴-۳-۵- شبکه های عصبی و کاربرد آنها در تشخیص گفتار
۹۴ ۴-۵- مراجع

۶- الگوریتم پیشنهادی برای شناسایی آواها و نتایج عملی

۹۵ ۱-۶- مقدمه
۹۶ ۲-۶- رابطه آواها و افراد مختلف

۹۹ ۳-۶- مراحل اجرای الگوریتم پیشنهادی
۱۰۰ ۱-۳-۶- تحلیل طیفی
۱۰۱ ۱-۱-۳-۶- کپستروم
۱۰۵ ۲-۱-۳-۶- روش ابداعی کپستروم بهبود یافته
۱۰۹ ۲-۳-۶- تبدیل میلین و تبدیل مقیاس
۱۱۴ ۳-۳-۶- محاسبه ضرایب MFCC و PLP
۱۱۸ ۱-۳-۳-۶- ضرایب طیف فرکانسی مل
۱۲۲ ۲-۳-۳-۶- محاسبات خطی ادراکی
۱۲۷ ۴-۳-۶- نتایج
۱۳۰ ۵-۳-۶- طبقه‌بندی ویژگیهای استخراج شده
۱۳۵ ۴-۶- مقایسه‌ای بین روشهای مختلف استخراج ویژگی
۱۳۷ ۵-۶- نتیجه‌گیری و کارهای آینده
۱۴۰ ۶-۶- مراجع
۱۴۲ Abstract

فهرست اشکال

۱۰ ۱-۲- مسیر انتگرالگیری در رابطه معکوس تبدیل میلین یک خط عمودی در صفحه مختلط p است
۱۴ ۲-۲- تابع $\Gamma(p)$ (بالا) و $1/\Gamma(p)$ (پایین) بر حسب تابعی از p (حقیقی و $-4 < p < 4$) ...
۲۵ ۱-۳- بلوک دیاگرام تبدیل میلین تابع پیوسته
۲۵ ۲-۳- نمایش دو بعدی تابع انتقال محور x
۲۶ ۳-۳- مراحل الگوریتم تبدیل میلین گسسته

۲۸	طیف فرکانسی سیگنالهای آنالوگ و دیجیتال	۴-۳
۲۹	تابع سینک، پاسخ ضربه فیلتر بازیابی ایده‌آل	۵-۳
۳۲	نمونه‌برداری یکنواخت و نمونه‌برداری دوباره بصورت نمایی از آن	۶-۳
۴۲	مدل زمانی جمله "your dark suit in greasy" و اسپکتروگرام آن	۱-۴
	مدل زمانی جمله فیلتر شده "your dark suit in greasy" و اسپکتروگرام آن که	۲-۴
۴۲	توسط یک فیلتر پایین گذر در حوزه میلیون صورت گرفته است	
۴۳	اسپکتروگرام سیگنال فیلتر شده پایین گذر (بالا) و بالا گذر (پایین) [۱]	۳-۴
۴۳	اسپکتروگرام سیگنال اصلی [۱]	۴-۴
۴۴	سیگنال بازیابی شده با استفاده از فاز (بالا) و با استفاده از دامنه (پایین)	۵-۴
	فیلترنمودن در حوزه میلیون در مقایسه با فیلتر نمودن در حوزه فوریه. (a) سیگنال	۶-۴
	اصلی (b) سیگنال نویز با توان نویزی ۳ دبی (c) کاهش نویز با استفاده از فیلتر	
	کردن سیگنال نویزی در حوزه فوریه (MSE=2.065*10 ⁻³). (d) اختلاف بین	
	سیگنال فیلتر شده در حوزه فوریه و سیگنال اصلی. (e) کاهش نویز با استفاده از	
	همان فیلتر در حوزه میلیون (MSE=1.59*10 ⁻³). (f) اختلاف بین سیگنال فیلتر	
۴۵	شده در حوزه میلیون و سیگنال اصلی	
	مراحل عملکرد تبدیل مقیاس (a) تصویر اصلی، (b) انتقال از حوزه دکارتی به	۷-۴
	لگاریتمی - قطبی، (c) دامنه تبدیل مقیاس، (d) تبدیل معکوس مقیاس. افزایش	
۴۸	رزولوشن فرکانسی بعلت نمونه برداری لگاریتمی از تصویر اصلی است. [۳]	
	(a) تصویر فیلتر پایین گذر شده در حوزه مقیاس (شعاع ۳۲) (b) فیلتر شده با باند	۸-۴
	میانی (شعاعهای ۱۶ و ۳۲) (c) فیلتر شده بالاگذر (مقیاسهای انرژی بیشتر از ۳۲).	
	(d-e) مشابه با قسمت بالا ولی ابعاد شعاع دو برابر شده است. عملیات فیلتر کردن با	
۵۰	حذف مقیاسهایی که خارج محدوده مورد نظر ماست، صورت گرفته است [۵]	
	فیلتر نمودن با استفاده از تبدیل مقیاس و مقایسه آن با تبدیل فوریه. (a) تصویر	۹-۴
	نویزی شده گوسی، (b) فیلتر نمودن با استفاده از یک فیلتر پایین گذر در حوزه	
۵۱	فوریه، (c) فیلتر کردن با همان فیلتر در حوزه مقیاس	
	دو مثال از فیلترنمودن در حوزه مقیاس در مقایسه با فیلتر نمودن در حوزه فوریه. (a)	۱۰-۴
	تصویر Cameraman با نویز گوسی. (b) فیلترنمودن در حوزه فوریه با استفاده از فیلتر	
	پایین گذر Butterworth. (شعاع ۳۲ پیکسل) (c) فیلترنمودن در حوزه مقیاس با فیلتر	
	قسمت b. کاهش نویز با استفاده از قسمت نمودن تصویر. (d) تصویر با نویز بسیار	
	بالا (e) کاهش نویز تصویر با استفاده از فیلتر Gabor در حوزه فوریه (۴ فرکانسهای و	
	۴ جهتی) (f) کاهش نویز تصویر با استفاده از حوزه مقیاس (چهار مقیاس و ۴ جهتی)	
۵۲	[۵]	

۱۱-۴	تجسم نويز در حوزه مقياس (a) تصوير اصلي (b) تبديل مقياس شكل (a)، طيف براي درك بهتر بصورت لگاريتمي مقياس بندي شده است. (c) تصوير بعد از حذف نمودن پيكهاي مربوط به (d) تبديل مقياس تصوير (a) بعد از حذف نمودن پيكها [۴]	۵۲
۱۲-۴	بلوك دياگرام تبديل فوريه - ميلين	۵۳
۱-۵	نمودار طبقه بندي واجهاي استاندارد انگليسي آمريكايي به كلاسهاي صوتي [۱]	۶۰
۲-۵	نمودار زماني چند آواي مختلف	۶۲
۳-۵	اسپكتروگرام آواهاي شكل ۵-۲	۶۳
۴-۵	فرکانسهاي اندازه گيري شده اولين و دومين فرکانس رزونانس براي رنج وسيعي از گويندها به ازاي چندين آواي مختلف [۴]	۶۳
۵-۵	مثلت آوايي كه هر نقطه بيانگر ميانگين فرکانس رزونانس هر آوا، ادا شده توسط افراد مختلف است [۴]	۶۴
۶-۵	نمودار زماني مربوط به اصوات تركيبي /a ^w /، /a ^y /، /o ^y / و /a ^y /	۶۶
۷-۵	اسپكتروگرام اصوات تركيبي شكل ۵-۶	۶۶
۸-۵	تغييرات اولين دو فرکانس رزونانس در اصوات تركيبي [۵]	۶۷
۹-۵	نمودار زماني مربوط به اصوات /m/، /n/ و /ŋ/	۶۸
۱۰-۵	اسپكتروگرام به اصوات /m/، /n/ و /ŋ/	۶۸
۱۱-۵	نمودار زماني (چپ) و اسپكتروگرام (راست) اصوات سايشي بيصدای /ʃ/ و /s/، /f/ و /sh/	۶۹
۱۲-۵	نمودار زماني و اسپكتروگرام اصوات سايشي صدا دار /v/ و /z/	۷۰
۱۳-۵	نمودار زماني شكل موج رشته ادا شده /i-b-u/	۷۱
۱۴-۵	مقايسه نمودار زماني و اسپكتروگرام اصوات مكثي صدا دار و بيصدای /p/، /b/ و /t/	۷۲
۱۵-۵	شبكة واجي براي يك سرى كلمه	۷۳
۱۶-۵	بلوك دياگرام سيستم تشخيص صوت با استفاده از روش واج شناسي شنيداري ...	۷۶
۱۷-۵	دسته بندي آواي صوت شناسي	۷۸
۱۸-۵	دسته بندي اصوات گفتار بوسيله درخت دودوبي	۷۹
۱۹-۵	جداسازي و برچسب زني جمله "seven-six"	۸۲
۲۰-۵	جداسازي و برچسب زني جمله "did you"	۸۲
۲۱-۵	بلوك دياگرام تشخيص دهنده گفتار با استفاده از روش تشخيص الكو	۸۴
۲۲-۵	روش پايين- بالا براي جمع آوري اطلاعات سيستم تشخيص گفتار	۸۸
۲۳-۵	روش بالا- پايين براي جمع آوري اطلاعات سيستم تشخيص گفتار	۸۸
۲۳-۵	روش تخته سياه براي جمع آوري اطلاعات سيستم تشخيص صوت [۶]	۸۹
۲۴-۵	بلوك دياگرام سيستم شنوايي گوش انسان	۹۰
۲۵-۵	مولفه هاي محاسباتي شبكه عصبي تاخير زماني [۷]	۹۱
۲۶-۵	ساختار داخلي TDNN براي تشخيص /g/ و /b/، /d/ [۷]	۹۱

۹۲ ترکیب شبکه عصبی و فیلتر تطبیقی برای تشخیص گفتار [۸]	۲۷-۵
۹۳ مثالی برای درک بهتر ترکیبی از شبکه عصبی با مجموعه‌ای از فیلترهای تطبیقی	۲۸-۵
۹۳ شبکه عصبی کنترل پنهانی [۹]	۲۹-۵
۹۸ نمودار زمانی چند آوای مختلف	۱-۶
 مقایسه‌ای بین نمودار زمانی آوای یکسان ادا شده توسط یک فرد در زمانهای	۲-۶
۹۸ مختلف و افراد مختلف	
۱۰۲ بلوک دیاگرام مراحل محاسبه روش کپستروم حقیقی	۳-۶
 نمودار سیگنال پنجره‌بندی شده، طیف فرکانسی بر حسب دبی (طول تبدیل فوریه	۴-۶
 ۱۰۲۴)، کپستروم حقیقی، کپستروم پنجره‌بندی شده (پارامتر قطع برابر ۲۰) و پوش	
۱۰۴ طیفی	
 پوش طیفی آوای "aa" شکل ۶-۲ با استفاده از روش کپستروم به ازای پارامتر	۵-۶
۱۰۶ قطع برابر با ۲۰ و استفاده از رابطه (۶-۶)	
۱۰۶ مقایسه پوش طیفی آوای "aa" ادا شده توسط افراد مختلف	۶-۶
۱۰۷ مقایسه پوش طیفی به ازای تغییرات پارامتر قطع با وزندار نمودن یکسان	۷-۶
 مقایسه پوش طیفی به ازای تغییرات پارامتر وزندار نمودن با پارامتر یکسان ($N_1=12$)	۸-۶
۱۰۷	
۱۰۸ بلوک دیاگرام الگوریتم محاسبه پوش طیفی با استفاده از کپستروم بهبود یافته	۹-۶
۱۰۹ پوش طیفی شکل ۲-۶ با استفاده از روش کپستروم بهبود یافته به ازای $\gamma=10$	۱۰-۶
 نمودار زمانی آوای /IY/ که بوسیله افراد مختلف ادا شده است آواها برگرفته از [۴]	۱۱-۶
۱۱۲ می‌باشند	
 پوش طیفی استخراج شده با استفاده از کپستروم بهبود یافته برای سیگنالهای شکل	۱۲-۶
۱۱۲ ۱۱-۶	
۱۱۳ تبدیل مقیاس منحنی‌های شکل ۶-۱۲	۱۳-۶
۱۱۳ تبدیل مقیاس به ازای مشتق مرتبه سوم شکل ۶-۱۲	۱۴-۶
۱۱۴ نمودار زمانی آوای /ao/ که توسط افراد مختلف ادا شده است	۱۵-۶
۱۱۵ پوش طیفی استخراج شده از روش کپستروم بهبود یافته متناظر با آوای شکل ۱۵	۱۶-۶
۱۱۵ دامنه تبدیل مقیاس مشتق مرتبه سوم شکل ۶-۱۶	۱۷-۶
۱۱۶ نمودار زمانی آوای /er/ و /aa/, /iy/, /ae/, /uh/, /eh/	۱۸-۶
۱۱۶ پوش طیفی آوای مختلف بدست آمده از روش کپستروم بهبود یافته	۱۹-۶
۱۱۷ دامنه تبدیل مقیاس مشتق مرتبه سوم پوش طیفی آوای شکل ۶-۱۹	۲۰-۶
 بلوک دیاگرام خلاصه شده الگوریتم پیشنهادی برای استخراج ویژگی و سپس	۲۱-۶
۱۱۷ دسته‌بندی آواها	
۱۱۹ بلوک دیاگرام پروسه محاسبه پارامترهای MFCC	۲۲-۶

۲۳-۶	چهار حالت مختلف برای بانک فیلتر مل. به مقیاسهای مختلف محورهای فرکانسی
۱۲۲	هر نمودار دقت شود
۲۴-۶	بلوک دیاگرام مراحل الگوریتم PLP
۲۵-۶	مجموعه‌ای از بارک‌ها به ازای فرکانس نمونه‌برداری ۸۰۰۰ و تعداد فیلتر ۱۷ با ۱۲۸=
۱۲۶ NFFT
۲۶-۶	اسپکتروگرام سیگنال اصلی (بالا) و اسپکتروگرام بازیابی شده با استفاده از ضرایب
۱۲۶ PLP
۲۷-۶	بردار ویژگی آوای /IY/ با استفاده از الگوریتم پیشنهادی به همراه (الف) روش MFCC
۱۲۸ (ب) روش PLP
۲۸-۶	بردار ویژگی آواهای مختلف با استفاده از الگوریتم پیشنهادی به همراه (الف) روش
۱۲۹ MFCC (ب) روش PLP
۲۹-۶	مقایسه‌ای بین درصد تشخیص نهایی سیستم به ازای روشهای مختلف برای ۵ کلاس
۱۳۶ آوایی
۳۰-۶	مقایسه‌ای بین درصد تشخیص نهایی سیستم به ازای روشهای مختلف برای ۸ کلاس
۱۳۶ آوایی
۳۱-۶	تغییرات دقت تشخیص سیستم به ازای تغییرات پارامتر γ

فهرست جداول

۱۷	توابع انتخاب شده $f(x)$ ، با تبدیل میلین آنها و محدوده تعریف تبدیل ...	۱-۲
۲۰	قضایای تبدیل میلین	۲-۲
۶۰	لیستی خلاصه شده از علامت اختصاری واجهای آمریکایی - انگلیسی	۱-۵
۶۵	فرکانسهای رزونانس چند آوا	۲-۵
۱۳۰	تعداد نمونه های استخراج شده در هر مرحله از الگوریتم پیشنهادی	۱-۶
	نتایج آزمایشات انجام شده با استفاده از الگوریتم پیشنهادی به همراه روش	۲-۶
۱۳۳	MFCC برای ۵ کلاس مختلف	

۳-۶	نتایج آزمایشات انجام شده با استفاده از الگوریتم پیشنهادی به همراه روش
۱۳۳ PLP برای ۵ کلاس مختلف
۴-۶	نتایج آزمایشات انجام شده با استفاده از الگوریتم پیشنهادی به همراه روش
۱۳۴ MFCC برای ۸ کلاس مختلف
۵-۶	نتایج آزمایشات انجام شده با استفاده از الگوریتم پیشنهادی به همراه روش
۱۳۴ PLP برای ۸ کلاس مختلف

Abstract

Human auditory system can recognized the same vowel pronounced by different people with different gender, different age. However this is a difficult task for machine to do it accurate. Many other transforms with different properties have been devised in order to make certain operations easier or certain features more easily visible. Among these, the Mellin Transform, and its restricted version called the Scale Transform, can represent a signal in terms of *scale*. The key property of the scale transform is the scale invariance.

This thesis proposed a new feature extraction algorithm for vowel modeling and recognition using scale transform and one of the known methods for feature extraction such as Mel frequency Cepstral Coefficient (MFCC). Experimental results indicate that the new features produced better accuracy than original feature extracted from the Mellin transform and other methods in application of vowel recognition.

Key words: scale transform, feature extraction, spectral envelope, vowels

فصل اول

بیان مسالہ

۱-۱- مقدمه

ارتباط گفتاری حالت بارزی از رد و بدل شدن اطلاعات بین انسانهاست. بعنوان یک شرط برای ارتباط بین انسانها، گفتار نقش مهمی را ایفا می‌کند. با گسترش علم و فناوری نیاز به ارتباط با ماشینهای مختلف برای انسانها ضروری بنظر می‌رسید. بعنوان نمونه چه بسیار افرادی هستند که تمایل دارند سرعت تایپ کردن آنها همانند سرعت صحبت کردن آنها باشد. از اینرو دانشمندان و محققین به دنبال یک زبان ارتباطی بین انسان و ماشین هستند تا از این طریق انسان بتواند همانگونه که با انسان دیگری ارتباط برقرار می‌سازد، با یک ماشین نیز رابطه برقرار کند و بدین وسیله بتواند مقداری از کارهای خود را به ماشین محول سازد.

هدف اصلی سیستمهای تشخیص گفتار اتوماتیک افزایش توانایی ماشین برای تشخیص و شناسایی جملات ادا شده از یک سیگنال صوتی است. هدف نهایی و غایی در سیستمهای تشخیص گفتار اتوماتیک، قادر ساختن ماشین برای اجرای این کار بزرگ بخوبی انسان و حتی بهتر از انسان است. بدست آوردن ماشینی با چنین قابلیتی آرزوی بشر در قرن اخیر بوده است، اما اجرا و تحقق آن در چهل و پنجاه سال اخیر، آن هم با گسترش تکنولوژی کامپیوترها ممکن شده است.

در طی گسترش سیستمهای تشخیص گفتار اتوماتیک، نتایج بسیار مهمی بدست آمد. اندازه واژگان از چند جمله به چندین میلیون و از سیستمهای وابسته به گوینده به سیستمهای مستقل از گوینده ارتقاء یافته‌اند، ولی هنوز چالشهای بزرگی در این رابطه وجود دارد. از تغییر پذیری سیگنال صوتی بعنوان یکی از

فاکتورهای بسیار مهم در عقب ماندگی این تکنولوژی می‌توان نام برد. هم گوینده و هم محیط می‌تواند باعث چنین تغییراتی گردد. بعنوان نمونه نویز یکی از عوامل تغییر سیگنالهای صوتی است.

تشخیص اتوماتیک گفتار بعلاوه تغییرات بسیار وسیع موجود در آن، مانند لهجه یا طرز تلفظ، میزان بلندی صدا، شرایط سلامتی و خستگی فرد، کاری بسیار دشوار است. در نتیجه در سیستمهای تشخیص گفتار اتوماتیک، برای کاربردهای خاص یا کلی، جنبه‌های زیادی از این تغییرات باید مد نظر قرار گیرد. بهمین علت تقسیم بندیهای زیادی برای طراحی یک سیستم تشخیص گفتار اتوماتیک بوجود می‌آید. در حالت کلی اصوات به چهار دسته تقسیم‌بندی می‌شوند که یکی از آنها آواها می‌باشد. بی شک، یکی از مهمترین کلاسهای صوتی در زبان انگلیسی و یا آمریکایی، آواها می‌باشند. اصولاً، آواهای تولید شده با توجه به موقعیت حفره دهان و زبان کوچک تعیین می‌شوند، اما موقعیت فک و لبها نیز می‌تواند در آوای ایجاد شده تاثیرگذار باشد. چندین روش برای توصیف و طبقه‌بندی آواها وجود دارد که از آن جمله می‌توان به نمودارهای زمانی و طیف فرکانسی اشاره نمود.

در این رساله، هدف ما استخراج یک سری ویژگیها از سیگنالهای آوا است، که برای اجرای آن از تلفیق تبدیل میلین و الگوریتمهای معروف استخراج ویژگی استفاده شده و یک الگوریتم جدید معرفی شده است. این ویژگیها بایستی اولاً مستقل از سن، جنسیت و نوع گوینده باشد و ثانياً دارای کارایی و دقت بسیار بالایی بوده تا بکمک آنها بتوان آواها را از یکدیگر متمایز نمود. برای این منظور ما در ابتدا سیگنالهای صوتی را به حوزه فرکانس انتقال می‌دهیم و برای انجام اینکار از روش کپستروم بهبود یافته که تعمیم یافته روش کپستروم معمولی است استفاده می‌کنیم. در روش کپستروم، هدف یافتن پوش طیفی یک سیگنال است. پوش طیفی هموار شده طیف فرکانسی سیگنال (اندازه تبدیل فوریه آن) است که از هرگونه تغییرات شدید جلوگیری می‌نماید، در عین اینکه سعی می‌کند شکل کلی طیف را حفظ نماید. در واقع، ما بدنبال پوش طیفی یک سیگنال بعنوان نشانه‌ای برای شناسایی و طبقه‌بندی آواها هستیم. آواها دارای طیف فرکانسی هارمونیک می‌باشند که پوش طیفی بر روی آن قرار می‌گیرد. همین ویژگی باعث می‌شود بتوان براحتی آواها را از بقیه اصوات جدا نمود.

در ادامه قصد داریم تا با محاسبه تبدیل مقیاس پوش طیفی بدست آمده و استفاده از روشهای معروف استخراج ویژگی، ویژگیهای لازم را از درون پوش طیفی سیگنال خارج سازیم. آزمایشات انجام شده نشان دادند که ویژگیهای استخراج شده از آواها (سیگنالهای ورودی) با توجه به الگوریتمهایی که برای دسته بندی

آواها از یکدیگر مورد استفاده قرار گرفته است، دارای کارایی بسیار بالایی برای تشخیص آواها از یکدیگر است.

۱-۲- نگاهی اجمالی به فصل‌ها

- **فصل دوم** در این فصل به معرفی تبدیل میلین و انواع مختلف آن می‌پردازیم و در مورد شرایط وجود و ویژگیهای این تبدیل صحبت می‌کنیم. همچنین در ادامه به یک سری از توابع که رابطه مستقیمی با تبدیل میلین دارند، اشاره می‌کنیم. با بدست آوردن رابطه تبدیل میلین با تبدیلات دیگر نظیر فوریه و لاپلاس سعی در ارائه الگوریتمی مناسب جهت محاسبه تبدیل میلین و انواع آن مخصوصاً تبدیل مقیاس داریم. در نهایت در این فصل مجموعه‌ای از ویژگیهای مهم و پرکاربرد تبدیل میلین به همراه تبدیل میلین مجموعه‌ای از توابع آورده شده است.

- **فصل سوم** یکی از مشکلاتی که در گذشته استفاده از تبدیل میلین را در کاربردهای مختلف محدود نموده است، فقدان یک الگوریتم مناسب برای محاسبه تبدیل میلین سیگنالها مخصوصاً سیگنالهای گسسته بوده است. در این فصل الگوریتم مناسبی برای محاسبه تبدیل میلین سیگنالهای گسسته ارائه شده است که به کمک آن می‌توان تبدیل میلین سیگنالهای گسسته را محاسبه نمود. این الگوریتم دارای دقت و سرعت بسیار بالایی می‌باشد و می‌توان بعنوان یکی از بهترین الگوریتمها برای محاسبه تبدیل میلین گسسته از آن نام برد. تمام مراحل این الگوریتم در محیط مطلب پیاده‌سازی شده است.

- **فصل چهارم** با توجه به ویژگیهای منحصر بفرد تبدیل میلین و نوع محدود شده آن یعنی تبدیل مقیاس، کاربردهای بسیار متنوعی را برای این تبدیل در زمینه تشخیص الگو می‌توان نام برد. در این فصل در ابتدا یک سری از کاربردهای تبدیل میلین در حوزه سیگنالهای یک بعدی و مخصوصاً در پردازش سیگنالهای صوتی آورده شده است. ضمناً در این فصل دو کاربرد جدید تبدیل میلین در زمینه حذف نویز از سیگنالهای صوتی و مبحث تشخیص امضا معرفی شده است. در ادامه این فصل در مورد کارایی این تبدیل در پردازش تصاویر دیجیتال از جمله حذف نویز از تصاویر، مثالهایی آورده شده است.

- **فصل پنجم** بررسی اصول سیستمهای تشخیص اتوماتیک گفتار و همچنین مقایسه این روشها با یکدیگر در این فصل انجام می‌گیرد. در ابتدای این فصل در مورد اصوات گفتار، تقسیم‌بندیهای انجام شده برای طبقه‌بندی و جداسازی اصوات از یکدیگر و ویژگیهای هر یک از آنها بطور کامل توضیح داده می‌شود. در ادامه این فصل در مورد مهمترین روشهایی که در یک سیستم تشخیص اتوماتیک گفتار بکار می‌رود، صحبت می‌کنیم و مزایا و معایب هر روش را بیان می‌کنیم.
- **فصل ششم** در این فصل بطور کامل در مورد الگوریتم پیشنهادی خود که برای استخراج ویژگی از آواها مورد استفاده قرار می‌گیرد، بحث شده است و بصورت قدم به قدم مراحل اجرای الگوریتم و اثرات آنرا با توجه به آزمایشات صورت گرفته مورد بررسی قرار می‌دهیم و نتایج آزمایشگاهی هر مرحله آورده شده است. در نهایت، در مورد مزایا و معایب الگوریتم پیشنهادی و کارهای آتی صحبت نموده و سپس کارایی الگوریتم پیشنهادی با الگوریتمهای معروف استخراج ویژگی مقایسه شده است.

۱-۳- مقالات استخراج شده از رساله

- [1] Mahdi Jamaati, Hossein Marvi, "Spectral Speech Representation by Modified Mellin Transform," International Conference Electrical Engineering (ICEE), 6-10 July 2008, Okinawa, Japan.
- [2] Mahdi Jamaati, Milad Lankarany and H. Marvi, "On the Use of Scale Transform in Modeling of the Spectral Envelope of Vowels," Acoustics08, June 29-July 4, 2008, France.
- [3] Mahdi Jamaati, M.lankarany and H. Marvi, "Feature extraction for Vowels recognition using Mellin transform," The Journal of the Society of America (JASA), vol 123, No 5, pt.2, May 2008, Page 3732.
- [4] Mahdi Jamaati and H. Marvi, "Vowels Recognition Using Mellin Transform and PLP-based Feature Extraction," The Journal of the Society of America (JASA), vol 123, No 5, pt.2, May 2008, Page 3177.

فصل دوم

معرفی تبدیل میلین

۲-۱- مقدمه

تبدیل میلین^۱ یک تبدیل کاملاً معروف در آنالیز الگوریتم‌هاست. این تبدیل بسیار شبیه تبدیلات لاپلاس و فوریه می‌باشد و با استفاده از همین روابط سعی در نوشتن برنامه‌ای مناسب جهت بکار بردن آن در برنامه‌هایی مانند مطلب داریم. این تبدیل در اواسط دهه ۶۰ میلادی برای حل یک سری مشتقات جزئی آورده شد و اصل کاربرد آن در مبحث تئوری اعداد می‌باشد. اخیراً تبدیل میلین کاربردهای فراوانی در تئوری اطلاعات پیدا نموده است. پرکاربرد بودن این تبدیل از دو ویژگی بسیار مهم آن ناشی می‌شود. این تبدیل امکان کاهش معادلات برخی توابع را به حالت جبری می‌دهد و نگاشت مستقیمی بین بسطهای مجانب تابع نزدیک صفر یا بینهایت و مجموعه نقاط تکین تبدیل در صفحه مختلط ایجاد می‌کند [۱].

همچنین، تبدیل میلین در مهندسی برق، بعنوان نمونه برای بررسی و تحلیل سیستمهای کنترل موتور استفاده می‌شود. تبدیل میلین را نامتغیر نسبت به اندازه^۲ می‌نامند. این عبارت بدین معنی است که اگر دو سیگنال از لحاظ مقیاس (انبساط یا فشردگی) تابع با ثابت نگه داشتن انرژی سیگنال اصلی) با هم فرق داشته باشند، اندازه تبدیل میلین دو تابع با هم برابر است. در این فصل قصد داریم تا در مورد برخی از خصوصیات تبدیل میلین صحبت کنیم.

۲-۲- تعریف تبدیل میلین

فرض کنید تابع حقیقی $f(x)$ به ازای مقادیر مثبت x تعریف شده است. تبدیل میلین تابع $f(x)$ با $M_f(p)$ نشان داده می‌شود. تبدیل میلین این تابع بصورت زیر تعریف می‌شود:

$$M_f(p) = \int_0^{\infty} x^{p-1} f(x) dx \quad (۱-۲)$$

که در رابطه فوق $p = -jc + \beta$ تعریف شده و عددی مختلط است و آنرا پارامتر تبدیل میلین می‌نامیم. معکوس تبدیل میلین نیز از رابطه زیر بدست می‌آید:

$$f(t) = \int_{c-i\infty}^{c+i\infty} M_f(p) t^{-p} dp \quad (۲-۲)$$

^۱ . Mellin transform
^۲ . Scale invariant

در حالت کلی، در صورتی بازه‌ی پایینی x برابر با صفر تعریف می‌شود که قسمت حقیقی متغیر p بزرگتر از مقدار معینی باشد و بهمین ترتیب در صورتی بازه بالای x را برابر با بینهایت قرار می‌دهیم که قسمت حقیقی p کوچکتر از مقدار مشخصی باشد. به این ترتیب برای متغیر تبدیل میلین یک بازه بدست می‌آید و آنرا اصطلاحاً محدوده تعریف تبدیل می‌نامند که در ادامه راجع به آن بحث می‌شود. بنابراین اگر تبدیل میلین تابع $f(x)$ به ازای تمام مقادیر x تعریف شده باشد، بر روی محور عمود بر صفحه مختلط p قرار می‌گیرد. بنابراین با در نظر گرفتن این شرایط برای تابع $f(x)$ می‌توان نشان داد که تبدیل میلین تابع $f(x)$ یا همان $M_f(p)$ یک تابع تحلیلی از p و قرار گرفته بر روی این محور است [۶].

وجود تبدیل میلین رابطه (۲-۱) به همگرایی انتگرال زیر بستگی دارد:

$$\int_0^{\infty} |f(t)| t^{p-1} dt < \infty \quad (2-3)$$

این یک شرط کلی برای وجود تبدیل میلین است. در صورتیکه $p = -jc + \beta$ در نظر گرفته شود، شرایط بیشتری را می‌توان بدست آورد، در عین اینکه معادله نیز ساده‌تر می‌شود.

پارامتر β به ما این امکان را می‌دهد تا با انتخاب مقادیر مختلف برای آن، تبدیلات مختلف با ویژگیهای متفاوت داشته باشیم. بعنوان نمونه اگر $\beta = 1/2$ باشد، تبدیل مقیاس^۳ ایجاد می‌شود که در آن دامنه تبدیل مقیاس توابع $f(t)$ و $g(t) = \sqrt{\alpha} f(\alpha t)$ با هم برابر است. اگر $\beta = 0$ باشد، تبدیلی بدست می‌آید که نسبت به فشردگی یا انبساط محور زمان بدون تغییر است. با این تبدیل دو تابع $f(t)$ و $g(t) = f(\alpha t)$ به یک تابع نگاشت داده می‌شوند. بهمین علت این مجموعه از تبدیلات را تبدیل β میلین می‌نامند.

۲-۳- محدوده تعریف تبدیل میلین

تبدیل میلین را هیچگاه نباید بدون در نظر گرفتن محدوده تعریف آن محاسبه نمود. این محدوده مکان و موقعیت همگرایی تابع را مشخص می‌کند. این محدوده نقش مهمی در پروسه تبدیل معکوس میلین دارد، مخصوصاً هنگامیکه از این تبدیل در تئوری اعداد استفاده گردد، این اهمیت بیشتر نیز خواهد شد. ایده اصلی، محاسبه تبدیل میلین یک سری و معکوس آن و در نهایت بدست آوردن یک بسط مجانب است.

³ . Scale transform

تبدیل معکوس میلین بر روی خط موازی با محور موهومی که در داخل محدوده تعریف قرار گرفته است، بدست می‌آید. بدون دانستن موقعیت محدوده تعریف، نمی‌توان انتگرال را محاسبه نمود، بعبارت دیگر نمی‌توان مشخص نمود که مقدار کدام مانده‌ها باید محاسبه گردد [۲].

محدوده تعریف تبدیل میلین از خصوصیت‌های همگرایی انتگرال این تبدیل نشئت گرفته است. با توجه به رابطه (۲-۱) با تقسیم نمودن انتگرال به دو قسمت، مشابه زیر

$$M_f(p) = \left(\int_0^1 + \int_1^\infty \right) x^{p-1} f(x) dx \quad (۲-۴)$$

با فرض اینکه تابع $f(x)$ بر روی محور حقیقی و مثبت انتگرال‌پذیر باشد، اولین انتگرال باید به صفر محدود گردد و دومی در بینهایت. با جایگزاری $p = -jc + \beta$ داریم:

$$\left| \int_0^1 x^{p-1} f(x) dx \right| \leq \int_0^1 x^{\beta-1} |f(x)| dx \quad (۲-۵)$$

و

$$\left| \int_1^\infty x^{p-1} f(x) dx \right| \leq \int_1^\infty x^{\beta-1} |f(x)| dx \quad (۲-۶)$$

حال با در نظر گرفتن اینکه $f(x)$ یک تابع چند جمله‌ای از درجه u در حول و حوش صفر باشد، $(f(x) = O(x^u))$ ، انتگرال اول در صورتی همگرا خواهد شد که

$$\beta + u - 1 > -1 \quad \rightarrow \quad \beta > -u \quad (۲-۷)$$

بهمین ترتیب فرض کنید که در بینهایت $f(x) = O(x^v)$. انتگرال دومی همگرا خواهد بود در صورتیکه

$$\beta + v - 1 < -1 \quad \rightarrow \quad \beta < -v \quad (۲-۸)$$

این دو شرط بیان‌کننده دو نیم صفحه می‌باشند. اولی نیم صفحه سمت چپ و دومی نیم صفحه سمت راست است. فصل مشترک این دو نیم صفحه، محدوده تعریف تبدیل است که آنرا با $(-u, -v)$ نشان می‌دهیم. خیلی اوقات اتفاق می‌افتد که تابع تصویر در کل صفحه پیوسته و تحلیلی است که این امر به ما

امکان محاسبه انتگرال معکوس را با استفاده از انتقال خط انتگرالگیری به سمت راست یا چپ، می‌دهد. شایان ذکر است که میزان جابجایی محدود شده به محدوده تعریف تبدیل است.

بطور خلاصه، اگر $f(x)$ تابعی انتگرال‌پذیر در طول محور حقیقی و مثبت باشد و

$$f(x)_{x \rightarrow 0^+} = O(x^u) \quad \text{و} \quad f(x)_{x \rightarrow \infty} = O(x^v) \quad (۹-۲)$$

در اینصورت تبدیل میلین آن در داخل محدوده تعریف $(-u, -v)$ همگرا بوده و انتگرال معکوس میلین آن بر روی خط موازی با محور موهومی این محدوده، بدست می‌آید.

۲-۴ - رابطه تبدیل میلین با تبدیلات دیگر

با توجه به تعریف تبدیل میلین، رابطه بسیار نزدیکی بین این تبدیل و تبدیل فوریه برقرار است. با جایگزاری $p = -jc + \beta$ در رابطه (۲-۱) و استفاده از تغییر متغیر $x = e^u$ داریم:

$$\begin{aligned} M_f(-jc + \beta) &= \int_0^{\infty} f(x) x^{-jc+\beta-1} dx = \int_0^{\infty} f(x) e^{(-jc+\beta-1)\ln x} dx \\ &= \int_{-\infty}^{\infty} f(e^u) e^{(-jc-1)u} e^{u\beta} (e^u du) = \int_{-\infty}^{\infty} e^{x\beta} f(e^x) e^{-jcx} dx \end{aligned} \quad (۱۰-۲)$$

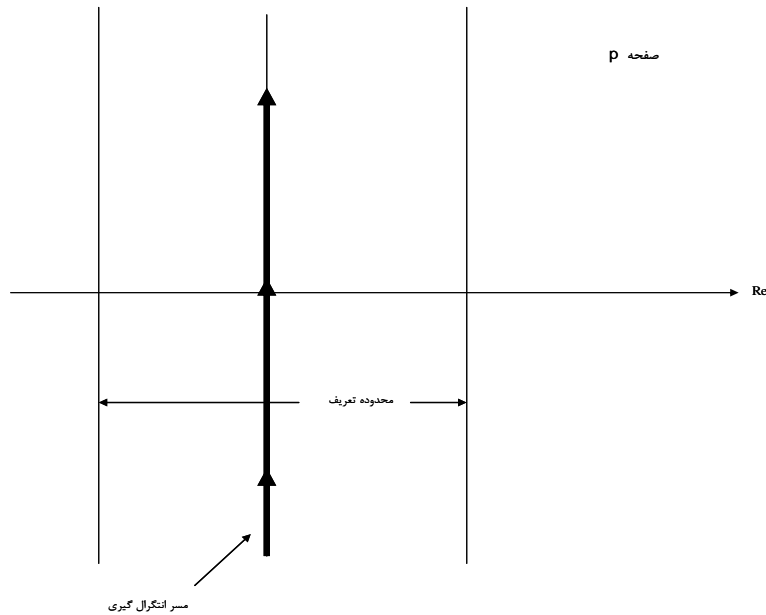
با مقایسه رابطه فوق با رابطه تبدیل فوریه به این نتیجه می‌رسیم که تبدیل میلین تابع $f(x)$ در واقع برابر با تبدیل فوریه تابع $g(x)$ است که در آن $g(x) = e^{\beta x} f(e^x)$ است. بعبارت دیگر

$$M_{f(x)}(p = -jc + \beta) = F\{e^{\beta x} f(e^x)\} \quad (۱۱-۲)$$

همچنین با توجه به تعریف تبدیل لاپلاس، می‌توان نشان داد که

$$M_{f(x)}(p) = L\{f(e^{-x})\} \quad (۱۲-۲)$$

در نتیجه، ویژگیهای زیادی را می‌توان با استفاده از روابط تعریف شده بین تبدیل میلین با تبدیلهای فوریه و لاپلاس بیان نمود [۳].



شکل ۲-۱. مسیر انتگرال‌گیری در رابطه معکوس تبدیل میلین یک خط عمودی در صفحه مختلط p است

با استفاده از فرمول معکوس لاپلاس یا فوریه می‌توان تبدیل معکوس میلین رابطه (۲-۲) را بدست آورد که مسیر انتگرال‌گیری، همانگونه که در شکل (۲-۱) نشان داده شده است، خط عمودی محور مختلط p است. می‌توان اثبات کرد که رابطه معکوس تبدیل میلین به ازای $M_f(p)$ داده شده، تابعی یکتا بدست می‌آورد [۷].

همانطور که گفته شد، پارامتر تبدیل میلین یا همان p یک پارامتر مختلط است. در صورتیکه p را بصورت $p = -jc + \beta$ نشان دهیم با فرض اینکه $\beta_L < \beta < \beta_R$ باشد، مستقیماً با استفاده از تعریف (۲-۱) می‌توان ویژگیهای اولیه زیر را برای تبدیل میلین بدست آورد.

$$\begin{aligned} M\{f(\alpha x)\} &= \alpha^{-p} M\{f(x)\} & \beta_L < \beta < \beta_R, \quad \alpha > 0 \\ M\{x^\alpha f(x)\} &= M_f(p + \alpha) & \beta_L - \text{Re}\{\alpha\} < \beta < \beta_R - \text{Re}\{\alpha\} \end{aligned} \quad (۲-۱۳)$$

9

$$M\{f(x^\alpha)\} = \frac{1}{|\alpha|} M_f\left(\frac{p}{\alpha}\right), \quad \begin{cases} \alpha \beta_L < \beta < \alpha \beta_R, & \text{if } \alpha > 0 \\ \alpha \beta_R < \beta < \alpha \beta_L, & \text{if } \alpha < 0 \end{cases} \quad (۲)$$

(۱۴)

۲-۵- فرمول پارسوال و ویژگیهای مربوط به آن

حاصلضرب تبدیل فوریه یا لاپلاس دو تابع، با استفاده از کانولوشن دو تابع بدست می‌آید. بهمین ترتیب

برای تبدیل میلین حاصلضرب دو تابع داریم:

$$\int_0^{\infty} g(y)b(y)y^{p-1}dy = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} M_g(\omega)M_b(p-\omega)d\omega \quad (2-15)$$

که در رابطه فوق c متعلق به $M_g(\omega)$ است. دوباره سمت راست بصورت کانولوشن در آمده است. برای نشان دادن درستی رابطه (۲-۱۵)، می‌توان از سمت چپ معادله شروع نمود و با استفاده از رابطه (۲-۲) برای معرفی $M_g(\omega)$ و عوض نمودن مرتبه انتگرال، آنرا اثبات نمود.

یکی از روابط ساده‌ای که با استفاده از رابطه (۲-۱۵) بدست می‌آید، رابطه زیر است:

$$\int_0^{\infty} g(xy)b(y)y^{p-1}dy = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} M_g(\omega)M_b(p-\omega)x^{-\omega}d\omega, \quad x > 0 \quad (2-16)$$

که رابطه فوق از ترکیب روابط (۲-۱۳) و (۲-۱۵) بدست آمده است. در حالت خاصی که در رابطه (۲-۱۶) ، $p=1$ باشد، داریم:

$$\int_0^{\infty} g(xy)b(y)dy = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} M_g(p)M_b(1-p)x^{-p}dp \quad (2-17)$$

که در رابطه فوق c متعلق به دو بازه تعریف توابع $M_g(p)$ و $M_b(1-p)$ است. برای برقراری رابطه (۲-۱۷) لازم است که این دو بازه روی یکدیگر قرار گیرند. دقت داشته باشید در حالتیکه در رابطه فوق ، $x=1$ باشد رابطه بدست آمده را رابطه پارسوال تبدیل میلین می‌نامند.

رابطه (۲-۱۷) هسته روش تبدیل میلین را شکل می‌دهد. عملیات انجام شده در سمت چپ معادله را کانولوشن میلین دو تابع $g(x)$ و $b(x)$ می‌نامند. فرق اساسی این رابطه با رابطه‌های مشابه کانولوشن این است که در این رابطه، حاصلضرب x و y را داریم، نه تفاضل آنها را. با استفاده از ویژگی رابطه معکوس، رابطه (۲-۱۷) را می‌توان بصورت زیر بازنویسی نمود:

$$M_{(conv(g,b))} = M_g(p)M_b(1-p) \quad (2-18)$$

در سمت راست معادله (۲-۱۸) ما حاصلضرب تبدیل میلین دو تابع را داریم (یکی از آنها معکوس شده و سپس انتقال یافته است) و در نتیجه رابطه (۲-۱۸) معکوس رابطه (۲-۱۵) می‌باشد. بعبارت دیگر رابطه فوق بیان می‌کند که تبدیل میلین کانولوشن دو تابع در حوزه x برابر با حاصلضرب تبدیل میلین این دو تابع، در حوزه p است [۶].

۲-۶- تابع گاما

تابع گاما $\Gamma(p)$ را می‌توان بصورت تبدیل میلین تابع e^{-x} بیان نمود. معمولاً هنگامیکه از تبدیل میلین استفاده می‌شود این تابع بکار می‌رود. طبق تعریف داریم:

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx, \quad \beta > 0 \quad (2-19)$$

در رابطه فوق لازم است که قسمت حقیقی پارامتر تبدیل میلین (β) بزرگتر از صفر باشد تا انتگرال فوق در $x=0$ همگرا گردد. این محدودیت بدین معنی است که بازه تعریف تبدیل میلین در اینحالت کل نیم صفحه مختلط سمت راست p است. می‌توان نشان داد که

$$\Gamma(p) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{1}{p+n} + \int_1^{\infty} x^{p-1} e^{-x} dx, \quad p \neq 0, -1, -2, \dots \quad (2-20)$$

رابطه فوق با تقسیم نمودن رابطه (۲-۱۹) به $\int_0^1 + \int_1^{\infty}$ و باز نمودن e^{-x} با استفاده از سری تیلور و انتگرالگیری جزء به جزء بدست آمده است. در رابطه بدست آمده (۲-۲۰) فرض بر این است که β بزرگتر از صفر است. بهر حال سمت راست معادله فوق به ازای تمام مقادیر $p = 0, -1, -2, \dots$ تجزیه‌پذیر است.

تابع گاما دارای ویژگیهای زیادی است. ما از چند نمونه آن برای تبدیل میلین استفاده می‌کنیم. از رابطه (۲-۲۰) می‌توان فهمید که در نقاط $p = 0, -1, -2, \dots$ تابع گاما دارای قطبهای ساده با مانده‌های

$$1, -1, \frac{1}{2}, \dots, \frac{(-1)^n}{n!}, \dots$$

رابطه بازگشتی

$$\Gamma(p+1) = p\Gamma(p) \quad (2-21)$$

که براحتی می توان رابطه فوق را با انتگرالگیری جزء به جزء از رابطه (۲-۱۹) بدست آورد. به ازای $\Gamma(1) = 1$ چنین بر می آید که

$$\Gamma(n+1) = n!, \quad n = 0, 1, 2, \dots \quad (2-22)$$

می توان نشان داد که

$$\Gamma(p)\Gamma(1-p) = \frac{\pi}{\sin \pi p} \quad (2-23)$$

که آنرا رابطه انعکاس می نامند. با توجه به نتیجه رابطه (۲-۲۳)، $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. با رابطه بازگشتی (۲-۲۱)، می توان مقادیر $\Gamma(\frac{3}{2}), \Gamma(\frac{5}{2}), \dots$ را تعیین نمود. یکی دیگر از نتایج رابطه (۲-۲۳) این است که $1/\Gamma(p)$ بر روی تمام صفحه p تحلیلی است. بنابراین، $\Gamma(p)$ هیچ صفری در صفحه p ندارد.

رابطه تکثیر یا تکرار بصورت زیر بیان می شود:

$$\Gamma(2p) = \frac{1}{2\sqrt{\pi}} 2^{2p} \Gamma(p)\Gamma(p + \frac{1}{2}) \quad (2-24)$$

که آنرا می توان بوسیله رابطه ضرب برای $\Gamma(3p), \Gamma(4p), \dots$ تعمیم داد.

$$\Gamma(np) = \frac{1}{\sqrt{n(2\pi)^{n-1}}} n^{np} \prod_{l=0}^{n-1} \Gamma(p + \frac{l}{n}), \quad n = 1, 2, \dots \quad (2-25)$$

در سمت راست رابطه (۲-۲۵) دقت داشته باشید که تمام ضرایب p متعلق به آرگومان توابع گاما، برابر یک است. در نهایت، رابطه استرلینگ

$$\Gamma(p) = \sqrt{\frac{2\pi}{p}} \left(\frac{p}{e}\right)^p \left[1 + O\left(\frac{1}{p}\right)\right], \quad |p| \rightarrow \infty \text{ with } |\arg p| < \pi \quad (2-26)$$

یک تقریب مجانب برای $\Gamma(p)$ به ازای آرگومانهای مختلط و بزرگ است. برای اینکه این رابطه به ازای $\beta < 0$ برقرار باشد، رابطه (۲-۲۶) را با رابطه (۲-۲۳) ترکیب می کنیم [۵].

نمودار $\Gamma(p)$ و $1/\Gamma(p)$ برای حالتیکه قسمت حقیقی پارامتر مستقل تبدیل میلین داده شده باشد، در شکل ۲-۲ نشان داده شده است.

۲-۷-۲ تابع PSI (مشتق لگاریتمی تابع گاما)

در تبدیل میلین، خیلی اوقات ممکن است با $\Gamma'(p)$ سر و کار داشته باشیم [۱۰]. این مقدار با استفاده از $\Gamma(p)$ و تابع PSI، یا $\psi(p)$ بدست می‌آید که بصورت زیر تعریف می‌شود:

$$\psi(p) = \frac{\Gamma'(p)}{\Gamma(p)} = \frac{d}{dp} \ln \Gamma(p) = -\gamma + \sum_{n=0}^{\infty} \left(\frac{1}{n+1} - \frac{1}{p+n} \right) \quad (27-2)$$

فرمول بازگشتی برای تابع psi، که نتیجه ساده‌ای از رابطه (۲-۲۱) است، برابر است با

$$\psi(p+1) = \frac{1}{p} + \psi(p) \quad \text{-2}$$

(28)

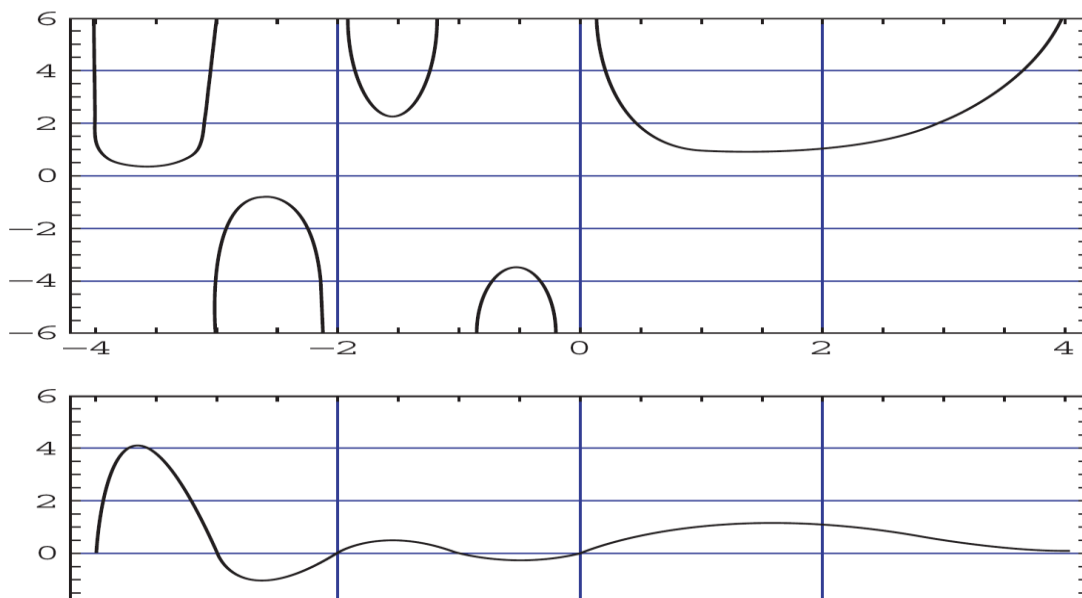
با استفاده از رابطه (۲-۲۸) می‌توان نشان داد که

$$\psi(n+1) = \psi(1) + \left(1 + \frac{1}{2} + \dots + \frac{1}{n}\right), \quad n = 1, 2, \dots \quad (29-2)$$

مقدار $\psi(1)$ را می‌توان از رابطه زیر بدست آورد:

$$-\psi(1) = \gamma = 0.5772156649 \quad (30-2)$$

که γ ثابت اولر است. در نتیجه معادله (۲-۲۹) به ما امکان محاسبه $\psi(2), \psi(3), \dots$ را می‌دهد.



شکل ۲-۲. تابع $\Gamma(p)$ (بالا) و $1/\Gamma(p)$ (پایین) بر حسب تابعی از p (حقیقی و $-4 < p < 4$)

۲-۸- نماد POCHHAMMER

نماد POCHHAMMER یا $(p)_n$ بصورت زیر بیان می‌شود [۱۲]:

$$(p)_n = \frac{\Gamma(p+n)}{\Gamma(p)}, \quad n = 0, 1, 2, \dots \quad (۲-۳۱)$$

رابطه فوق در رابطه بازگشتی زیر صدق می‌کند:

$$(p)_{n+1} = (p+n)(p)_n, \quad n = 0, 1, 2, \dots \quad (۲-۳۲)$$

۲-۹- انتگرالهای میلین - بارنز

هنگامیکه یک انتگرال را با استفاده از تبدیل میلین محاسبه می‌کنیم، در ابتدا بایستی تبدیلات میلین توابع قرار گرفته در انتگرال را بدست آوریم. معمولاً این امر با استفاده از برنامه‌های قدرتمندی مانند مطلب انجام می‌گیرد یا اینکه از جداول منتشر شده تبدیل میلین استفاده می‌گردد. در جدول (۲-۱)، چند نمونه تابع $f(x)$ ، تبدیل میلین آن، و محدوده تعریف آن نشان داده شده است [۸].

هدف بعدی تبدیل میلین، یافتن انتگرال میلین - بارنز جهت محاسبه انتگرال است. بهمین علت در اینجا قصد داریم تا در مورد انتگرال میلین - بارنز صحبت نماییم. نکته اولی را که می توان با مشاهده جدول ۲-۱ فهمید این است که هر تبدیل میلین را می توان بصورت حاصلضرب نوشت که در آن ضرایب بفرم $\Gamma(a+Ap)^{-1}, \Gamma(a+Ap)$ یا α^{-p} در می آید که تمام ضرایب A حقیقی می باشند. برای درک بهتر، فرض کنید که آنرا ضرایب استاندارد بنامیم. توابع زیر انتگرال انتگرالهای معکوس، برابر با حاصلضرب ضرایب استاندارد در x^{-p} می شوند. بعنوان نمونه، با توجه به رابطه (۱۱) جدول ۲-۱، فرمول تبدیل معکوس (۲-۲) و خاصیت (۲-۱۳) داریم:

$$\left(\frac{\sin 3x}{3x}\right)^2 = -\frac{\sqrt{\pi}}{4} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\Gamma(\frac{p}{2}-1)}{\Gamma(\frac{3}{2}-\frac{p}{2})} (3x)^{-p} dp, \quad 0 < \beta < 2 \quad (2-33)$$

انتگرالهای همگرای مشابه سمت راست معادله (۲-۳۳) را انتگرالهای میلین - بارنز می نامند. این انتگرالها برای ما بسیار مهم می باشند، چرا که بسیاری از توابع $f(x)$ را می توان بصورت انتگرالهای میلین - بارنز یا بصورت ترکیب خطی از انتگرالهای میلین - بارنز نوشت. بعبارت دیگر، تبدیلات میلین آنها یا $M_f(p)$ را می توان بصورت ترکیبات خطی از حاصلضربهای استاندارد نوشت.

حاصلضرب استاندارد را می توان بسیاری از مواقع بصورت غیر استاندارد نوشت. با توجه به اینکه انتگرالهای میلین - بارنز برای ما مهم می باشند، بهتر است در صورت امکان از ضرایب استاندارد استفاده گردد. این امر موجب می گردد که جداول منتشر شده تبدیل میلین، یکسان گردند [۱۸].

۲-۱۰- خواص تبدیل میلین

در جدول (۲-۲) برخی از خصوصیات پرکاربرد تبدیل میلین نشان داده شده است. روابط فوق را می توان براحتی با استفاده از تعریف تبدیل میلین رابطه (۲-۱) بدست آورد. بعنوان مثال برای رابطه ۱۲ جدول (۲-۲) داریم:

$$M\left[x \frac{d}{dx} f(x); p\right] = f(x)x^p \Big|_0^\infty - pM_f(p) = -pM_f(p) \quad (2-34)$$

این روابط و قضایا بر اساس رابطه‌ای که بین تبدیل میلین با تبدیلات فوریه و لاپلاس وجود دارد بدست آمده است [۴،۹،۱۱،۱۵].

جدول ۱-۲: توابع انتخاب شده $f(x)$ ، با تبدیل میلین آنها و محدوده تعریف تبدیل

تابع اصلی $f(x), x > 0$	تبدیل میلین $M_f(p) = \int_0^{\infty} f(x)x^{p-1}dx$	محدوده تعریف
$(1+x)^{-1}$	$\pi / \sin(\pi p)$	$0 < \beta < 1$
$(a+x)^{-1}, \arg a < \pi$	$\pi a^{p-1} \csc(\pi p)$	$0 < \beta < 1$
$(1+x)^{-a}$	$\Gamma(p)\Gamma(a-p)/\Gamma(a)$	$0 < \beta < \text{Re}(a)$

$(1-x)^{-1}$	$\pi \cot(\pi p)$	$0 < \beta < 1$
$(a-x)^{-1}, a > 0$	$\pi a^{p-1} \cot(\pi p)$	$0 < \beta < 1$
$u(1-x)(1-x)^{b-1}, \text{Re}(b) > 0$	$\Gamma(p)\Gamma(b)/\Gamma(p+b)$	$\beta > 0$
$u(x-1)(x-1)^{-b}$	$\Gamma(b-p)\Gamma(1-b)/\Gamma(1-p)$	$\beta < \text{Re}(b) < 1$
$(x^2 + a^2)^{-1}, \text{Re}(a) > 0$	$\frac{1}{2} \pi a^{p-2} \csc\left(\frac{\pi p}{2}\right)$	$0 < \beta < 2$
$(x^n + a), \arg a < \pi$	$(\pi/n) \csc(\pi p/n) a^{p/n-1}$	$0 < \beta < n$
$\begin{cases} x^v & 0 < x < 1 \\ 0 & x > 1 \end{cases}$	$(p+v)^{-1}$	$\beta > -\text{Re}(v)$
$\begin{cases} (1-x^h)^{v-1} & -1 < x < 0 \\ 0 & x \geq 0 \\ h > 0 & \text{Re}(v) > 0 \end{cases}$	$h^{-1} \frac{\Gamma(v)\Gamma(p/h)}{\Gamma(v+p/h)}$	$\beta > 0$
$(1-x^a)(1-x^{na})^{-1}$	$\frac{\pi}{na} \sin\left(\frac{\pi}{n}\right) \csc\left(\frac{\pi p}{na}\right) \csc\left(\frac{\pi p + \pi a}{na}\right)$	$0 < \beta < (n-1)a$

توابع نمایی

$e^{-sx}, s > 0$	$s^{-p} \Gamma(p)$	$\beta > 0$
$(e^x - 1)^{-1}$	$\Gamma(p) \zeta(p)$ $\zeta(p) = \text{Zeta Function}$	$\beta > 1$
$(e^{ax} + 1)^{-1}, \text{Re}(a) > 0$	$a^{-p} \Gamma(p) (1 - 2^{1-p}) \zeta(p)$	$\beta > 0$
$(e^{ax} - 1)^{-1}, \text{Re}(a) > 0$	$a^{-p} \Gamma(p) \zeta(p)$	$\beta > 1$
$e^{-ax} (1 - e^{-x})^{-1}, \text{Re}(a) > 0$	$\Gamma(p) \zeta(p, a)$	$\beta > 1$
$(e^x - 1)^{-2}$	$\Gamma(p) [\zeta(p-1) - \zeta(p)]$	$\beta > 2$
$e^{-ax^h}, \text{Re}(a) > 0, h > 0$	$h^{-1} a^{-p/h} \Gamma(p/h)$	$\beta > 0$
$(xe^{-t})^{-1}$	$\Gamma(1-p)$	$-\infty < \beta < 1$
e^{-x^2}	$\frac{1}{2} \Gamma(p/2)$	$0 < \beta < \infty$
$1 - e^{-ax^h}, \text{Re}(a) > 0, h > 0$	$-h^{-1} a^{-p/h} \Gamma(p/h)$	$-h < \beta < 0$
e^{iax}	$a^{-p} \Gamma(p) e^{i\pi(p/2)}$	$0 < \beta < 1$

توابع لگاریتمی

تبدیل میلین

تابع اصلی $f(x), x > 0$

$$M_f(p) = \int_0^{\infty} f(x)x^{p-1} dx$$

محدوده تعریف

$$u(s-1)\ln(s-x) \quad -s^{-p} p^{-1} [\psi(p+1) + s^{-1} \ln \gamma] \quad \beta > 0$$

$$\ln \left| \frac{1+x}{1-x} \right| \quad (\pi/p) \tan(\pi p/2) \quad -1 < \beta < 1$$

$$\begin{cases} \ln x & 0 < x < a \\ 0 & x > a \end{cases} \quad p^{-1} a^{-p} (\ln a - p^{-1}) \quad \beta > 0$$

$$\begin{cases} x^v \ln x & 0 < x < 1 \\ 0 & x > 1 \end{cases} \quad -(p+v)^{-2} \quad \beta > -\operatorname{Re}(v)$$

$$e^{-x} (\ln x)^n \quad \frac{d^n \Gamma(p)}{dp^n} \quad \beta > 0$$

$$u(x-1) \sin(a \ln x) \quad a/(p^2 + a^2) \quad \beta < -|\operatorname{Im}(a)|$$

$$u(1-x) \sin(-a \ln x) \quad a/(p^2 + a^2) \quad \beta > |\operatorname{Im}(a)|$$

$$(u(x) - u(x-s)) \ln(s/x), s > 0 \quad (s/p)^2 \quad \beta > 0$$

توابع مثلثاتی

$$\sin(ax), a > 0 \quad a^{-p} \Gamma(p) \sin(\pi p/2) \quad -1 < \beta < 1$$

$$e^{-ax} \sin(\lambda x), \operatorname{Re}(a) > |\operatorname{Im}(\lambda)| \quad (a^2 + \lambda^2)^{-p/2} \Gamma(p) \sin\left(p \tan^{-1} \frac{\lambda}{a}\right) \quad \beta > -1$$

$$\sin^2(ax), a > 0 \quad -2^{-p-1} a^{-p} \Gamma(p) \cos(\pi p/2) \quad -2 < \beta < 0$$

$$\cos(ax), a > 0 \quad a^{-p} \Gamma(p) \cos(\pi p/2) \quad 0 < \beta < 1$$

$$\tan^{-1}(x) \quad -\pi/2p \cos(\pi p/2) \quad -1 < \beta < 0$$

$$\operatorname{co} \tan^{-1}(x) \quad \pi/2p \cos(\pi p/2) \quad 0 < \beta < 1$$

توابع دیگر

$$J_\nu(ax) \quad \left(\frac{1}{2}\right)^{p-1} \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{a^p \Gamma\left(1 + \frac{\nu-p}{2}\right)} \quad -\operatorname{Re}\{v\} < \beta < \frac{3}{2}$$

$$J_\mu(x)J_\nu(x) \quad \frac{1}{2}\left(\frac{1}{2}\right)^{-p}\Gamma(1-p)\Gamma\left(\frac{\nu+\mu+p}{2}\right) \times \left[\Gamma\left(1+\frac{\nu-\mu-p}{2}\right)\Gamma\left(1+\frac{\nu+\mu-p}{2}\right)\right]^{-1} \quad -\operatorname{Re}\{\nu+\mu\} < \beta < 1$$

$$\left(\frac{\sin x}{x}\right)^2 \quad -\frac{\sqrt{\pi}}{4}\frac{\Gamma\left(\frac{p}{2}-\frac{1}{2}\right)}{\Gamma\left(\frac{3}{2}-\frac{p}{2}\right)} \quad 0 < \beta < 2$$

تبدیل میلین

تابع اصلی $f(x), x > 0$ محدوده تعریف

$$M_f(p) = \int_0^\infty f(x)x^{p-1}dx$$

$$\sin(ax)J_\nu(ax), a > 0 \quad \frac{2^{v-1}\Gamma\left(\frac{1}{2}-p\right)\Gamma\left(\frac{1+v+p}{2}\right)}{a^p\Gamma(1+v-p)\Gamma\left(\frac{2-v-p}{2}\right)} \quad -1 < \operatorname{Re}(v) < \operatorname{Re}(p) < 1/2$$

$$\delta(x-a), a > 0 \quad a^{p-1} \quad \text{Whole plane}$$

$$\sum_{n=1}^\infty \delta(x-sn), s > 0 \quad s^{p-1}\zeta(1-p) \quad \beta < 0$$

$$\sum_{n=-\infty}^\infty s^{-nr}\delta(x-s^n) \quad \frac{1}{\ln s} \sum_{n=-\infty}^\infty \delta\left(c-\frac{n}{\ln s}\right) \quad c = \operatorname{Im}(p)$$

$$\log\left(\frac{1-e^{-x}}{x}\right) \quad -\frac{\Gamma(p+1)\xi(p+1)}{p} \quad -1 < \beta < 0$$

$$\operatorname{erfc} x \quad \frac{\Gamma\left(\frac{1}{2}p+\frac{1}{2}\right)}{\pi^2 p} \quad 0 < \beta < \infty$$

$$\operatorname{Si} x \quad -\frac{\Gamma(p)\sin\frac{1}{2}\pi p}{p} \quad -1 < \beta < 0$$

جدول ۲-۲. قضایای تبدیل میلین

حالت	تابع	تبدیل میلین
۱	$M\{f(ax)\}, a > 0$	$a^{-p} M_f(p)$
۲	$M\{x^a f(x)\}, a > 0$	$M_f(p+a)$
۳	$M\{f(x^a)\}, a > 0$	$a^{-1} M_f(p/a), a > 0$
۴	$M\{x^{-1} f(x^{-1})\}$	$M_f(1-p)$
۵	$M\{\ln x \cdot f(x)\}$	$\frac{d}{dp} M_f(p)$
۶	$M\{(\ln x)^k f(x)\}$	$\frac{d^k}{dp^k} M_f(p)$
۷	$M\left\{\frac{d^k}{dp^k} f(x)\right\}$	$(-1)^k (p-k)_k M_f(p-k)$ $(p-k)_k \equiv (p-k)(p-k+1)\dots(p-1)$ $= \frac{(p-1)!}{(p-k-1)!} = \frac{\Gamma(p)}{\Gamma(p-k)}$
۸	$M\left\{x^k \frac{d^k}{dp^k} f(x)\right\}$	$(-1)^k \frac{\Gamma(p+k)}{\Gamma(p)} M_f(p)$
۹	$M\{f(x) \cdot g(x)\}$	$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} M_f(p) M_g(p-z) dz$
۱۰	$M\{conv(f, g)\}$	$M_f(p) M_g(p)$
۱۱	$\frac{d}{dx} f(x)$	$-(p-1) M_f(p-1)$

١٢	$x \frac{d}{dx} f(x)$	$-pM_f(p)$
١٣	$\int_0^x f(t)dt$	$-\frac{1}{p} M_f(p+1)$

٢-١١- مراجع

- [1] Bertrand, Jacqueline, Pierre Bertrand, and Jean-Philippe Ovarlez, " The Mellin Transform, in Transforms and Applications Handbook ", ed. Alexander Poularikas, CRC Press, Boca Raton, Florida, 1996.
- [2] Davies, G, " Integral Transforms and Their Applications ", 2nd ed., Springer-Verlag, New York, NY, 1984.
- [3] Erdelyi, A., W. Magnus, F. Oberhettinger, and F. G. Tricomi," Tables of Integral Transfer", McGraw-Hill Book Co., New York, NY, 1954.
- [4] Oberhettinger, F., " Tables of Mellin Transform ", 2nd ed., Springer-Verlag, New York, NY, 1974.
- [5] Sneddon, Ian N., "The Use of Integral Transform ", McGraw-Hill Book Co., New York, NY, 1972.
- [6] George Fikioris, " Mellin-Transform Method for Integral Evaluation ", A Publication in the Morgan & Claypool Publishers series, First Edition, 2007.

- [7] Ronald N. Bracewell, " The Fourier Transform and Its Application ". McGraw-Hill International Editions. Electrical Engineering Series. Third Edition. Printed in Singapore. 2002. Chapter 13, Pages 343 – 347.
- [8] Paris, R. B., and Kaminsky, D., " Asymptotics and Mellin-Barnes Integrals ", Cambridge University Press, 2001.
- [9] A. D. Polyanin and A. V. Manzhirov, " Handbook of Integral Equations ", CRC Press, Boca Raton, 1998. ISBN 0-8493-2876-4.
- [10] P. Flajolet, X. Gourdon, P. Dumas, " Mellin transforms and asymptotics: Harmonic sums ", Theoretical Computer Science, 144(1-2):3-58, June 1995.
- [11] Tables of Integral Transforms (<http://eqworld.ipmnet.ru/en/auxiliary/aux-inttrans.htm>) at EqWorld: The World of Mathematical Equations.
- [12] G.H. Hardy and J.E. Littlewood, "Contributions to the Theory of the Riemann Zeta-Function and the Theory of the Distribution of Primes," *Acta Mathematica*, 41(1916) pp.119-196. (*See notes therein for further references to Cahen's and Mellin's work, including Cahen's thesis.*)
- [13] Arfken, G. " Mathematical Methods for Physicists ", 3rd ed. Orlando, FL: Academic Press, p.795, 1985.
- [14] Gradshteyn, I. S. and Ryzhik, I. M. " Mellin Transform." §17.41 in Tables of Integrals, Series, and Products, 6th ed. San Diego, CA: Academic Press, pp. 1193-1197, 2000.
- [15] Morse, P. M. and Feshbach, H. " Methods of Theoretical Physics ", Part I. New York: McGraw-Hill, pp. 469-471, 1953.
- [16] Prudnikov, A. P.; Brychkov, Yu. A.; and Marichev, O. I. " Evaluation of Integrals and the Mellin Transform." *Itogi Nauki i Tekhniki, Seriya Matemat. Analiz* **27**, 3-146, 1989.

[17] Titchmarsh, E. C. " The Theory of the Riemann Zeta Function ", 2nd ed. New York: Clarendon Press, 1987.

[18] Zwillinger, D. (Ed.). " CRC Standard Mathematical Tables and Formulae ". Boca Raton, FL: CR Press, p. 567, 1995.

فصل سوم

تبدیل میلین سریع

۳-۱- مقدمه

در این فصل به دنبال ارائه الگوریتمهایی جهت محاسبه تبدیل میلین سیگنالهای آنالوگ و دنبال آن محاسبه تبدیل میلین سیگنالهای گسسته می‌باشیم. محاسبه تبدیل میلین پیوسته کار دشواری نیست، بدین علت که با توجه به معادلات ارائه شده در فصل ۲ بخصوص معادلات (۲-۱۱) و (۲-۱۲) که رابطه بین تبدیل میلین را با تبدیلات فوریه و لاپلاس بیان می‌نمود، می‌توان با استفاده از روابط انتگرالی و همچنین جدول ۲-۱، این انتگرالها را محاسبه نمود.

در کاربردهای واقعی، مانند پردازش صوت و تصویر، تابع پیوسته یا سیگنال اصلی به سختی در اختیار ما قرار دارد. بعبارت دیگر اکثرا با یک سری رشته عددی یا بعبارت دیگر یک تابع گسسته که نمونه‌برداری شده

تابع اصلی است، سر و کار داریم. بهمین علت در این فصل در ابتدا در مورد چگونگی محاسبه تبدیل میلین پیوسته بحث می‌کنیم و سپس تمرکز خود را بیشتر بر روی چگونگی محاسبه تبدیل میلین سیگنالهای گسسته می‌گذاریم.

۳-۲- محاسبه تبدیل میلین توابع پیوسته

ایده اصلی محاسبه تبدیل میلین از رابطه بین این تبدیل و تبدیل فوریه گرفته شده است. تبدیل میلین را می‌توان بصورت تبدیل فوریه انتقال یافته محور x در نظر گرفت. با توجه به رابطه (۲-۱) تبدیل میلین بصورت زیر بیان می‌شود:

$$M_f(p) = \int_0^{\infty} x^{p-1} f(x) dx \quad (۱-۳)$$

با جایگزاری $\tau = \ln x$ یا معکوس آن $x = e^\tau$ ، که τ یک متغیر جدید است، محور x انتقال می‌یابد. در اینحالت

$$dx = e^\tau d\tau \quad (۲-۳)$$

بعلاوه با این تعریف، محدود انتگرالگیری تغییر می‌کند. هنگامیکه $x = 0$ باشد، $\tau = \ln 0 = -\infty$ و بهمین ترتیب هنگامیکه $x = \infty$ شود، $\tau = \ln \infty = \infty$ خواهد شد. با اعمال این تغییرات در رابطه (۳-۱) داریم:

$$M_f(p) = \int_{-\infty}^{\infty} e^{\tau(p-1)} f(e^\tau) (e^\tau d\tau) \quad (۳-۳)$$

با مرتب سازی دوباره رابطه (۳-۳) و جایگزاری $p = -jc + \beta$ خواهیم داشت:

$$M_f(p) = \int_{-\infty}^{\infty} e^{\beta x} f(e^x) e^{-jcx} dx \quad (۴-۳)$$

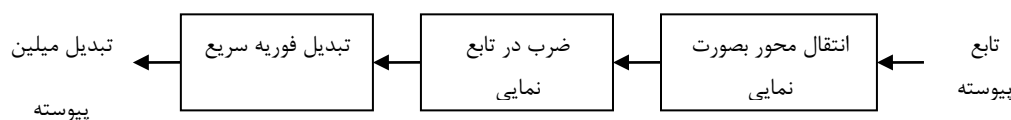
رابطه فوق بیان می‌کند، تبدیل میلین تابع پیوسته $f(x)$ برابر با تبدیل فوریه تابع $g(x)$ است، که در آن $g(x) = e^{\beta x} f(e^x)$. در بلوک دیاگرام نشان داده شده در شکل ۳-۱ تبدیل میلین یک تابع پیوسته معادل با انتقال محور x بصورت نمایی ($f(x) \rightarrow f(e^x)$) و سپس ضرب نمودن آن در تابع نمایی $e^{\beta x}$ و در نهایت

محاسبه تبدیل فوریه آن است. برای درک بهتر مفهوم انتقال متغیر x نگاشت دو بعدی تبدیل از x به τ در شکل ۳-۲ نشان داده شده است.

۳-۳- محاسبه تبدیل میلین توابع گسسته

در اکثر کاربردهای که ما با آنها سر و کار داریم، تابع اصلی یا پیوسته به سختی در اختیار ما قرار دارد و اکثراً با توابع گسسته که نمونه برداری شده تابع پیوسته هستند، سر و کار داریم. نمونه برداری پروسه‌ای است که در طی آن یک سیگنال آنالوگ (تابع پیوسته) به یک رشته عددی (تابع گسسته) تبدیل می‌شود. محاسبه تبدیل میلین گسسته امری نسبتاً ضروری بنظر می‌رسد. یکی از راههای محاسبه این تبدیل، تقریب زدن انتگرال تبدیل رابطه (۳-۱) با استفاده از مجموع ریمن^۴ است [۱۲]. اما استفاده از این الگوریتم دارای پیچیدگی محاسباتی بسیار بالایی می‌باشد و در کاربردهای عملی مانند تشخیص الگو^۵ نمی‌توان از این الگوریتم استفاده نمود.

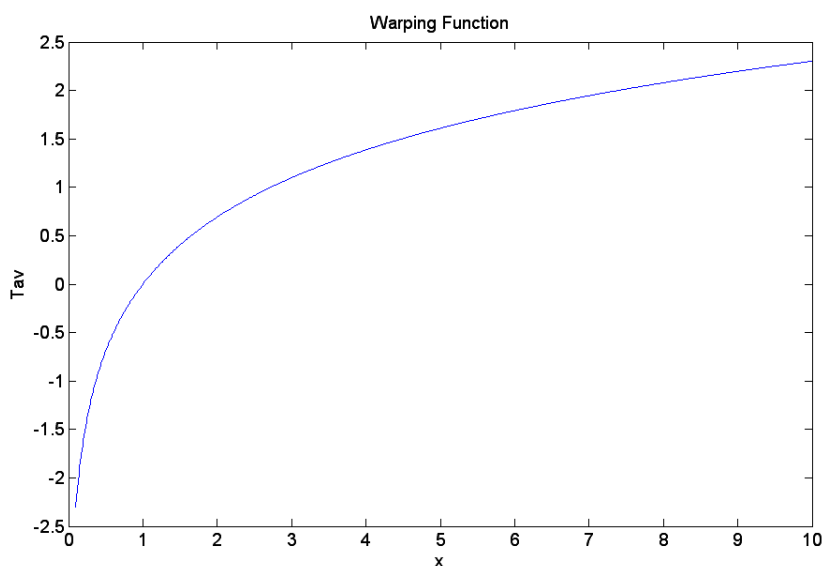
یکی از راههای دیگری که می‌توان از آن برای محاسبه تبدیل میلین گسسته استفاده نمود، تقریب زدن مقدار انتگرال رابطه (۳-۴) است. اینکار را می‌توان به کمک یک تابع که بصورت یکنواخت نمونه برداری شده است، انتقال آن بصورت نمایی، ضرب نمودن آن در یک تابع نمایی و در نهایت محاسبه تبدیل فوریه تابع حاصل، انجام داد. بلوک دیاگرام مراحل انجام اینکار در شکل ۳-۳ نشان داده شده است.



شکل ۳-۱. بلوک دیاگرام تبدیل میلین تابع پیوسته

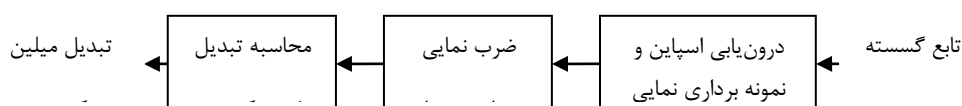
^۴ . Riemann sum

^۵ . Pattern recognition



شکل ۳-۲. نمایش دو بعدی تابع انتقال محور x

همانطور که در این شکل نشان داده شده است، مشکل عمده جهت تقریب زدن این انتگرال، عملیات انتقال^۶ (درون‌یابی اسپاین^۷ و نمونه برداری نمایی) است. در این قسمت باید یک تابع یا بعبارت بهتر یک رشته عددی که بصورت یکنواخت نمونه برداری شده است را دوباره بصورت نمایی نمونه برداری کنیم.



شکل ۳-۳. مراحل الگوریتم تبدیل میلین گسسته

۳-۳-۱- تئوری نمونه برداری

برای اجرای هر گونه تحلیلی با استفاده از کامپیوتر، بایستی تابع یا سیگنال اصلی به نمونه‌های گسسته در حوزه زمان گسسته کاهش یابد [۱]. پروسه‌ای که یک سیگنال را از حالت زمان پیوسته به حالت زمان گسسته تبدیل می‌کند را نمونه برداری گویند و این عمل با انتخاب نمونه‌های زمان پیوسته تابع اصلی در

^۶ . Wrapping Operation
^۷ . Spline Interpolation

بازه‌های زمانی مشخص و ثابت T ، که آنرا بازه نمونه برداری می‌نامیم، انجام می‌گیرد. کمیت $F_s = 1/T$ را نرخ نمونه برداری می‌نامیم.

شالوده اصلی تئوری نمونه برداری را می‌توان در چند قانون خلاصه نمود. اولین قانون مربوط به نمایش فرکانسی متغیرهای زمان گسسته با استفاده از تبدیل فوریه است.

قانون اول : تبدیل فوریه یک تابع گسسته تابعی از متغیرهای پیوسته ω است که پریودیک و با دوره تناوب 2π است.

قانون دوم به ما این امکان را می‌دهد تا با سیگنالهای نمونه برداری شده، مانند تابعی از متغیرهای گسسته رفتار نماییم.

قانون دوم : با نمونه برداری از سیگنال زمان پیوسته $x(t)$ با بازه نمونه برداری T یک تابع از متغیرهای گسسته n که در آن $x(n) \equiv x(nT)$ است، تولید می‌شود.

اگر طیف یک سیگنال را تبدیل فوریه انتقال یافته آن بنامیم، قانون اصلی نمونه برداری بصورت زیر بیان می‌گردد:

قانون سوم : نمونه برداری یک سیگنال زمان پیوسته با نرخ نمونه برداری F_s یک سیگنال زمان گسسته تولید می‌کند که طیف فرکانسی آن یک تابع پریودیک و تکرار پذیر از طیف سیگنال اصلی است و پریود تکرار آن برابر با F_s است.

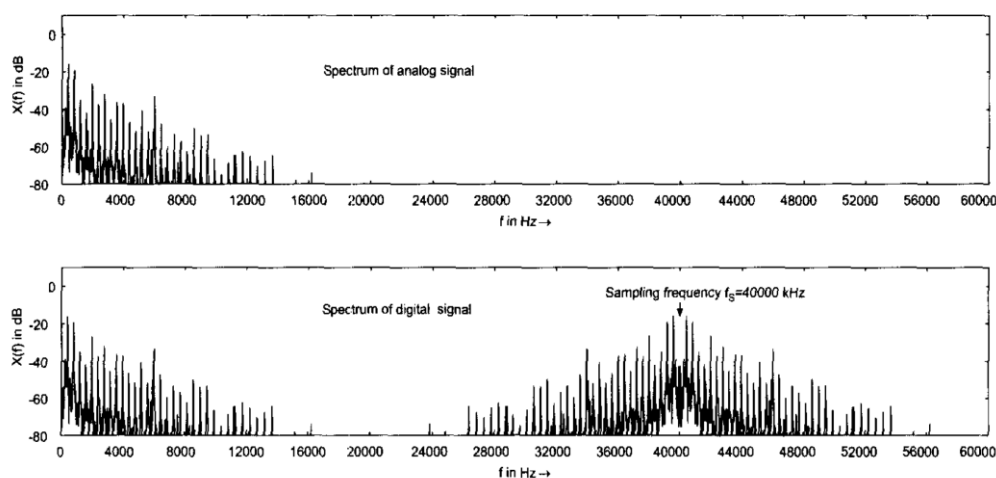
طیف سیگنال، توزیع انرژی را بر روی محدوده فرکانسی نشان می‌دهد. بعنوان نمونه قسمت بالایی شکل ۳-۴ طیف زمان کوتاه یک سیگنال آنالوگ را نشان می‌دهد [۱۳]. محدوده فرکانسی آن تا ۲۰ کیلو هرتز است. نمونه برداری و کوانتیزه نمودن سیگنال آنالوگ با فرکانس نمونه برداری $f_s = 40 \text{ KHz}$ منجر به ایجاد سیگنال دیجیتال یا گسسته می‌گردد. طیف فرکانسی این سیگنال گسسته در همان بازه زمانی در قسمت پایینی شکل ۳-۴ نشان داده شده است. عملیات نمونه برداری باعث تکرار طیف سیگنال آنالوگ می‌گردد.

مولفه های فرکانسی از صفر هرتز تا ۲۰ کیلو هرتز سیگنال آنالوگ، دوباره از ۴۰ کیلو هرتز تا ۶۰ کیلو هرتز ظاهر می گردند. همانطور که در این شکل نشان داده شده است، این طیف نسبت به فرکانس نمونه برداری متقارن است. اما نکته ای که به آن باید دقت نمود این است که طیف سیگنال گسسته از صفر تا ۲۰ کیلو هرتز دقیقاً مشابه با طیف سیگنال آنالوگ است. در نتیجه برای بازیابی سیگنال آنالوگ با استفاده از سیگنال نمونه برداری شده، فقط کافیست که سیگنال گسسته را از یک فیلتر پایین گذر عبور دهیم.

با قوانین ساده ای که در بالا آنها را معرفی کردیم، براحتی می توان تئوری نمونه برداری زیر را فهمید، این تئوری توسط نایکوئیست در قرن بیستم معرفی گردید و بوسیله شانون و در اوایل دهه چهل بصورت جهانی مطرح گردید.

تئوری شانون - نایکوئیست : سیگنال زمان پیوسته $x(t)$ ، که مولفه های طیف فرکانسی آن

کوچکتر از فرکانس F_b باشد یا بعبارت بهتر باند محدود شده به F_b باشد را می توان از روی نوع



شکل ۳-۴. طیف فرکانسی سیگنالهای آنالوگ و دیجیتال

نمونه برداری شده آن $\hat{x}(n) = x(nT)$ بازیابی نمود، اگر فرکانس نمونه برداری $F_s = 1/T$ بصورت زیر باشد:

$$F_s > 2F_b \quad (۳-۵)$$

همانطور که در بالا توضیح داده شد، براحتی می‌توان با استفاده از یک فیلتر خطی که فرکانسهای بالاتر از $F_N = F_s / 2$ را حذف نماید، سیگنال اصلی را بازیابی نمود. فیلتر بازیابی را می‌توان در حوزه زمان، پیوسته و بصورت پاسخ ضربه در نظر گرفت، که بصورت زیر تعریف می‌گردد:

$$h(t) = \text{sinc}(t) = \frac{\sin(\pi t / T)}{\pi t / T} \quad (3-6)$$

به تابع فوق سینک گویند که در شکل ۳-۵ نشان داده شده است.

در حالت ایده‌ال، بازیابی سیگنال زمان پیوسته با استفاده از سیگنال نمونه برداری شده در دو مرحله مطابق زیر صورت می‌گیرد:

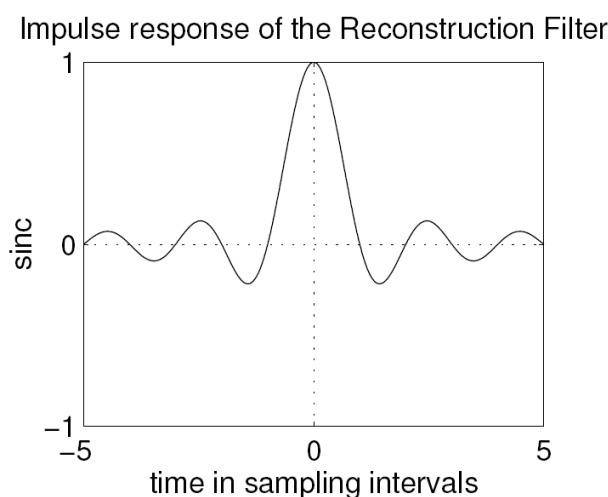
۱. تبدیل از حالت زمان گسسته به پیوسته با نگه داشتن مقادیر ثابت سیگنال در بازه‌های زمانی بین دو نمونه گرفته شده. این عمل با استفاده از وسیله‌ای بنام نگه‌دارنده انجام می‌گیرد. با اتصال سری و پشت سر هم نمونه‌بردار و نگه‌دارنده این تبدیل را می‌توان انجام داد.
۲. کانالو نمودن آن با تابع سینک^۸ ایده‌ال

تابع سینک از آن جهت ایده‌ال است که بسط زمانی آن از دو سمت تا بینهایت ادامه دارد، بنابراین در حالت عادی، با توجه به بینهایت بودن نمونه‌های تابع سینک، پروسه بازیابی نمی‌تواند بطور کامل یک سیگنال آنالوگ را بازیابی نماید. اما بهر حال، با استفاده از تقریب زدن تابع سینک بوسیله مجموعه‌ای از توابع ضربه می‌توان به دقت خوبی دست یافت.

با در نظر گرفتن شرایط فوق، اگر تابع گسسته بصورت یکنواخت نمونه‌برداری شده باشد و از قضیه شانون - نایکوئیست پیروی نماید، می‌توان تابع اصلی را بازیابی نمود. با توجه به بلوک دیاگرام ارائه شده در شکل ۳-۳ هدف نمونه برداری نمایی از یک سیگنال گسسته است که از قضیه شانون - نایکوئیست تبعیت می‌کند. برای انجام اینکار از دو نوع درونیابی می‌توان استفاده نمود. اولین درونیابی بر اساس تابع نمونه برداری شده نمایی است که در آن ما می‌دانیم که این تابع باید در طول محور زمانی انتقال یافته شده قرار

^۸ . sinc function

گیرد. در اینحالت تابع حاصل، دارای پهنای باند میلین محدود⁹ است. در اینحالت، بعنوان نمونه یک موج سینوسی با یک دوره تناوب را می توان مشابه با شکل تابع اصلی رسم نمود ولی در اینحالت تراکم نمونه ها در نقاط شروع تابع بسیار بالا می باشد. درون یاب بعدی



شکل ۳-۵. تابع سینک، پاسخ ضربه فیلتر بازیابی ایده‌ال

بصورت تابع نمونه برداری شده غیر یکنواخت انتقال یافته زمانی است. در اینحالت، سیگنال انتقال یافته دارای پهنای باند فوریه محدود است. در اینحالت، بعنوان نمونه، شکل تابع سینوسی با یک دوره تناوب، بسیار با شکل اصلی خود متفاوت خواهد شد و همچنین دارای اعوجاج بسیار بالایی است. فرضهایی را که برای تابع ورودی جهت اجرای الگوریتم تبدیل میلین گسسته در نظر گرفته ایم عبارتند از:

- دارای محدودیت زمانی است، بخاطر اینکه بتوان آنرا در یک سیستم ذخیره سازی با ابعاد محدود ذخیره نمود.
- دارای تبدیل فوریه با پهنای باند محدود است، بدین علت که با نمونه برداری یکنواخت و تحت شرایط شانون - نایکوئیست بدست آمده است.
- دارای تبدیل میلین با پهنای باند محدود است تا تعداد نقاط محدودی برای نمایش تبدیل میلین آن وجود داشته باشد.

⁹ . Mellin Band - limited

این شرایط در صورتی برقرار می‌باشند که تابع اصلی دارای تبدیل فوریه با پهنای باند محدود باشد

[۱۲].

۳-۲-۳ - نمونه برداری نمایی از تابع گسسته

هنگامیکه می‌خواهیم از یک تابع (سیگنال) که بصورت یکنواخت نمونه‌برداری شده است، دوباره بصورت

نمایی نمونه‌برداری کنیم، چندین سوال مطرح می‌گردد. این سوالات عبارتند از:

چه تعداد نمونه لازم است؟

این نمونه‌ها چگونه باید در طول زمان توزیع شوند؟

چگونگی تاثیر زمان شروع سیگنال بر روی پروسه نمونه برداری؟

چگونگی بازیابی سیگنال و سوالات بسیار دیگر.

اول از همه باید شرط نمونه برداری نایکوئیست برقرار گردد. بنابراین، فاصله بین دو نمونه در نمونه‌برداری نمایی نمی‌تواند بیشتر از فاصله بین دو نمونه در حالت نمونه‌برداری یکنواخت باشد. این عبارت بدین معنی است که اگر T_s دوره تناوب نمونه برداری نمایی باشد، بیشترین مقداری را که می‌تواند داشته باشد، برابر با حد فاصل بین دو نمونه در حالت نمونه‌برداری یکنواخت (T_s) است. دومین محدودیت این است که پروسه نمونه‌برداری باید تمام سیگنال را پوشش دهد، از نقطه شروع تا نقطه پایان.

این دو محدودیت ما را وادار می‌سازد تا نمونه‌های بیشتری را در نمونه برداری نمایی از یک سیگنال گسسته داشته باشیم. از طرف دیگر، نقطه شروع سیگنال (t_0) بسیار مهم است. در واقع هر چقدر t_0 به صفر نزدیک باشد، تعداد نمونه‌های بیشتری برای نمونه‌برداری دوباره نمایی نیاز است. در نتیجه اگر $t_0 = 0$ باشد، نیاز به بینهایت نمونه داریم. بنابراین برای استفاده از این الگوریتم باید نقطه شروع حتما بزرگتر از صفر باشد. می‌توان نمونه برداری دوباره نمایی را بصورت دنباله زیر در نظر گرفت:

$$\left\{ (\tau_s)^k \right\}_{k=-\infty}^{\infty} \quad (۷-۳)$$

که k اندیس نمونه و τ_s را می توان پرشهای نمایی نامید. بنابراین با استفاده از اولین محدودیت ($\tau_s^{k_e} - \tau_s^{k_e-1} = T_s$) باید:

$$\tau_s = \frac{t_0 + nT_s}{t_0 + (n-1)T_s} \quad (۸-۳)$$

که n تعداد نمونه های نمونه برداری یکنواخت است. نکته ای که در این قسمت به آن باید توجه شود این است که در حالت کلی باید داشته باشیم $\tau_s^{k_e} - \tau_s^{k_e-1} \leq T_s$. اما اگر بخواهیم کمترین نمونه ممکن بدست آید، می توان از رابطه فوق استفاده نمود. $\tau_s^{k_e}$ آخرین نمونه و k_e اندیس نمونه آخر (نمونه آخرین نقطه زمانی) است. در رابطه فوق اگر فرض کنیم زمان شروع سیگنال برابر با عکس فرکانس نمونه برداری یکنواخت است، یا عبارت دیگر $t_0 = T_s$ ، رابطه فوق ساده تر خواهد شد و داریم:

$$\tau_s = 1 + \frac{1}{n} \quad (۹-۳)$$

حال با استفاده از دومین محدودیت ($\tau_s^{k_0} = t_0 \wedge \tau_s^{k_e} = t_0 + nT_s = t_0$) تعداد نمونه های نمایی مورد نیاز برابر است با:

$$eN = \left[\ln((t_0 + nT_s) / t_0) / \ln((t_0 + nT_s) / (t_0 + (n-1)T_s)) \right] + 1 \quad (۱۰-۳)$$

اگر از T_s بعنوان نقطه شروع نمونه برداری استفاده شود، رابطه فوق ساده تر خواهد شد:

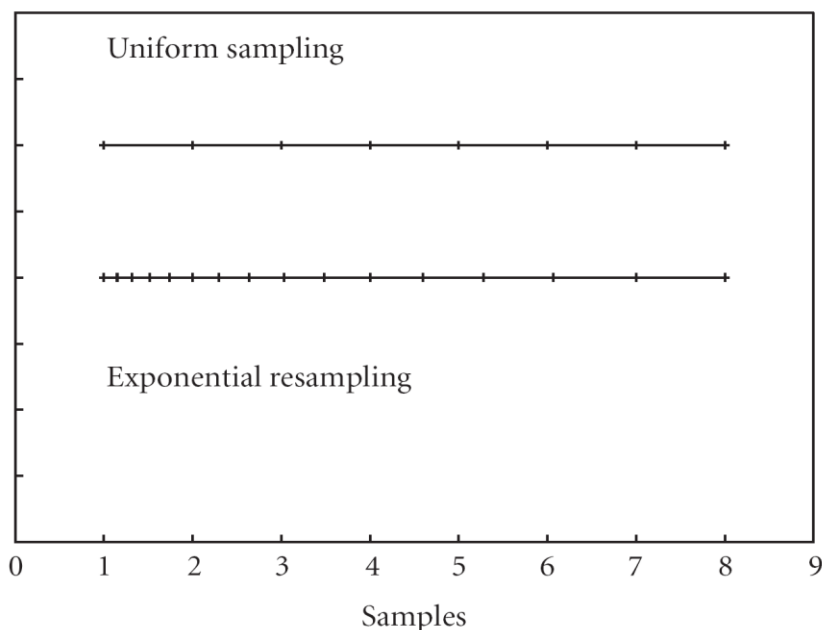
$$eN = \ln(n+1) / \ln((n+1) / n) \quad (۱۱-۳)$$

بعلاوه اگر تعداد نمونه های سیگنال گسسته ورودی زیاد باشد (در عمل تعداد بیشتر از ۱۶)، می توان رابطه (۱۱-۳) را بصورت زیر نوشت:

$$eN = n \ln n \quad (۱۲-۳)$$

با محاسبات فوق توانستیم قدمهای نمونه برداری و تعداد نمونه های نمایی مورد نیاز را برای نمونه برداری نمایی از یک سیگنال گسسته، که خود نمونه برداری شده یکنواخت سیگنال اصلی است، را بدست آوریم [۲]. در شکل ۳-۶ موقعیت زمانی نمونه ها در نمونه برداری یکنواخت و چگونگی توزیع نمونه ها در حالت

نمونه برداری نمایی از آن، نشان داده شده است. همانطور که در شکل نشان داده شده است، در حالت نمونه برداری نمایی، چگالی توزیع نمونه‌ها در ابتدای سیگنال بالا است و بتدریج با فاصله گرفتن از نقطه شروع سیگنال این چگالی کاهش می‌یابد، اما فاصله بین دو نمونه در این حالت نباید بیشتر از فاصله بین دو نمونه در حالت نمونه برداری یکنواخت گردد.



شکل ۳-۶. نمونه برداری یکنواخت و نمونه برداری دوباره بصورت نمایی از آن

۳-۳-۳- درون‌یابی

نمونه برداری و درون‌یابی در واقع رابط بین توابع آنالوگ و دیجیتال می‌باشند. همانطور که گفته شد، نمونه برداری یک سیگنال آنالوگ را به یک سیگنال دیجیتال تبدیل می‌کند، در حالیکه درون‌یابی یک سیگنال گسسته (دیجیتال) را به سیگنال آنالوگ تبدیل می‌کند. در پروسه نمونه برداری، بسیاری از اطلاعات سیگنال آنالوگ حذف می‌گردد، بدین علت که تعداد محدودی از مقادیر سیگنال آنالوگ نگه داشته می‌شود. بهمین علت، در روش درون‌یابی یک سری فرضیات در مورد نحوه قرارگیری نمونه‌ها بین دو نمونه گرفته شده، در نظر گرفته می‌شود. بنابراین برای بازیابی کامل یک سیگنال آنالوگ از روی نمونه‌های آن، هم پروسه نمونه برداری مهم است و هم پروسه درون‌یابی. با توجه به شرایطی که برای یک تابع پیوسته در فاصله بین

دو نمونه آن در نظر گرفته می شود، چندین حالت برای درون‌یابی می‌توان در نظر گرفت که در ادامه راجع به هر یک از آنها بحث می‌شود.

در تحلیل‌های ریاضی، درون‌یابی روشی برای ساخت نقاط جدید در داخل محدوده‌ای از داده‌های گسسته است. در مباحث مهندسی بعضی از مواقع مانند حالت فوق تعدادی نقاط داده داریم که با نمونه‌برداری بدست آمده‌اند و سعی داریم تابعی را بسازیم که بیشترین مطابقت را با داده‌های ورودی داشته باشد. این عمل را تطبیق منحنی¹⁰ یا آنالیز برگشتی می‌نامند [۷]. درونیابی حال خاصی از تطبیق منحنی است، که در آن تابع باید دقیقاً از میان نقاط داده عبور کند.

مشکلاتی که در رابطه با درون‌یابی وجود دارد، تقریب تابعی پیچیده با استفاده از یک تابع ساده است. فرض کنید که تابع درون‌یاب را می‌دانیم ولی محاسبه آن خیلی پیچیده و مشکل است، بنابراین می‌توان با انتخاب یک سری از نقاط داده مشخص از تابع پیچیده، ایجاد یک جدول مراجعه¹¹، و با درون‌یابی این نقاط، تابع ساده‌تری را تخمین زد. البته، در صورت استفاده از یک تابع ساده‌تر بجای تابع اولیه نمی‌توان انتظار داشت که نتایج مشابه زمانی باشد که از تابع اصلی استفاده می‌کنیم، اما بسته به نوع مسئله و روش درون‌یابی استفاده شده می‌توان مقدار خطای ایجاد شده را کاهش داد.

نکته‌ای که باید به آن اشاره نمود این است که روشهای متعددی برای درون‌یابی در ریاضیات وجود دارد، که اصطلاحاً آنها را اپراتورهای درون‌یابی می‌نامند. در این روشها بدنبال تطبیق نقاط با استفاده از یک سری توابع هستیم. بعنوان نمونه، ساده‌ترین فرم درون‌یابی محاسبه میانگین بین دو نقطه و یافتن نقطه بین آن دو است. این حالت دارای نتیجه‌ای مشابه با درون‌یابی خطی است. انواع درون‌یابی عبارتند از:

۱- درون‌یابی نزدیکترین همسایه

۲- درون‌یابی خطی

۳- درون‌یابی چند جمله‌ای

۴- درون‌یابی اسپاین

۴-۱- اسپاین خطی

¹⁰ . Curve fitting

¹¹ . Lookup table

۲-۴- اسپاین درجه دوم

۳-۴- اسپاین درجه سوم

۳-۳-۴- اجرا و نتایج

با استفاده از ایده‌های مطرح شده در قسمت‌های قبل، تبدیل میلیون سریع اجرا گردید. الگوریتم یک سیگنال یا تابع گسسته با نمونه برداری یکنواخت را می‌گیرد و عملیات نمونه برداری نمایی را بر روی آن انجام می‌دهد. فرض بر این است که تابع بصورت یکنواخت و طبق شرایط نایکوئیست نمونه برداری شده است و برای بدست آوردن دقت و سرعت، همزمان با هم، تعداد نمونه‌های مورد استفاده در حالت نمایی $2eN$ است که eN از رابطه (۳-۱۲) بدست می‌آید. در اینجا فرض بر این است که لحظه شروع سیگنال T_s (دوره تناوب نمونه برداری یکنواخت) است، اما می‌توان به ازای زمانهای شروع دیگر نیز این الگوریتم را اجرا نمود. در این الگوریتم می‌توان از درون‌یابی اسپاین درجه سوم معمولی یا کاردینال استفاده نمود. در اینحال پیچیدگی الگوریتم جهت محاسبه درون‌یابی $O(N)$ است که N تعداد نمونه‌های نمایی است. بعد از نمونه برداری نمایی، ضرب نمایی نقطه به نقطه صورت می‌گیرد که پیچیدگی آن $O(N)$ است. و پس از آن تبدیل فوریه سریع محاسبه می‌گردد. مرتبه پیچیدگی الگوریتم تبدیل فوریه سریع برابر با $O(N \ln N)$ است. در نتیجه پیچیدگی کلی الگوریتم برابر با $O(N \ln N)$ است. در نهایت انرژی خروجی حاصل، نرمالیزه می‌گردد که یک عملگر خطی است. در نتیجه پیچیدگی کلی الگوریتم برابر با پیچیدگی محاسباتی الگوریتم تبدیل فوریه است. در صورتیکه بخواهیم پیچیدگی کلی را بر حسب تعداد نمونه‌های سیگنال نمونه برداری شده یکنواخت بدست آوریم، داریم:

$$O(n \ln n^2) = \text{مرتبه پیچیدگی الگوریتم تبدیل میلیون سریع} \quad (۳-۲۶)$$

٣-٤- مراجع

- [1] Rocchesso. "Introduction to sound processing". Wiley 2005 – Chapter 1, Systems, sampling and quantization, pages 1 – 22.
- [2] Antonio De Sena And Davide Rocchesso. "A Fast Mellin transform With Application IN DAFx". Proc of the 7 th int. conference on digital Audio Effect (DAFx04), Naples. Italy, October 5-8,2004
- [3] Cardinal Splines at Microsoft Developer Network (<http://msdn2.microsoft.com/enus/library/ms536358.aspx>)
- [4] [Http://en.wikipedia.org/wiki/Cubic_Hermite_spline#Cardinal_spline](http://en.wikipedia.org/wiki/Cubic_Hermite_spline#Cardinal_spline) Categories: Splines Interpolation
- [5] Carl de boor. "Spline Toolbox ", for use with MATLAB. user guide version 2.
- [6] Frank Candocia and Jose C. Principe, "Comments on Sinc Interpolation of Discrete Periodic Signals ". To appear in the IEEE Transactions on Signal Processing.
- [7] Schaum A. "Theory and Design of Local Interpolators", Comp. Vision Graph. Image Proc.: Models and Image Proc., Vol. 55, No. 6, pp 464-481, Nov. 1993.
- [8] Oppenheim A.V. and Schafer R. W., Discrete-Time Signal Processing, Prentice Hall, Englewood Cliffs, NJ, Chap. 8, 1989.
- [9] David Kidner, Mark Dorey and Derek Smith (1999). What's the point? Interpolation and extrapolation with a regular grid DEM (http://www.geovista.psu.edu/sites/geocomp99/Gc99/082/gc_082.htm). IV International Conference on GeoComputation, Fredericksburg, VA, USA.
- [10] Kincaid, David; Ward Cheney (2002). Numerical Analysis (3rd edition). Brooks/Cole. ISBN 0-534-38905-8. Chapter 6.
- [11] Schatzman, Michelle (2002). Numerical Analysis: A Mathematical Introduction. Clarendon Press, Press, Oxford. ISBN 0-19-850279-6. Chapters 4 and 6.

- [12] Antonio De Sena and Davide Rocchesso. "A Fast Mellin and Scale Transform". EURASIP Journal on advances in signal Processing Volume 2007.
- [13] Udo Zolzer, " DAFx: Digital Audio Effects ", Chapter 1,pages 1-18,John Willy and Sons, LTD,2002.
- [14] J. Bertrand, P. Bertrand, and J. Ovarlez, "Discrete Mellin transform for signal analysis," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP-90., vol. 3. Albuquerque, NM, USA: IEEE, May 1990, pp. 1603–1606.
- [15] T. Schanze, "Sinc interpolation of discrete periodic signals," IEEE Transactions on Signal Processing, vol. 43, no. 6, pp. 1502–1503, June 1995.

فصل چهارم

کاربردهای تبدیل

میلین

۴-۱- مقدمه

همانطور که در فصل قبل بیان گردید، یکی از ویژگیهای مهم تبدیل میلین و نوع محدود شده آن یعنی تبدیل مقیاس، نامتغیر بودن این تبدیل نسبت به تغییرات مقیاس است. اخیراً با ترکیب این تبدیل با تبدیلات دیگر، تبدیلاتی حاصل شده است که هم خصوصیات تبدیل میلین را دارد و هم خصوصیات تبدیل اضافه شده. بعنوان نمونه ویژگی تبدیل فوریه، نامتغیر بودن نسبت به تغییرات انتقال است. با ترکیب این تبدیل با تبدیل میلین تبدیلی حاصل می‌شود که آنرا تبدیل فوریه - میلین¹² می‌نامند و نسبت به تغییرات مقیاس و انتقال ثابت است. از آن جمله می‌توان به تبدیلات میلین - فوریه و میلین - میلین اشاره نمود [۸، ۱۱].

با توجه به خصوصیات ذکر شده می‌توان گفت که تبدیل میلین کاربرد وسیعی در زمینه تشخیص الگو دارد که از جمله کاربردهای مهم آن می‌توان به شناسایی هدف، مخفی سازی اطلاعات در تصاویر¹³ یا سیگنالهای صوتی، حذف نویز از تصاویر، شناسایی افراد با توجه به سیگنالهای صوتی ادا شده هر فرد، تحلیل و پردازشهای اولیه بر روی تصاویر اثر انگشت، و بازیابی تصاویر [۱۰] اشاره نمود. موارد فوق تنها نمونه‌ای کوچک از کاربردهای تبدیل میلین است. در این فصل قصد داریم تا در مورد کاربردهای جدید تبدیل میلین در زمینه حذف نویز از سیگنالهای صوتی صحبت کنیم.

¹² . Fourier – Mellin

¹³ . watermarking

یکی از مشکلاتی که در دهه‌های گذشته استفاده از تبدیل میلین را در زمینه‌های مختلف محدود می‌نمود، نبود یک الگوریتم مشخص و دقیق برای محاسبه این تبدیل، مخصوصاً در بحث سیگنالهای گسسته بود. اما با استفاده از الگوریتم تبدیل میلین سریع که در فصل ۳ در مورد آن صحبت نمودیم و الگوریتم‌های دیگری که در ادامه در مورد آنها صحبت خواهیم نمود، کاربرد این تبدیل روز به روز افزونتر می‌شود.

در این فصل ابتدا در مورد کاربردهای تبدیل میلین در حوزه‌های یک بعدی و بیشتر بر روی سیگنالهای صوتی و در ادامه در مورد کارایی آن در سیگنالهای دو بعدی و بخصوص پردازش تصاویر دیجیتال صحبت خواهیم نمود.

۴-۲- کاربرد تبدیل میلین در سیگنالهای صوتی

یکی از نکات مهمی که در تشخیص الگو مطرح می‌شود، استخراج ویژگی از الگو است. برای تحلیل سیگنالها معمولاً از حوزه‌هایی نظیر زمان و فرکانس استفاده می‌شود. بررسی سیگنالها فقط از لحاظ زمانی، نمی‌تواند در برگیرنده تمام اطلاعات و ویژگیهای موجود در یک سیگنال صوتی باشد. از اینرو بررسی سیگنالها در حوزه‌های دیگر (نظیر فرکانس، مقیاس و ...) برای استخراج ویژگیهای بیشتر می‌تواند مفید باشد [۱۵]. استفاده از تبدیلات فرکانسی، مانند تبدیل فوریه، به ما این امکان را می‌دهد تا بجای استفاده از حوزه زمان جهت نمایش سیگنال، از حوزه فرکانس استفاده نماییم. بعنوان نمونه با بررسی زمانی یک موج سینوسی که بوسیله نویز سفید خراب شده است، براحتی نمی‌توان اثرات نویز را بر روی سیگنال، مشاهده نمود، اما با بررسی این سیگنال در حوزه فرکانس، با استفاده از تبدیل فوریه سیگنال نویزی شده، می‌توان به وجود نویز پی برد.

همانطور که می‌دانید، زمان و فرکانس تنها حوزه‌های نمایش سیگنال نمی‌باشند، از طرفی بعضی از مواقع استفاده از یک حوزه به تنهایی نمی‌تواند برای نمایش سیگنال موثر باشد. بهمین دلیل از آنالیزهای پیچیده‌تری استفاده می‌شود. مثلاً استفاده از تحلیل فوریه زمان کوتاه^{۱۴} زمان و فرکانس را بهم مرتبط می‌سازد. با استفاده از این تبدیل می‌توان انرژی سیگنال را بر حسب زمان و فرکانس نمایش داد [۲].

¹⁴ . Short Time Fourier Transform (STFT)

یکی از سوالاتی که برای اکثر محققین مطرح می‌شود این است که آیا آنالیزهای زمان - فرکانس که بر اساس تبدیل فوریه می‌باشند، در همه کاربردها مناسب است. بعنوان نمونه هنگام بررسی دقت شنوایی انسان، آیا استفاده از تبدیلی که فرکانسهای منفی را حذف کند، درست است؟ بعضی‌ها به دنبال راه‌حلی برای این مشکل می‌باشند و عده‌ای دیگر سعی در وفق دادن تبدیل فوریه برای رفع این نیازها می‌باشند. بعنوان نمونه انتقال حوزه فرکانس^{۱۵} جهت بدست آوردن تبدیل Mel - Scale نتایج بهتری را، نسبت به تبدیل فوریه، در اختیار ما قرار می‌دهد.

یک سیگنال را می‌توان با استفاده از حوزه‌های مختلف مورد تجزیه و تحلیل قرار داد. بعنوان نمونه سیگنال صوت یک پدیده بسیار پیچیده است که می‌توان آنرا بر حسب تغییرات فشار هوا نسبت به زمان بیان نمود. به کمک این نمایش می‌توان زمان اتفاق افتادن وقایع مهم صوتی را محاسبه نمود. تبدیل میلین و نوع خاص آن بنام تبدیل مقیاس یک سیگنال را بر حسب مقیاس نشان می‌دهند. حوزه مقیاس می‌تواند همانند حوزه‌های دامنه، زمان و فرکانس تفسیر کننده یک سیگنال صوتی باشد. در این قسمت قصد داریم نشان دهیم که چگونه می‌توان بعضی از اثرات اصوات دیجیتال^{۱۶} را با استفاده از حوزه مقیاس^{۱۷} شناسایی نمود.

۴-۲-۱- تبدیل میلین زمان کوتاه

همانطور که می‌دانید ما با این فرض سیگنالهای صوت را سیگنالهای ایستا^{۱۸} در نظر گرفتیم که آنها را در بازه‌های زمانی کوچک (معمولا کمتر از ۲۰ میلی ثانیه) مورد بررسی قرار دهیم. با این فرض است که ما می‌توانیم از روابط آماری در سیگنالهای صوتی استفاده نماییم. با استفاده از تبدیل فوریه زمان کوتاه ما می‌توانیم در هر زمان دلخواه تبدیل فوریه سیگنال را بدست آوریم. در اینحالت تبدیل فوریه یک تابع دو متغیری بر حسب زمان و فرکانس است. مشابه با تبدیل فوریه زمان کوتاه، می‌توان از تبدیل میلین زمان کوتاه^{۱۹} استفاده نمود. در عمل، با استفاده از تابع پنجره $h(t)$ می‌توان تبدیل میلین زمان کوتاه را تعریف نمود که در آن:

$$M_f(p) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} f(\tau) h(\tau - t) e^{(p-1)\log\tau} d\tau. \quad (۴-۱)$$

¹⁵ . frequency warping

¹⁶ . Digital Audio Effect

¹⁷ . Scale domain

¹⁸ . stationary

¹⁹ . Short Time Mellin Transform (STMT)

تبدیل میلین زمان کوتاه، نمونه‌ای از هم پیوستن توزیعهای زمان - مقیاس است و دارای مشخصه‌های مشابه با تبدیل فوریه زمان کوتاه است. با استفاده از این تبدیل می‌توان مباحثی مانند فیلتر نمودن و حذف نویز از یک سیگنال صوتی را در حوزه میلین مورد بررسی قرار داد.

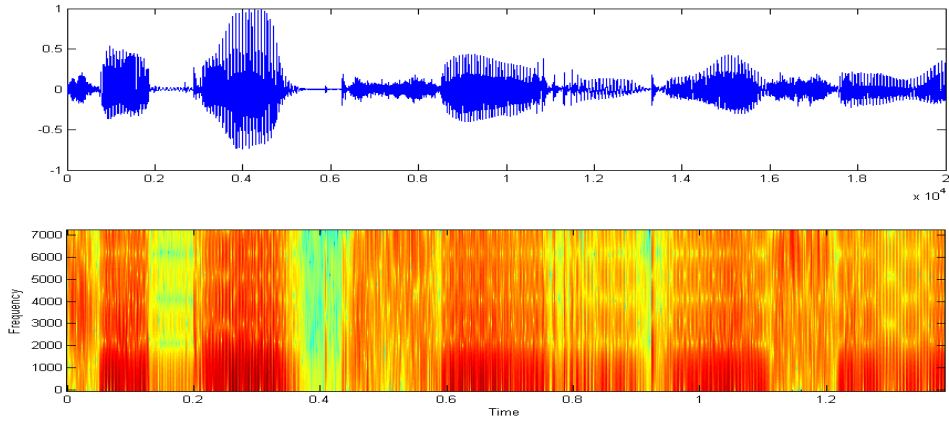
۴-۲-۲- فیلتر کردن در حوزه میلین

با تعریف تبدیل میلین زمان کوتاه می‌توان از مبحث فیلتر کردن در حوزه میلین استفاده نمود. در اینحالت فیلتر نمودن یک سیگنال مشابه با فیلتر نمودن آن در حوزه فرکانس است. بعنوان نمونه برای بدست آوردن یک سیگنال مقیاس علی، کفیسست تبدیل میلین سیگنال را در تابعی که به ازای مقیاسهای منفی صفر و در بقیه نقاط یک است، ضرب نمود، یا می‌توان یک فیلتر مقیاس باند میانی را با صفر نمودن تمام مقادیر مقیاس بجز آنهایی که در بازه مشخصی قرار دارند، طراحی نمود.

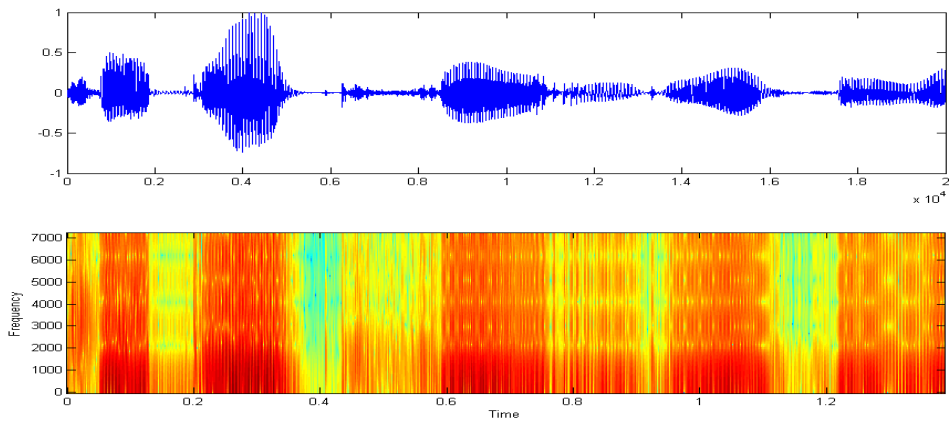
بطور مشابه، برای یک فیلتر مقیاس پایین‌گذر طراحی می‌توان با صفر نمودن تمام مولفه‌های دامنه‌ای که بین مقیاس قطع و ماکزیمم مقیاس سیگنال می‌باشند، یک فیلتر پایین‌گذر ایجاد نمود. بعنوان نمونه، همانطور در شکل ۴-۱ مدل زمانی و اسپکتروگرام جمله "your dark suit in greasy" که از مجموعه داده TIMIT گرفته شده است، نشان داده شده است. در شکل ۴-۲ نتیجه فیلتر نمودن در حوزه میلین برای سیگنال صوتی شکل ۴-۱ نشان داده شده است. با مشاهده نتایج حاصل از تبدیل میلین می‌توان به راحتی حذف فرکانسهای بالای، در قسمت‌های فرکانس بالا و عبارت دیگر در واحهای بیصدا، این جمله را با مقایسه طیف فرکانسی این دو جمله مشاهده نمود. اما نکته‌ای که در مقایسه با فیلتر نمودن در حوزه میلین و حوزه فوریه وجود دارد این است که در حوزه میلین فرکانسهای بالا با شدت کمتری کاهش می‌یابند.

در شکل ۴-۳ بالایی، فیلتر پایین‌گذر را می‌توان مشابه یک فیلتر پایین‌گذر متغیر با زمان در نظر گرفت. در اینحالت فرکانس قطع فیلتر بصورت نمایی به سمت صفر میل می‌کند. سرعت همگرایی به مقیاس قطع²⁰ بستگی دارد. فیلتر بالا‌گذر معکوس فیلتر پایین‌گذر است، یعنی دامنه تمام مولفه‌هایی که بین مقیاس شروع و مقیاس قطع می‌باشند، برابر با صفر می‌شود (شکل ۴-۳). پایینی). در مورد اثرات دیگر فیلتر کردن در حوزه میلین و فواید آن نسبت به فیلتر کردن در حوزه فوریه در ادامه و در بخش حذف نویز سیگنالهای صوتی، بیشتر صحبت خواهیم نمود.

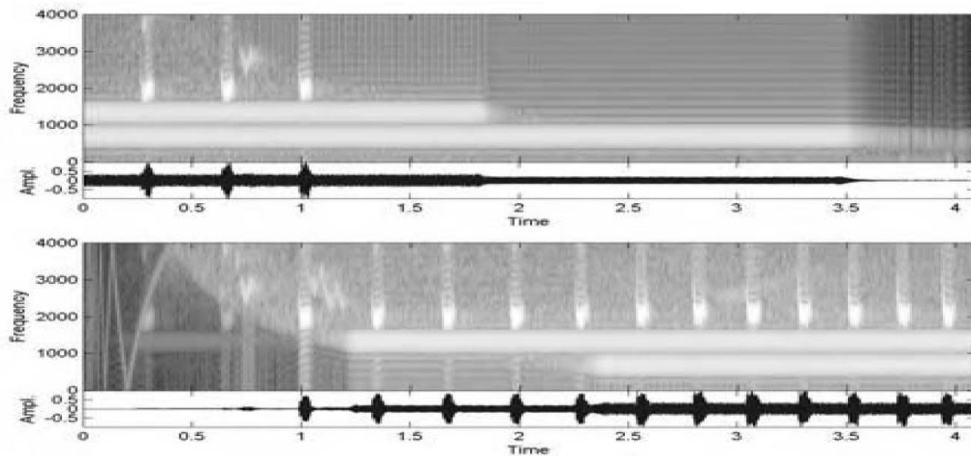
²⁰ . cutoff scale



شکل ۴-۱ مدل زمانی جمله "your dark suit in greasy" و اسپکتروگرام آن



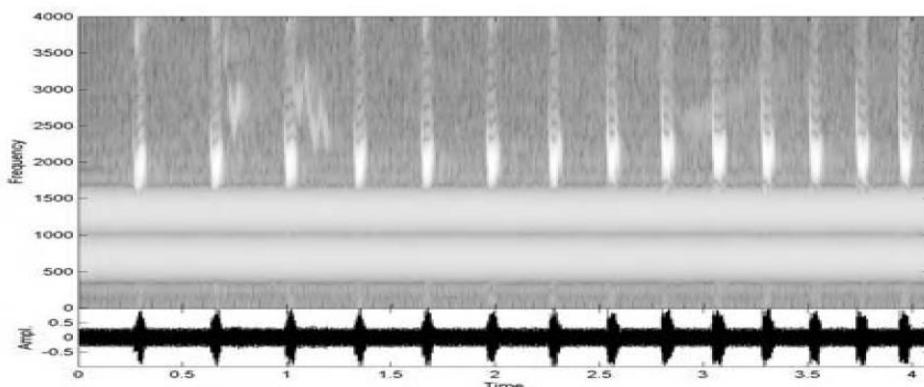
شکل ۴-۲ مدل زمانی جمله فیلتر شده "your dark suit in greasy" و اسپکتروگرام آن که توسط یک فیلتر پایین گذر در حوزه میلین صورت گرفته است.



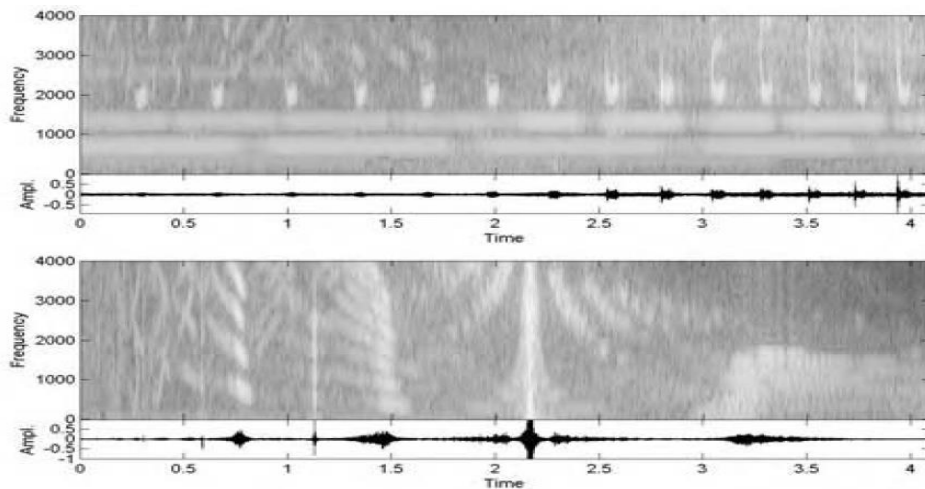
شکل ۴-۳ اسپکتروگرام سیگنال فیلتر شده پایین گذر (بالا) و بالا گذر (پایین) [۱].

۴-۲-۳- بازبایی سیگنال با استفاده از دامنه و یا فاز

یکی از خواص جالب تبدیل میلین این است که می توان با استفاده از فقط دامنه و یا فقط فاز تبدیل میلین سیگنال، آنرا بازبایی نمود. شکل ۴-۴ اسپکتروگرام مربوط به یک سیگنال را نشان می دهد. شکل ۴-۵ اسپکتروگرام سیگنال بازبایی شده را با جایگزاری پاسخ دامنه با یک ثابت و با جایگزاری پاسخ فاز نشان می دهد. دقت داشته باشید که بازبایی با استفاده از فاز، موقعیت زمانی وقایع مهم را که بخوبی آشکار هستند را حفظ می کند. این عمل شبیه قسمت سایه روشن تصویر بازبایی شده از تبدیل میلین است که در ادامه راجع به آن صحبت خواهیم نمود.



شکل ۴-۴ اسپکتروگرام سیگنال اصلی [۱]



شکل ۴-۵ سیگنال بازیابی شده با استفاده از فاز (بالا) و با استفاده از دامنه (پایین)

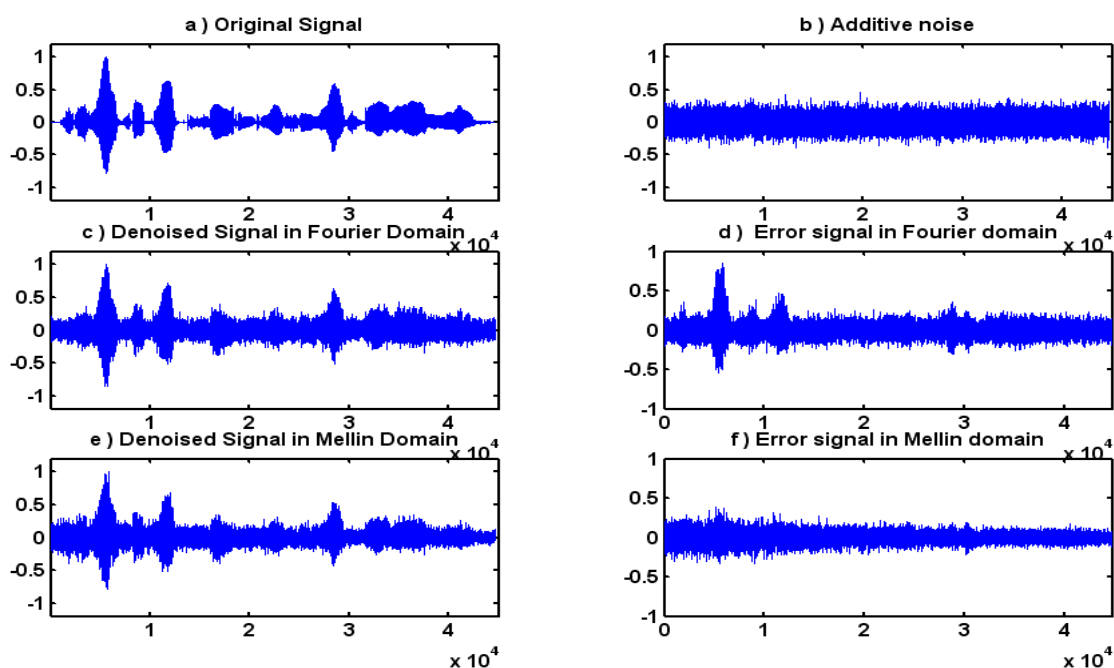
۴-۲-۴ کاهش نویز در سیگنالهای صوتی با استفاده از تبدیل میلین

یکی از روشهای ساده‌ای که برای کاهش و یا حذف نویز بکار می‌رود، استفاده از فیلترهای بالاگذر، میان‌گذر و یا پایین‌گذر است. فیلتر نمودن در حوزه زمان با کانوالو کردن سیگنال ورودی در فیلتر مورد نظر انجام می‌گیرد. این روش دارای پیچیدگی محاسباتی بالایی است و برای رفع این مشکل بجای آنکه یک سیگنال را در حوزه زمان فیلتر نماییم، با انتقال آن به حوزه فرکانس، در عین اینکه محاسبات ساده‌تر می‌شود، سرعت محاسبات نیز بیشتر می‌شود.

اما همانطور که می‌دانید استفاده از این فیلترها ممکن است باعث حذف یک سری از اطلاعات مهم سیگنال شود که این اطلاعات می‌تواند جهت استخراج ویژگی از یک سیگنال صوتی مهم باشد. یکی از مزایای مهم تبدیل میلین آن است که سعی در حفظ اطلاعات مهم یک سیگنال در زمینه نویزی دارد. برای این منظور در این قسمت سعی داریم تا با مقایسه نتایج فیلتر نمودن در حوزه فوریه و حوزه میلین کارایی این تبدیل را برای فیلتر نمودن نشان دهیم.

در شکل ۴-۶ مثالی از یک سیگنال صوتی نویزی آورده شده است که در آن یک سیگنال صوتی را با نویز سفیدی که دارای توزیع نرمال و توان ۳ دبی است، شکل ۴-۶. a، جمع نموده‌ایم. در این حالت نسبت سیگنال به نویز برابر با ۱۰ دبی است. با توجه به اینکه معمولاً در سیگنالهای صوت فرکانسهای پایین از درجه اهمیت بیشتری نسبت به فرکانسهای بالا برخوردارند، در این قسمت از یک فیلتر پایین گذر برای حذف نویز استفاده شده است. شکل ۴-۶. b، اختلاف بین سیگنال اصلی (سیگنال بدون نویز) و سیگنال فیلتر شده در حوزه فوریه را نشان می‌دهد. در این حالت اختلاف توان بین سیگنال بازیابی شده و سیگنال اصلی برابر با $1/3$ دبی است. ما کاهش نویز را با استفاده از همین فیلتر در حوزه میلیون انجام دادیم، که اختلاف بین سیگنال بازیابی شده و سیگنال اصلی در شکل ۴-۶. c، نشان داده شده است. در این حالت اختلاف توان بین سیگنال اصلی (سیگنال بدون نویز) و سیگنال بازیابی شده از حوزه میلیون برابر با 0.35 دبی است. میانگین مجذور خطا^{۲۱} در حوزه فوریه برابر با 2.065×10^{-3} و در حوزه مقیاس 1.59×10^{-3} است.

همانطور که در بالا به آن اشاره گردید، یکی از مزایای فیلتر کردن در حوزه میلیون در مقایسه با فیلتر کردن در حوزه فوریه، حفظ اطلاعات مهم یک سیگنال می‌باشد که این عامل را می‌توان با استفاده از نتایجی که در شکل ۴-۶ نشان داده شده است، براحتی مشاهده نمود.



²¹ . Mean Square Error (MSE)

شکل ۴-۶. فیلتر نمودن در حوزه میلین در مقایسه با فیلتر نمودن در حوزه فوریه. (a) سیگنال اصلی (b) سیگنال نویز با توان نویزی ۳ دبی (c) کاهش نویز با استفاده از فیلتر کردن سیگنال نویزی در حوزه فوریه ($MSE=2.065 \cdot 10^{-3}$). (d) اختلاف بین سیگنال فیلتر شده در حوزه فوریه و سیگنال اصلی. (e) کاهش نویز با استفاده از همان فیلتر در حوزه میلین ($MSE=1.59 \cdot 10^{-3}$). (f) اختلاف بین سیگنال فیلتر شده در حوزه میلین و سیگنال اصلی

۴-۳- تبدیل میلین دو بعدی و کاربردهای آن

همانطور که قبلا به آن اشاره گردید، اکثر کاربردهایی که در ابتدا در رابطه با تبدیل میلین وجود داشت، محدود به حوزه‌های یک بعدی می‌گردید، اما با گسترش سیستم‌های کامپیوتری و بالا رفتن سرعت آنها، ما به این قابلیت دست یافتیم که با استفاده از الگوریتم‌های مناسب بتوان تبدیل میلین دو بعدی را محاسبه نمود. تبدیل میلین دو بعدی بر اساس رابطه زیر بدست می‌آید:

$$M_{f(x,y)}(s_1, s_2) = \int_0^{\infty} \int_0^{\infty} x^{s_1-1} y^{s_2-1} f(x, y) dx dy \quad (۲-۴)$$

که در آن s_1 و s_2 متغیرهای مختلط‌اند. تحت شرایط مناسب تبدیل معکوس آن بصورت زیر بیان می‌شود:

$$f(x, y) = \frac{1}{2\pi d} \int_{h-i\infty}^{h+i\infty} \int_{k-i\infty}^{k+i\infty} M_{f(x,y)}(s_1, s_2) x^{-s_1} y^{-s_2} ds_1 ds_2 \quad (۳-۴)$$

با انتقال رابطه (۲-۴) از مختصات دکارتی به مختصات قطبی داریم:

$$M\{f(r)\} = \int_0^{\infty} f(r) r^{-js-1} dr, \quad r = \sqrt{x^2 + y^2} \quad (۴-۴)$$

و با تغییر متغیر $r = e^\gamma$ و بازنویسی رابطه فوق

$$M\{f(e^\gamma)\} = \int_0^{\infty} f(e^\gamma) e^{-js\gamma} d\gamma = F\{f(e^\gamma)\} \quad (۵-۴)$$

رابطه فوق بیان می‌کند که تبدیل فوریه دو بعدی برابر با تبدیل فوریه تابع انتقال یافته از مختصات دکارتی به مختصات لگاریتمی - قطبی²² است.

با جایگزاری $s_1 = -jc_1 + 1/2$ و $s_2 = -jc_2 + 1/2$ در رابطه (۴-۲) تبدیل مقیاس دو بعدی بدست می‌آید. بنابراین:

$$D(c_1, c_2) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \int_0^\infty f(x, y) \frac{e^{-jc_1 \ln x - jc_2 \ln y}}{\sqrt{xy}} dx dy \quad (۴-۶)$$

با توجه به اینکه اکثر پردازشهایی که برای تحلیل تصاویر صورت می‌گیرد، از طریق کامپیوتر انجام می‌گیرد، بهمین علت در بیشتر مواقع با سیگنالهای گسسته سرو کار داریم. از اینرو بایستی تبدیل میلین گسسته دو بعدی محاسبه گردد. یکی از روشهایی که برای محاسبه تبدیل میلین گسسته دو بعدی (یا حالت خاص آن یعنی تبدیل مقیاس) صورت می‌گیرد، با استفاده از رابطه (۴-۵) بدست می‌آید که شامل مراحل زیر می‌باشد:

قدم ۱: با استفاده از عملگر انتقال (انتقال حوزه زمان به یک حوزه جدید) سیگنال تصویر اصلی را به مختصات لگاریتمی - قطبی انتقال می‌دهیم. اگر $I(x, y)$ ماتریس تصویر در مختصات دکارتی به مرکز (x_0, y_0) باشد، در اینصورت تبدیل لگاریتمی - قطبی بصورت زیر بیان می‌گردد:

$$I^*(\rho, \phi) = \ell\{I(x, y); (x_0, y_0)\} \quad (۴-۷)$$

که در رابطه فوق ℓ نگاشت از (x, y) به (ρ, ϕ) بیان می‌گردد که در آن:

²² . Log – Polar coordinate

$$I^*(\rho(x, y), \phi(x, y)) = \frac{1}{\#(A_{\eta, \xi})} \sum_{(x, y) \in A_{\eta, \xi}} I(x, y)$$

$$A_{\eta, \xi} = \{(x, y) \mid (\rho(x, y), \phi(x, y)) = (\eta, \xi)\},$$

$$\rho(x, y) = \rho(r(x, y)) = [M \log(r(x, y) + \alpha) - \beta + \frac{1}{2}], \text{ with}$$

$$\beta \geq M \log \alpha$$

$$r(x, y) = \sqrt{(x - x_0)^2 + (y - y_0)^2}, \text{ and}$$

$$\phi(x, y) = \left[\frac{n}{2\pi} \tan^{-1} \left(\frac{y - y_0}{x - x_0} \right) + \frac{1}{2} \right].$$

α, M و β پارامترهای مثبتی هستند که باید تعیین گردند [۹].

قدم ۲: بعد از انتقال تصویر از مختصات دکارتی به مختصات لگاریتمی - قطبی بایستی تبدیل فوریه دو بعدی آن محاسبه گردد. تبدیل فوریه گسسته دو بعدی از رابطه زیر بدست می آید:

$$F\{I(m, n)\} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(m, n) e^{-j(2\pi/M)(km)} e^{-j(2\pi/N)(ln)} \quad (۹-۴)$$

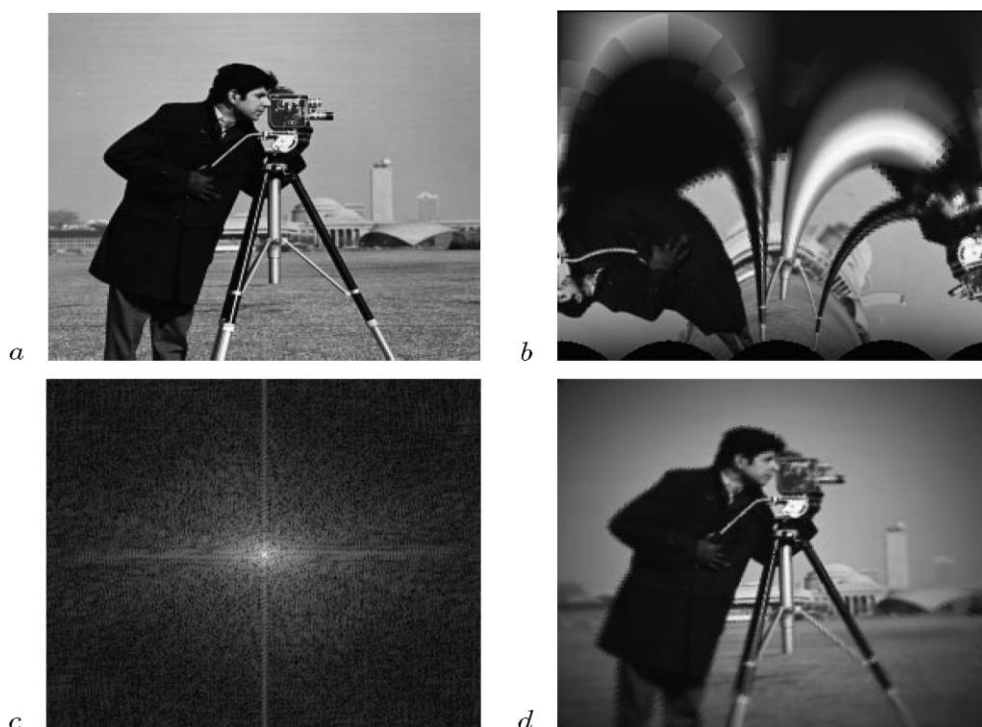
در حالت گسسته باید دقت نمود که از هر خطایی که منجر به اعوجاج می گردد، جلوگیری شود. بدین منظور قبل از محاسبه تبدیل فوریه از پنجره همینگ استفاده می شود [۵].

در شکل ۴-۷ مثالی از محاسبه تبدیل مقیاس یک تصویر نشان داده شده است. شکل ۴-۷ a، تصویر اصلی را نشان می دهد. مبدا را مرکز تصاویر در نظر می گیریم. تبدیل فوریه نسبت به انتقال ثابت می باشد، عبارت دیگر، دامنه تبدیل فوریه تابع $f(t)$ و $f(t+T)$ با هم برابر است، در نتیجه با توجه به توضیحات فوق تبدیل مقیاس مستقل از مبدا است. شکل ۴-۷ b، تصویر انتقال یافته از مختصات دکارتی به مختصات قطبی و سپس نمونه برداری شده لگاریتمی آنرا نشان می دهد که در نهایت در یک پنجره همینگ ضرب شده است. برای نمونه برداری لگاریتمی از درون بای دو سوپه²³ استفاده شده است. درونیایی دو سوپه، بسط داده شده درونیایی خطی برای توابع درونیاب دو متغیری یک شبکه منظم²⁴ است که درونیایی در دو راستای x و y انجام می گیرد. با استفاده از این درون بایی همسایه های نزدیک پیدا می گردد. شکل ۴-۷ c، دامنه

²³ . Bilinear Interpolation

²⁴ . regular network

تبدیل مقیاس را بعد از محاسبه تبدیل فوریه شکل ۴-۷. b، نشان می‌دهد و در نهایت شکل ۴-۷. d، نتایج تبدیل معکوس مقیاس را نشان می‌دهد. در این تصویر می‌توان براحتی افزایش رزولوشن فرکانسی را در مرکز مشاهده نمود.



شکل ۴-۷ مراحل عملکرد تبدیل مقیاس (a تصویر اصلی، b) انتقال از حوزه دکارتی به لگاریتمی - قطبی، c) دامنه تبدیل مقیاس، d) تبدیل معکوس مقیاس. افزایش رزولوشن فرکانسی بعلاوه نمونه برداری لگاریتمی از تصویر اصلی است. [۳]

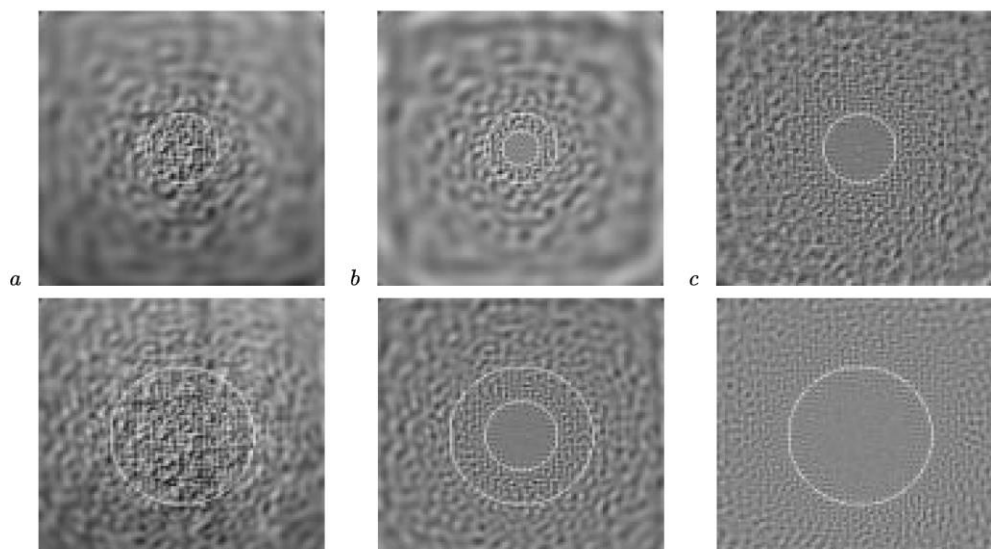
۴-۳-۱- فیلتر نمودن تصاویر در حوزه میلین

یکی از روشهایی که برای بررسی تبدیل میلین در تصاویر می‌توان انجام داد، فیلتر کردن تصاویر در حوزه میلین است. بنابراین، اگر تبدیل میلین یک تصویر را محاسبه نماییم و سپس مقادیر مقیاس بالای آنرا حذف نماییم، با معکوس نمودن تبدیل میلین براحتی می‌توان دید که مقیاسهای بزرگ تصویر حذف شده‌اند.

شکل ۴-۸ مثالی از فیلتر نمودن تصویر را در حوزه مقیاس (میلین) با سه فیلتر متفاوت با سه محدوده مقیاس متفاوت و دو سایز مختلف، نشان می‌دهد. شکل ۴-۸. a-c، فیلتر کردن تصویر متنی را در مقیاسهای پایین، بالا و میانی، نشان می‌دهد. فیلترهای بالاگذر و پایین‌گذر، بترتیب اجازه عبور مقیاسهای قرار گرفته بین دایره داخلی (a,d) و دایره خارجی (e,f) را می‌دهند و باند میانی امکان عبور مقیاسهای بین دو دایره را در (b,c) می‌دهد. در همین قسمت می‌توان بعضی از تفاوت‌های فیلتر نمودن در حوزه فوریه را دید. اول اینکه در حوزه مقیاس، فیلتر پایین‌گذر فرکانسهای بالایی را حذف نمی‌کند. فیلتر نمودن مقیاس پایین‌گذر، یک فیلتر متغیر زمانی تولید می‌کند که بیشترین رزولوشن آن در (a,d) است و در اطراف آن کاهش می‌یابد. فیلتر کردن مقیاس بالاگذر معکوس حالت پایین‌گذر عمل می‌کند و در اطراف مرکز تصویر رزولوشن کاهش می‌یابد. برای فیلتر باند میانی رزولوشن بیشتری در ناحیه حلقه‌ای بدست می‌آید.

۴-۳-۲- کاهش نویز در تصاویر

همانطور که در قسمتهای قبلی دید، تبدیل میلین سعی در حفظ اطلاعات مهم سیگنال دارد، بهمین علت بنظر می‌رسد کاهش نویز در حوزه مقیاس نتایج بهتری را نسبت به حوزه فوریه دارد. برای اولین بار موسز به این نکته اشاره کرد که می‌توان با استفاده از تبدیل فوریه و میلین اطلاعات بسیار مهم را از زمینه نویزی بیرون کشید. اخیرا کیزر اشاره کرده است که رابطه تجربی بسیار قدرتمندی در مورد فواید کار کردن در حوزه مقیاس وجود دارد، چرا که این تبدیل سعی در هموار نمودن طیف بدون وارد نمودن لطمه‌ای به مقیاس دارد. بعنوان نمونه، شکل ۴-۹. a، یک تصویر



شکل ۴-۸ a) تصویر فیلتر پایین گذر شده در حوزه مقیاس (شعاع ۳۲) b) فیلتر شده با باند میانی (شعاعهای ۱۶ و ۳۲) c) فیلتر شده بالاگذر (مقیاسهای انرژی بیشتر از ۳۲). d-e) مشابه با قسمت بالا ولی ابعاد شعاع دو برابر شده است. عملیات فیلتر کردن با حذف مقیاسهایی که خارج محدوده مورد نظر ماست، صورت گرفته است [۵].

نویزی شده گوسی با واریانس ۹۰۰ را نشان می‌دهد. شکل ۴-۹ b، نتیجه فیلتر نمودن را در حوزه فوریه با استفاده از یک فیلتر پایین گذر نشان می‌دهد. اگرچه در این حالت نویز مقداری کم شده است، اما جزئیات لبه‌ها از بین رفته است. شکل ۴-۹ c، نتیجه فیلتر نمودن را در حوزه مقیاس نشان می‌دهد. در قسمت مرکزی تصویر که رزولوشن افزایش یافته است، مشاهده می‌شود که کارایی بیشتری نسبت به شکل ۴-۹ b، دارد.

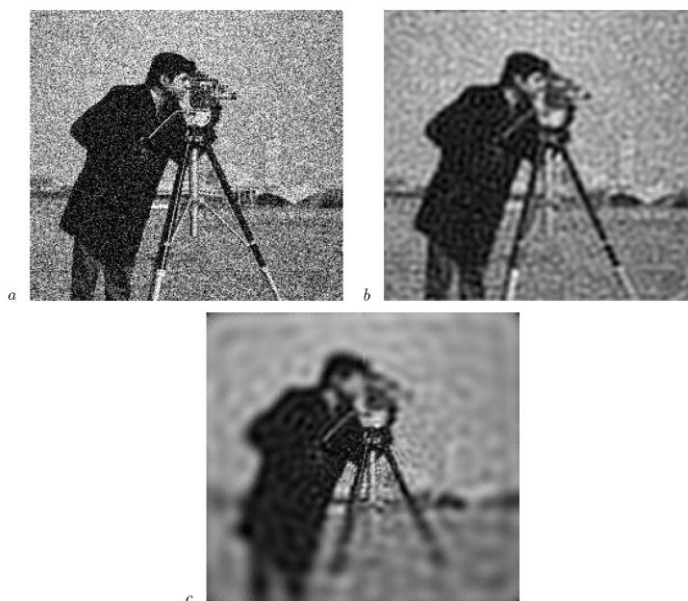
یکی از مشکلات اصلی استفاده از تبدیل مقیاس برای کاهش نویز، تغییرات رزولوشن مکانی تصویر بازیابی شده است. این امر باعث تغییر نقطه تمرکز ما برای افزایش کارایی کاهش نویز تصویر می‌شود (مرکز تبدیل). برای رفع این مشکل تصویر را به ابعاد کوچکتری در می‌آوریم و در هر قسمت تبدیل میلین (مقیاس) را محاسبه می‌کنیم. شکل ۴-۱۰ مثالی از تصویر نویزی شده Cameraman را نشان می‌دهد که با استفاده از نویز گوسی با واریانس ۹۰۰ و نسبت سیگنال به نویز ۱۳,۵۳ دبی انجام گرفته است. شکل ۴-۱۰ b، نتیجه فیلتر نمودن پایین گذر (Butterworth) را در حوزه فوریه نشان می‌دهد. اگرچه با اینکار نویز زمینه حذف شده است، اما باعث شده است که جزئیات لبه‌ها از بین رود. شکل ۴-۱۰ c، نتیجه همان فیلتر را در حوزه مقیاس نشان داده شده است.

کاهش نویز در شکل ۴-۱۰ a، با تغییر مرکز تصویر از بین ۶۴ پنجره بدون هیچگونه تداخل با ابعاد ۳۲*۳۲ پیکسل و در نهایت اجرای فیلتر مقیاس صورت گرفته است. نسبت سیگنال به نویز برای تصویر

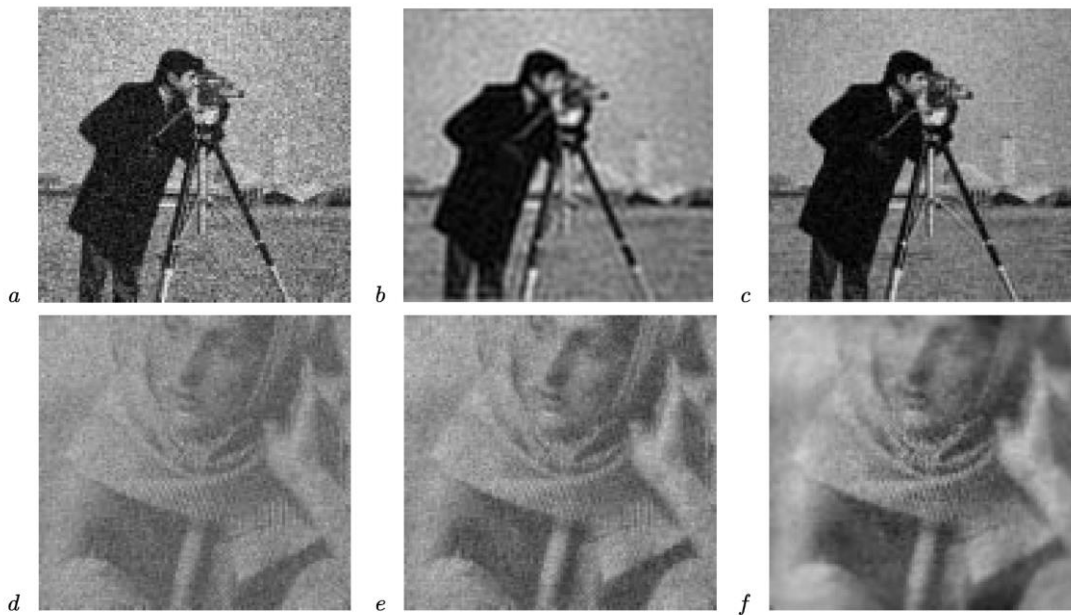
بازیابی شده ۱۷,۳۲ دیبی است. اگر ابعاد پنجره‌ها تصادفی انتخاب گردد، نتایج مقداری بهتر شده و نسبت سیگنال به نویز در حدود ۱۷,۵۵ دیبی می‌شود. شکل ۴-۱۰. d، مثال دیگری از کاهش نویز را برای یک تصویر با نویز زیاد، نشان می‌دهد که نسبت سیگنال به نویز آن در حدود صفر است. شکل ۴-۱۱ تصویر اصلی و نویزی شده را به همراه طیف مقیاس هر یک از این تصاویر نشان می‌دهد.

۴-۳-۳- تشخیص الگوهای نامتغیر با چرخش، مقیاس و انتقال

تبدیل فوریه - میلین یک ابزار ریاضی مفید برای تشخیص الگو است، چرا که نتیجه حاصل از آن نسبت به تغییرات چرخش، انتقال و مقیاس ثابت است. تبدیل فوریه خود به تنهایی نسبت به انتقال بدون تغییر است و با انتقال آن از مختصات دکارتی به مختصات لگاریتمی-قطبی، اختلافهای مقیاس و چرخش به افست‌های افقی و عمودی تبدیل می‌کند که قابل اندازه‌گیری می‌باشند. بلوک دیاگرام تبدیل فوریه - میلین در شکل ۴-۱۲ نشان داده شده است

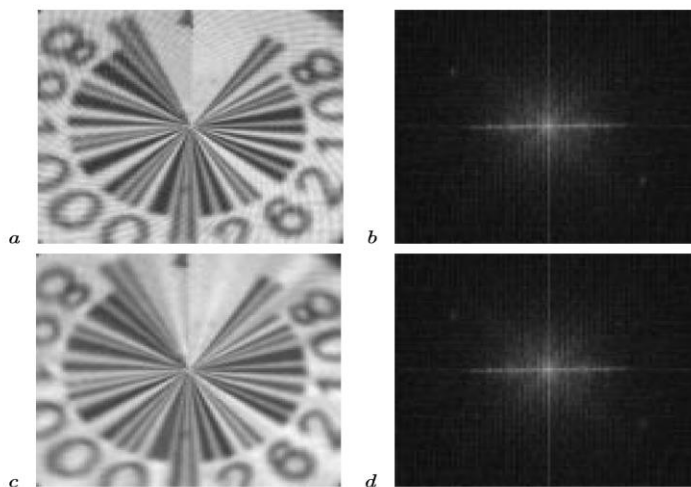


شکل ۴-۹ فیلتر نمودن با استفاده از تبدیل مقیاس و مقایسه آن با تبدیل فوریه. (a) تصویر نویزی شده گوسی، (b) فیلتر نمودن با استفاده از یک فیلتر پایین گذر در حوزه فوریه، (c) فیلتر کردن با همان فیلتر در حوزه مقیاس



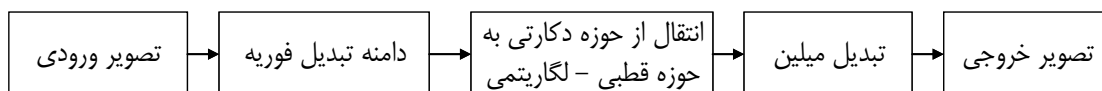
شکل ۴- ۱۰ دو مثال از فیلترنمودن در حوزه مقیاس در مقایسه با فیلتر نمودن در حوزه فوریه. (a) تصویر Cameraman با نویز گوسی. (b) فیلترنمودن در حوزه فوریه با استفاده از فیلتر پایین گذر Butterworth (شعاع ۳۲ پیکسل) (c) فیلترنمودن در حوزه مقیاس با فیلتر قسمت b. کاهش نویز با استفاده از قسمت نمودن تصویر. (d) تصویر با نویز بسیار بالا (e) کاهش نویز تصویر با استفاده از فیلتر Gabor در حوزه فوریه (۴ فرکانسهای و ۴ جهتی) (f) کاهش نویز تصویر با استفاده از حوزه مقیاس (چهار مقیاس و ۴ جهتی)

[۵]



شکل ۴-۱۱ تجسم نويز در حوزه مقياس (a) تصوير اصلي (b) تبديل مقياس شكل (a)، طيف براي درك بهتر بصورت لگاريتمي مقياس‌بندی شده است. (c) تصوير بعد از حذف نمودن پيکهای مربوط به (a) (d) تبديل مقياس تصوير (a) بعد از حذف نمودن پيکها

[۴]



شکل ۴-۱۲ بلوک دياگرام تبديل فوريه - ميلين

با توجه به اينکه نگاهت مختصات دکارتی به فضای قطبی - لگاریتمي یک به یک نیست، بايد میانگين پيکسلهای مجاور محاسبه شود. روشهای متداول براي انجام اينکار شامل نزديکترین همسايه، Bilinear و نمونه‌برداری دوباره Bicubic است. تبديل فوريه - ميلين براي اهدافی مانند شناسایی تصوير استفاده می‌شود. همچنين می‌توان از آن براي تشخيص چهره با مقایسه طيف بدست آمده از تصوير با مجموعه طيفهای از قبل محاسبه شده استفاده نمود.

٤-٤- مراجع

- [1] Antonio De Sena and Davide Rocchesso. "A Fast Mellin transform With Application IN DAFx," Proc of the 7 th int. conference on digital Audio Effect (DAFx04), Naples. Italy, October 5-8, 2004.
- [2] Antonio De Sena and Davide Rocchesso. "A Fast Mellin and Scale Transform". EURASIP Journal on advances in Signal Processing Volume 2007.
- [3] G. Cristobal and L. Cohen. "Scale in images". In SPIE Proc. Advanced Signal Processing. Algorithms, Architectures and Implementations, volume 2846, pages 251-261, 1996.
- [4] Riaan van den dool, "Image Processing Tools, Fourier-Mellin Transform," Stellenbosch University, Faculty of Engineering, May2004.
- [5] G. Cristobal, J. Cuesta, "Image filtering and denoising through the scale transform," In SPIE Proc. Advanced Signal Processing. Algorithms, Architectures and Implementations, 1996.
- [6] H. Sundarm, S. Joshi, and R. Bhatt. "Scale periodicity and its sampling theorem," IEEE Trans. On signal Processing, Vol 45, no 7, pp.1862-1864, July 1997.

- [7] L. Cohen. "The Scale representation," IEEE Trans. On signal Processing, vol. 41, no.12, pp. 3275-3291, December 1993.
- [8] J.R. Martinez-de Diosy and A. Ollero, "Real-Time Image Stabilization System Based on Fourier-Mellin Transform, Lecture Notes in Computer Science, Vol. 3211, Pp376-383.
- [9] Richard Alan Peters II, "On the Computation of the Discrete Log-Polar Transform," Department of Electrical and Computer Engineering, 22 April 2007.
- [10]Stephane Derrode and Faouzi Ghorbel, "Robust and efficient Fourier-Mellin transform approximation for gray-level image reconstruction and complete invariant description," Pattern Recognition Letters, 15:1043-1051, October 2004.
- [11]Y.Sheng and C. Lejeune, "Invariant pattern recognition using Fourier-Mellin transform and neural network," J. of Optics, 22(5):223-228, 1991.
- [12]L- Cohen, "Time-frequency analysis," Prentice Hall, Englewood Cliffs, 1995.
- [13]G. Kaiser, "Wavelet filtering with the Mellin transform," Appl. Math. Letters, to appear.
- [14]T. Irino and R. D. Patterson, "Segregating information about the size and the shape of the vocal tract using time domain auditory model: The stabilized wavelet-Mellin transform," Speech Communication, vol 36, no. 3, pp. 181-203, March 2002.
- [15]Antonio De Sena, "The Scale domain," PhD Proposal, November 1, 2005.

سیگنال‌های گفتار و سیستم‌های تشخیص اتوماتیک گفتار

۵-۱- مقدمه

تشخیص اتوماتیک گفتار با استفاده از ماشین، هدف اصلی بسیاری از محققان در چهار دهه اخیر بوده است. گرچه، تمام سعی محققین بر این است تا بتوانند ماشینی بسازند که بتواند تمام اصوات را از یکدیگر تشخیص دهد، اما هنوز از آنچه که مدنظر ماست، یعنی شناسایی اصوات در هر محیطی و گفته شده توسط هر فردی، فاصله زیادی داریم. بنابراین سوال اصلی در این فصل این است که معنی تشخیص صوت با استفاده از ماشین چیست.

برای جداسازی کاربردهای تحلیل زبان از سیستم‌های تحلیل داده دیگر، از آموخته‌های اولیه هر زبان استفاده می‌شود. تشخیص اتوماتیک گفتار²⁵ یکی از تکنولوژیهای موفق است که امکان ارتباط بین ماشین و انسان را مهیا می‌سازد تا از این طریق بتوان بعضی از اطلاعات را از آن درخواست نمود یا درخواست انجام بعضی از کارهای مشخص را، با استفاده از ارتباط زبانی، داشت. حوزه هوش مصنوعی یکی از زمینه‌های گسترش الگوریتم‌های تشخیص اتوماتیک گفتار است.

الگوها و نمونه‌های بکار رفته در سیستم‌های تشخیص اتوماتیک گفتار دارای یک سری اطلاعات زبانی از سیگنالهای گفتار می‌باشند، که این اطلاعات در زبانهای مختلف با یکدیگر تفاوت دارند، اما بدون در نظر گرفتن یک سری تغییرات، مانند پایینی و بلندی صدا، مدت زمان، و توالی، بیشتر از ۳۰ نوع واج²⁶ وجود ندارد. واجها کوچکترین بخش هر زبانی می‌باشند. با در نظر گرفتن تمام تغییرات فوق تعداد واجها افزایش می‌یابد. اخیرا بعضی از محققان برای ارتقاء سیستمهای تشخیص اتوماتیک گفتار از سیلابها²⁷ بجای واجها استفاده می‌کنند، چون معمولا روشی که مغز انسان برای شناسایی زبان استفاده می‌کند، ذخیره‌سازی و شناسایی سیلابهاست نه واجها [۲].

تشخیص اتوماتیک گفتار بعلاوه تغییرات بسیار وسیع موجود در آن، مانند لهجه یا طرز تلفظ، میزان بلندی صدا، شرایط سلامتی و خستگی فرد، کاری بسیار دشوار است. در نتیجه در سیستمهای تشخیص گفتار اتوماتیک، برای کاربردهای خاص یا کلی، جنبه‌های زیادی از این تغییرات باید مد نظر قرار گیرد.

اگر بخواهیم نگاه کلی بر سیستمهای تشخیص اتوماتیک گفتار داشته باشیم، می‌توان آنها را بصورت زیر خلاصه نمود:

در اوایل سال ۱۹۵۰ آزمایشگاه بل یک سیستم تشخیص اتوماتیک گفتار ساخت که قادر به شناسایی شماره‌ها از یکدیگر بود. آزمایشگاههای RCA یک تشخیص اتوماتیک گفتار تک گوینده را ساخت که می‌توانست ۱۰ سیلاب را از یکدیگر تشخیص دهد. دانشگاهی در انگلیس یک تشخیص دهنده واج ساخت و در آزمایشگاه ایالت لینکلن شایر انگلستان یک تشخیص دهنده آوای مستقل از گوینده ساخته شد. در طی دهه ۱۹۷۰، چندین سیستم جهت تشخیص جملات جدا شده ساخته شد. در طی دهه ۱۹۸۰ اولین الگوریتم

²⁵ .Automatic Speech Recognition (SAR)

²⁶ .phoneme

²⁷ .syllables

برای سیستمهای تشخیص گفتار با واژگان زیاد اجرا گردید. همچنین در این زمان مدل مخفی مارکوف²⁸ و شبکه های عصبی²⁹ برای گسترش و ارتقاء سیستمهای تشخیص اتوماتیک گفتار معرفی گردیدند. یکی از این سیستمها، سیستم SPHINX می باشد. سیستمهای تشخیص اتوماتیک گفتار در دهه ۹۰ در سیستمهای مخابراتی و اطلاع رسانی بکار رفتند. در نهایت در سالهای اخیر، برای ما استفاده از سیستمهای تشخیص اتوماتیک گفتار در سیستمهای عامل، سیستمهای ارتباط تلفنی، و سایتهای اینترنتی ممکن شده است. موتورهای جستجوی شبکه های وب که بر اساس صدای انسان عمل می کنند یکی از این کاربردها است.

همانطور که در بالا به آن اشاره گردید، یکی از مشکلات اساسی انجام تحقیقات در زمینه تشخیص صوت با استفاده از ماشین، وسیع بودن محدوده تحقیق است و همین امر باعث می شود تا برای حل یک مسئله تحقیقات زیادی در زمینه های مختلف صورت گیرد. موضوعاتی که برای یک سیستم تشخیص صوت باید مد نظر قرار گیرند، بترتیب عبارتند از:

۱- **تحلیل سیگنال** پروسه ای شامل جداسازی اطلاعات مطلوب از سیگنال صوت است. بعنوان نمونه تحلیل طیفی یکی از روشهایی است که برای طبقه بندی خصوصیات سیگنال صوت مورد استفاده قرار می گیرد تا بدین وسیله سیگنال صوت نسبت به تغییرات محیط ضبط شده، حساس نباشد.

۲- **فیزیک (صوت شناسی)** درک علمی صحیحی از رابطه بین فیزیک سیگنال صوت و مکانیزمهای فیزیولوژیکی (مکانیزم حفره صوتی انسان) که صوت را تولید نموده و مکانیزمی که بوسیله آن صوت دریافت می شود (مکانیزم شنوایی انسان).

۳- **تشخیص الگو** مجموعه ای از الگوریتمها که برای داده ها بکار می رود تا یک یا چند نمونه اولیه از الگو تولید نماید و بر اساس اندازه گیری یک سری ویژگیها، الگوها را با یکدیگر مقایسه نماید.

۴- **تئوری اطلاعات و ارتباطات** مراحل تخمین پارامترهای مدل های آماری؛ روشهایی برای شناسایی وجود الگوهای صوت، مجموعه ای از الگوریتمهای کد کننده و رمزگشای پیشرفته (مانند برنامه های

²⁸ .Hidden Markov Model (HMM)

²⁹ .neural network

دینامیکی، الگوریتم‌های پشت‌های³⁰ و رمزگشای ویتربای³¹ برای یافتن یک شبکه بزرگ ولی محدود برای ارائه بهترین جواب از بین مسیرهای تشخیص داده شده از کلمات، استفاده می‌گردد.

۵- زبان‌شناسی³² رابطه بین اصوات، کلمات در زبان، معنی جملات ادا شده و احساس دریافت شده از معانی. به موارد فوق می‌توان روش شناسایی گرامر و تجزیه زبان را اضافه نمود.

برای داشتن یک سیستم تشخیص صوت کارآمد بایستی اطلاعات کاملی در مورد هر یک از مسایل فوق داشته باشیم. بهمین علت دانستن یکسری مسایل اساسی برای درک مناسب از تشخیص صوت برای محققین امری ضروری بنظر می‌رسد.

۵-۲- اصوات گفتار و ویژگیهای آنها

تعداد اصوات صوت، مستقل از جنبه زبانی (واج) در هر زبانی با زبان دیگر متفاوت است. جدول ۵-۱ لیستی خلاصه شده از علامتهای اختصاری واجها در زبان امریکایی انگلیسی را نشان می‌دهد، و برای هر یک از آنها یک کلمه بعنوان نمونه آورده شده است. همانطور که در این جدول نشان داده شده است، ۴۸ صوت داریم که ۱۸ تای آنها آوا³³ یا ترکیبی از آوا (آنها را اصطلاحاً صدای ترکیبی³⁴ می‌نامند)، ۴ حرف صامت شبه آوا، ۲۱ حرف صامت استاندارد، ۴ صوت هجایی و یک واج بنام مکث حنجره‌ای³⁵ است.

تعداد زیادی از اصوات یا واجهای نشان داده شده در جدول ۵-۱ بصورت استاندارد نمی‌باشند. حالت کلی‌تر و عبارت دیگر نمایشی که بیشتر برای طبقه‌بندی واجها بکار می‌رود، در شکل ۵-۱ نشان داده شده است که برای کلاس اصوات نگلیسی امریکایی است. همانطور که در این شکل می‌بینید، در اینجا ما ۱۱ آوا، که شامل جلو، وسط و عقب می‌باشند، ۴ آوای ترکیبی؛ ۴ شبه آوا که به روان و نرم تقسیم می‌شوند؛ حروف

³⁰ . Stack algorithms

³¹ .Viterbi decoding

³² .linguistic

³³ .vowel

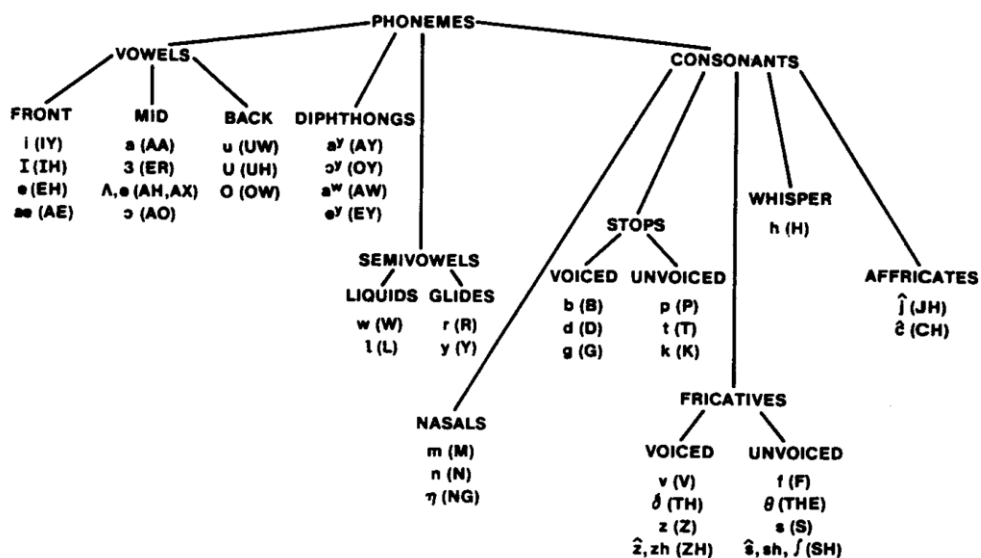
³⁴ .Diphthong

³⁵ .glottal stop

صامت دماغی، حروف صامت صدا دار و بی صدا؛ مکثهای صدا دار و بی صدا؛ صداهای نجوایی و سایشی داریم.
در مجموع ۳۹ تا از ۴۸ صوت جدول ۵-۱ در شکل ۵-۱ نشان داده شده‌اند.

جدول ۵-۱ لیستی خلاصه شده از علامت اختصاری واجهای آمریکایی - انگلیسی

واحد	علامت اختصاری	مثال	واحد	علامت اختصاری	مثال
/i/	IY	beat	/ŋ/	NX	sing
/I/	IH	bit	/p/	P	pet
/e/ (eʲ)	EY	bait	/t/	T	ten
/ɛ/	EH	bet	/k/	K	kit
/æ/	AE	bat	/b/	B	bet
/ɑ/	AA	Bob	/d/	D	debt
/ʌ/	AH	but	/g/	H	get
/ɔ/	AO	bought	/h/	HH	hat
/o/ (oʷ)	OW	boat	/f/	F	fat
/U/	UH	book	/θ/	TH	thing
/u/	UW	boot	/s/	S	sat
/ə/	AX	about	/ʃ/ (sh)	SH	shut
/i/	IX	roses	/v/	V	vat
/ɜ/	ER	bird	/ð/	DH	that
/ā/	AXR	butter	/z/	Z	zoo
/aʷ/	AW	down	/ʒ/ (zh)	ZH	azure
/aʲ/	AY	buy	/tʃ/ (tsh)	CH	church
/ɔʲ/	OY	boy	/dʒ/ (dzh,j)	JH	judge
/y/	Y	you	/w/	WH	which
/w/	W	wit	/I/	EL	battle
/r/	R	rent	/em/	EM	bottom
/l/	L	let	/ŋ/	EN	button
/m/	M	met	/ɾ/	DX	batter
/n/	N	net	/ʔ/	Q	(glottal Stop)



شکل ۵-۱ نمودار طبقه‌بندی واجهای استاندارد انگلیسی آمریکایی به کلاسهای صوتی [۱].

۵-۲-۱- آواها

مطمئناً، یکی از مهمترین کلاسهای صوتی در زبان انگلیسی آواها می‌باشند. در صحبت، آواها با تغییر شکل دادن حفره صوتی توسط پالسهای شبه پریودیک³⁶ هوا، باعث لرزش تارهای صوتی³⁷ می‌گردند. تغییرات سطح چند مقطعی³⁸ حفره صوتی تعیین کننده فرکانسهای تارهای عصبی (فرکانس رزونانس) است که در نتیجه آن صوت ایجاد می‌گردد. اصولاً، آواهای تولید شده با توجه به موقعیت حفره دهان و زبان کوچک تعیین می‌شوند، اما موقعیت فک و لبها نیز می‌تواند در آوای ایجاد شده تاثیرگذار باشد.

معمولاً، طول آواها از لحاظ مدت زمانی بلند و طولانی می‌باشد (در مقایسه با اصوات صامت) و یکی از بهترین روشها برای نمایش آنها، استفاده از طیف فرکانسی آنهاست. چندین روش برای توصیف و طبقه‌بندی آواها وجود دارد که از آن جمله می‌توان به نمودارهای زمانی و طیف فرکانسی اشاره نمود. بعنوان نمونه، شکل ۵-۲ نمودار زمانی چند آوا و شکل ۵-۳ اسپکتروگرام³⁹ این آواها را نشان می‌دهد. یکی از ساده‌ترین روشهایی که می‌توان برای طبقه‌بندی آواها از آن استفاده نمود، موقعیت زبان کوچک است (بعنوان نمونه، جلو، عقب، وسط). بر اساس این طبقه‌بندی آواهای /æ/, /i/, /ɪ/ و /ε/ آواهای جلویی، /a/, /ʌ/ و /ə/ آواهای میانی و /u/, /U/ و /o/ آواهای عقبی می‌باشند.

همانطور که در نمودار زمانی آواها (برحسب میلی ثانیه) در شکل ۵-۲ نشان داده شده است، آواهای جلویی دارای فرکانسهای رزونانس بالاتر، آواهای میانی دارای انرژی یکسان در طول باند فرکانسی و آواهای عقبی دارای اطلاعات طیفی در فرکانسهای پایین می‌باشند [۳]. این رفتار را می‌توان در نمودار اسپکتروگرام آواها که در شکل ۵-۳ نشان داده شده است، بخوبی مشاهده نمود. همانطور که در این شکل می‌بینید،

³⁶.quasi-periodic pulses

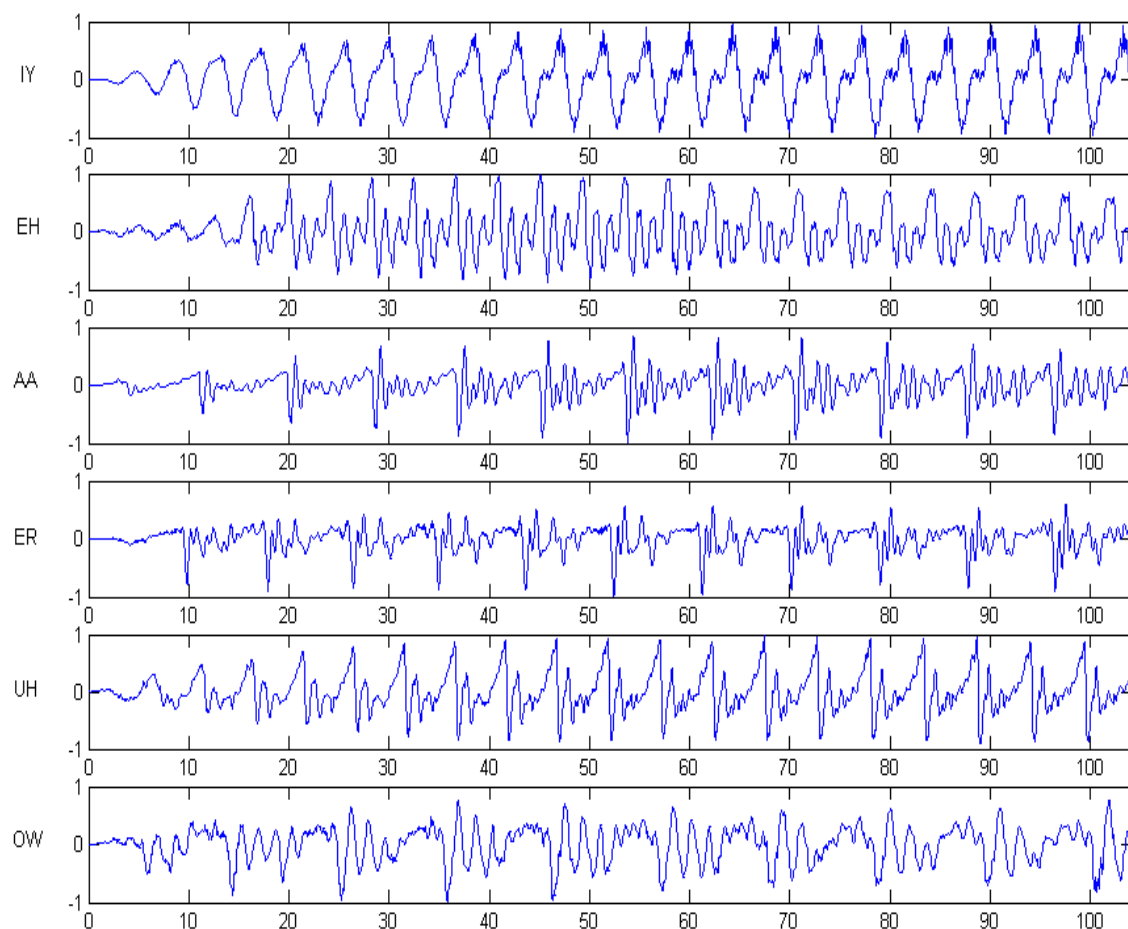
³⁷.vocal cords

³⁸.cross-sectional area

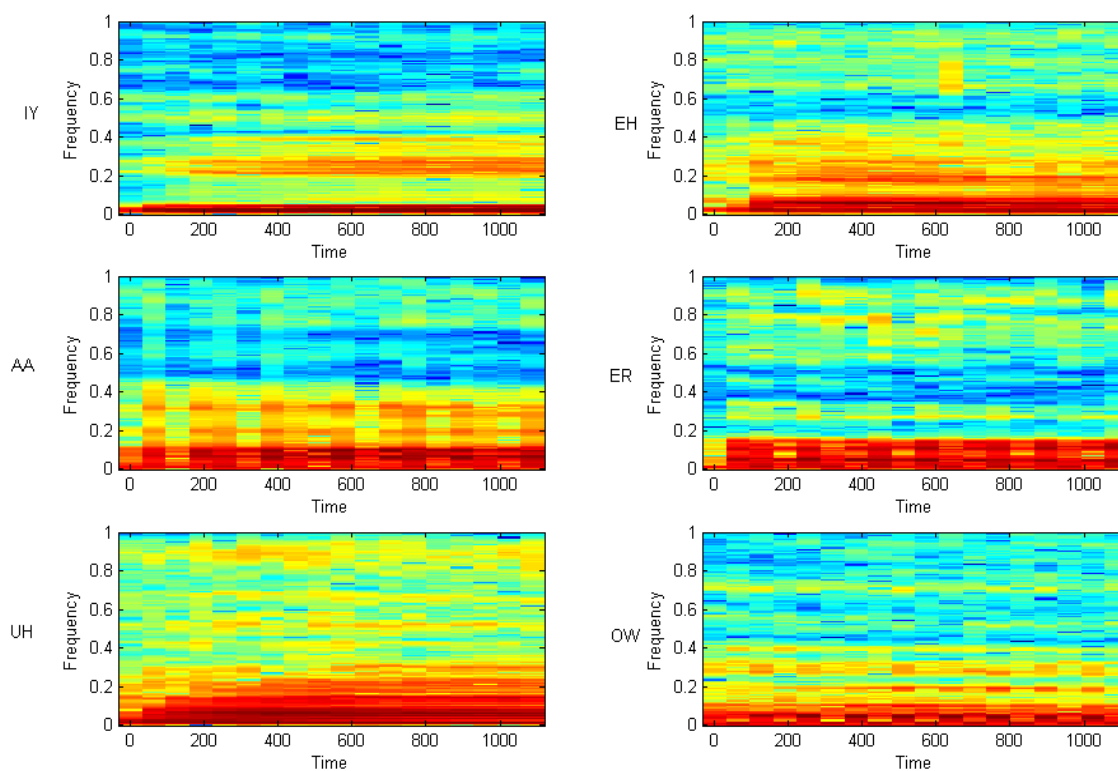
³⁹.Spectrogram

آواهای جلویی دارای فرکانسهای رزونانس دوم و سوم نسبتا بالایی می‌باشند، در حالیکه آواهای میانی دارای فرکانسهای رزونانس کاملا متمایز از یکدیگر می‌باشند و آواهای عقبی (مخصوصا /u/) در محدوده فرکانس پایین، تقریبا هیچگونه انرژی ندارد. نکته‌ای که باید به آن دقت نمود این است که آواهای گفته شده توسط جنسیت‌های مختلف نیز با هم متفاوت می‌باشند. برای روشنتر شدن این مطلب، شکل ۵-۴ نمودار کلاسیک ، بدست آمده توسط Paterson و Barney را نشان می‌دهد که در آن مقادیر اولین و دومین فرکانس رزونانس برای ۱۰ آوا ادا شده توسط محدوده وسیعی از زنها، مردها و بچه‌ها محاسبه شده است.

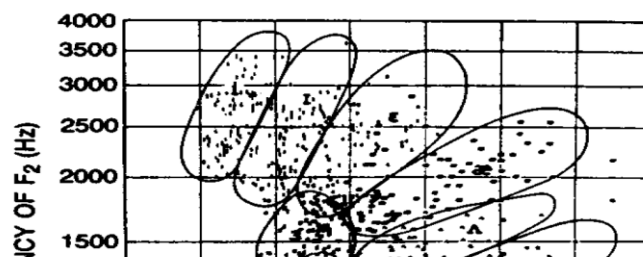
همانطور که در این شکل نشان داده شده است، محدوده تغییرات زیادی برای فرکانسهای رزونانس یک آوای مشخص وجود دارد، و همچنین بین آواهای مختلف ادا شده توسط افراد مختلف همپوشانی وجود دارد. بیضی رسم شده در این شکل بیانگر مشخصه‌های کلی یک ناحیه است که در آن تمام آواهای ادا شده یکسان قرار گرفته‌اند. این شکل به ما پیام ساده‌ای را می‌دهد و آن این است که اندازه‌گیری فرکانسهای رزونانس یا پیک‌های طیف فرکانسی نمی‌تواند معیار مناسبی برای طبقه‌بندی آواها باشد.



شکل ۵-۲ نمودار زمانی چند آوای مختلف

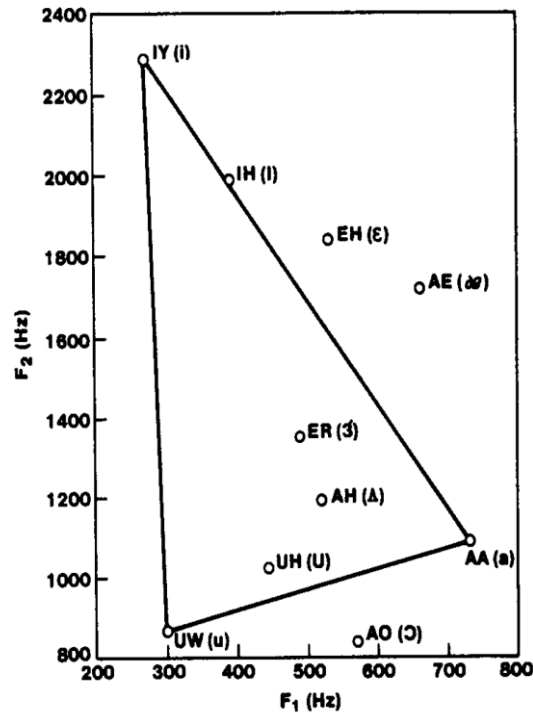


شکل ۵-۳ اسپکتروگرام آوای شکل ۵-۲



شکل ۴-۵ فرکانسهای اندازه‌گیری شده اولین و دومین فرکانس رزونانس برای رنج وسیعی از گوینده‌ها به ازای چندین آوای مختلف [۴]. یکی از روشهایی که از آن می‌توان برای استخراج ویژگی⁴⁰ از تصاویر ۳-۵ و ۴-۵ استفاده نمود، نشان دادن هر آوا با فرکانسهای رزونانس مرکزی فضای رزونانس است، بعبارت دیگر هر آوا را با میانگین محدوده فرکانسهای رزونانس آن آوا نشان دهیم. نتیجه انجام چنین کاری در شکل ۵-۵ آورده شده است و در جدول ۵-۲ هر آوا بر حسب فرکانسهای رزونانس نمایش داده شده است.

⁴⁰.feature extraction



شکل ۵-۵ مثلث آوایی که هر نقطه بیانگر میانگین فرکانس رزونانس هر آوا، ادا شده توسط افراد مختلف است [۴].

۵-۲-۲- صداهای ترکیبی

اگرچه مقداری ابهام در مورد اینکه چه صوتی، صدای ترکیبی است، وجود دارد، اما یک تعریف قابل قبول این است که صدای ترکیبی یک صوت تک سیلابی است. بر اساس این تعریف در زبان آمریکایی انگلیسی شش صدای ترکیبی وجود دارد که عبارتند از: /a^y/ (مانند کلمه buy)، /a^w/ (مانند down)، /e^y/ (مانند bait)، /ə^y/ (مانند boy)، /o/ (مانند boat) و /ju/ (مانند you). شکل ۵-۶ نمودار زمانی ۴ صدای ترکیبی که توسط چهار گوینده مرد گفته شده است را نشان می‌دهد و در شکل ۵-۷ اسپکتروگرام مربوط به این اصوات ترسیم شده است.

جدول ۵-۲ فرکانسهای رزونانس چند آوا

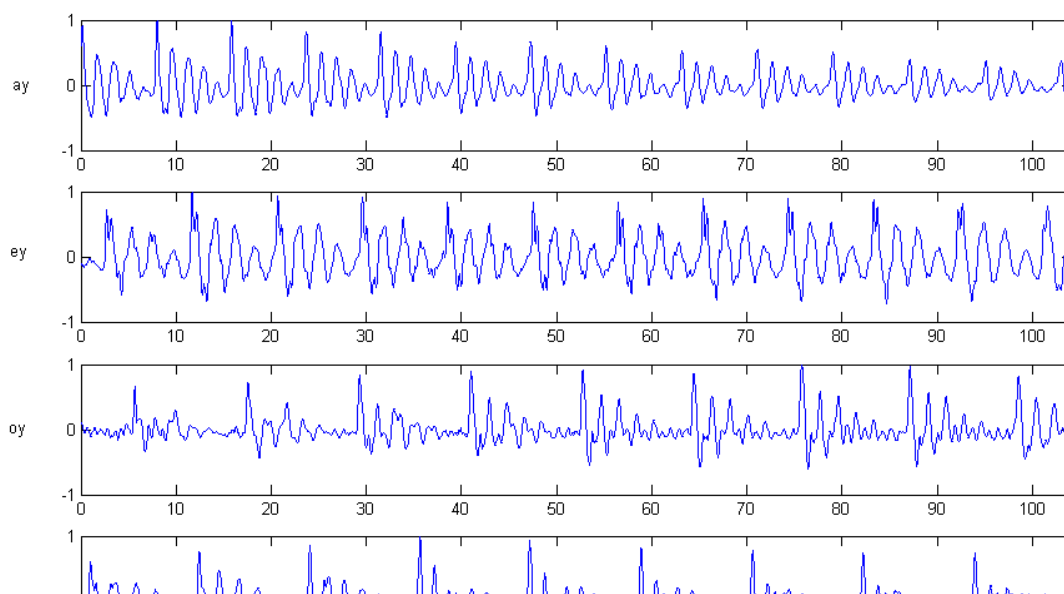
علامت اختصاری	نماد	کلمه نمونه	F ₁	F ₂	F ₃
هر آوا					
IH	/I/	bit	390	1990	2550
IY	/i/	beet	270	2290	3010
EH	/ε/	bet	530	1840	2480

AE	/æ/	bat	660	1720	2410
AH	/^/	but	520	1190	2390
AA	/a/	hot	730	1090	2440
AO	/ə/	bought	570	840	2410
UH	/U/	foot	440	1020	2240
UW	/u/	boot	300	870	2240
ER	/ɜ̃/	bird	490	1350	1690

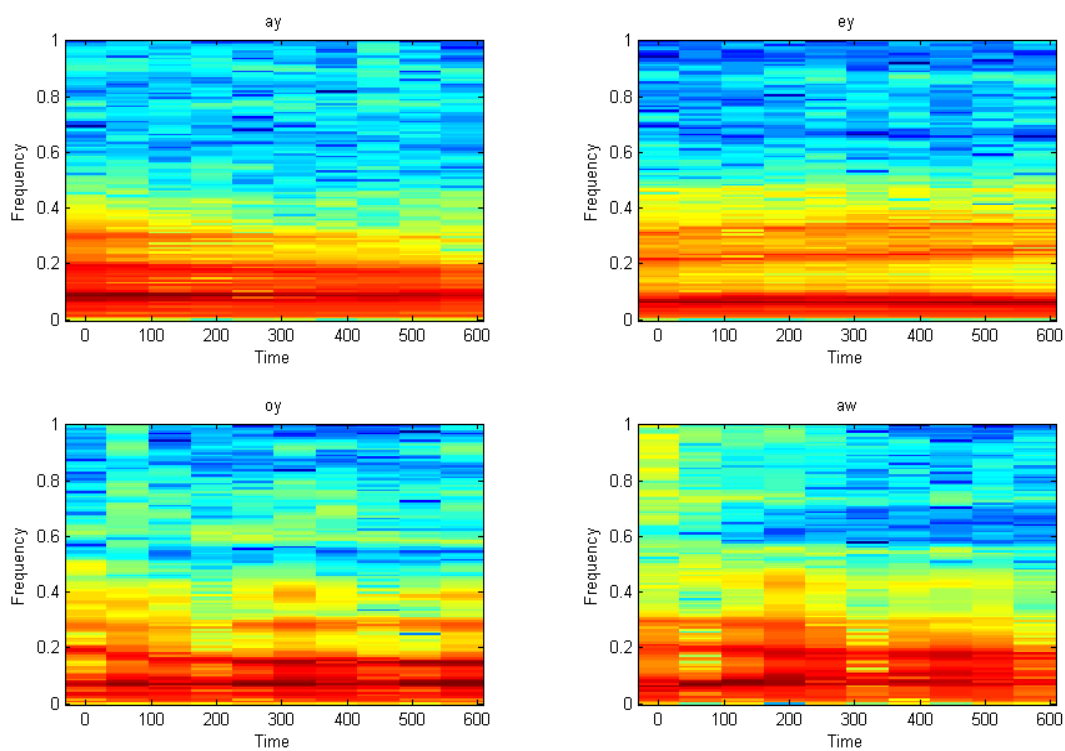
همانطور که در شکل ۵-۷ نشان داده شده است، میزان جابجایی فرکانسهای رزونانس برای اصوات $/a^w/$ ، $/a^y/$ و $/ə^y/$ قابل توجه است، در حالیکه این مقدار برای $/e^y/$ خیلی کم است. یکی از روشهایی نمایش تغییرات زمانی مشخصه‌های طیفی اصوات ترکیبی، رسم مقادیر فرکانس رزونانس دوم به ازای فرکانس رزونانس اول است که در شکل ۵-۸ نشان داده شده است. پیکانهای تصویر، جهت حرکت فرکانسهای رزونانس (در صفحه $(F_1 - F_2)$) را با افزایش زمان نشان می‌دهند. دایره‌های خط‌چین در این شکل موقعیت متوسط آواها را نشان می‌دهند. بر اساس این داده‌ها، و اندازه‌گیریهای دیگر، اصوات ترکیبی را می‌توان بر اساس تغییرات زمانی تابع مساحت حفره صوتی که بین دو آوا تغییر می‌کند، طبقه‌بندی نمود.

۵-۲-۳- شبه آواها⁴¹

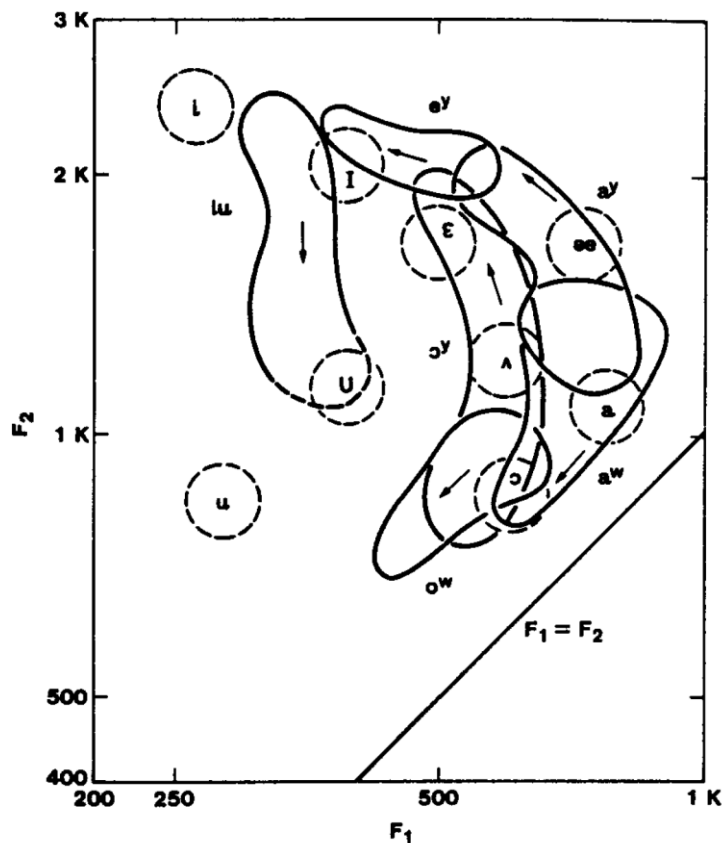
طبقه‌بندی گروهی از اصوات مانند $/r/$ ، $/l/$ ، $/w/$ و $/y/$ بسیار مشکل می‌باشد. این اصوات را از آن جهت که ماهیتی شبیه آواها دارند، شبه آوا می‌نامند. این دسته از اصوات را معمولاً بر اساس تغییر تابع مساحت حفره صوتی بین واجهای نزدیک بهم طبقه‌بندی می‌کنند. ماهیت این اصوات بسیار شبیه به آواها و اصوات ترکیبی است.



شکل ۵-۶ نمودار زمانی مربوط به اصوات ترکیبی /aʷ/، /əʷ/، /oʷ/ و /aʷ/



شکل ۵-۷ اسپکتروگرام اصوات ترکیبی شکل ۵-۶



شکل ۵-۸ تغییرات اولین دو فرکانس رزونانس در اصوات ترکیبی [۵]

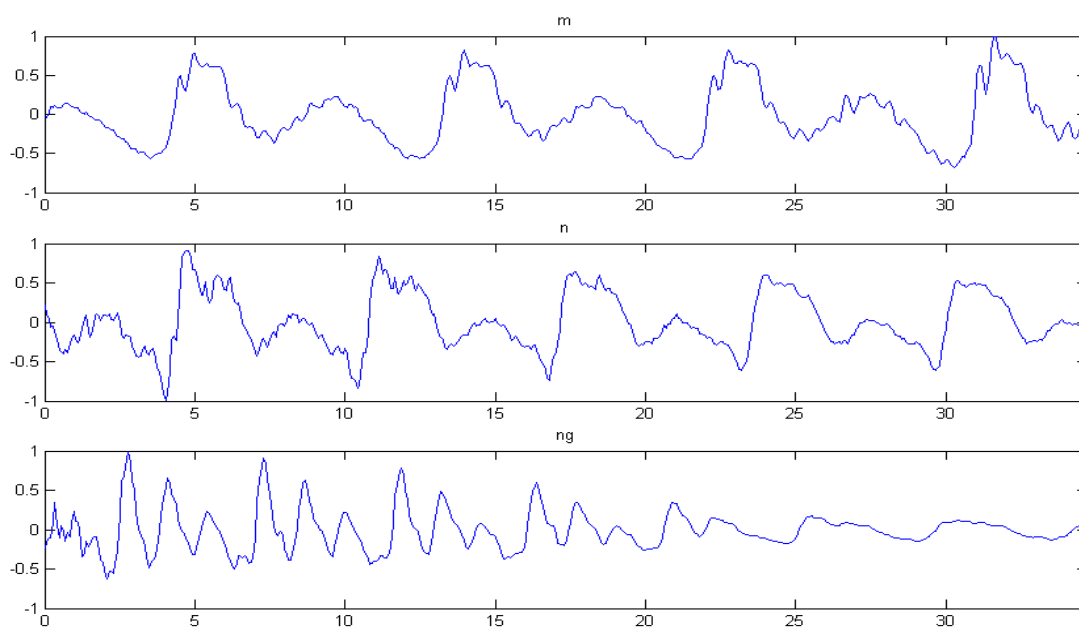
۵-۲-۴ - حروف بیصدای تو دماغی^{۴۲}

حروف بیصدای تو دماغی /m/، /n/ و /ŋ/ با تحریک دهانه نای^{۴۳} و منقبض شدن حفره صوتی در بعضی از نقاط مسیر دهان ایجاد می‌گردد. یکی از راههای تشخیص این سه حرف بیصدا مکانی است که بیشترین انقباض در آنجا صورت می‌گیرد. برای /m/ انقباض در لبهاست؛ برای /n/ انقباض در پشت دندانها و برای /ŋ/ انقباض در جلوی دهان است.

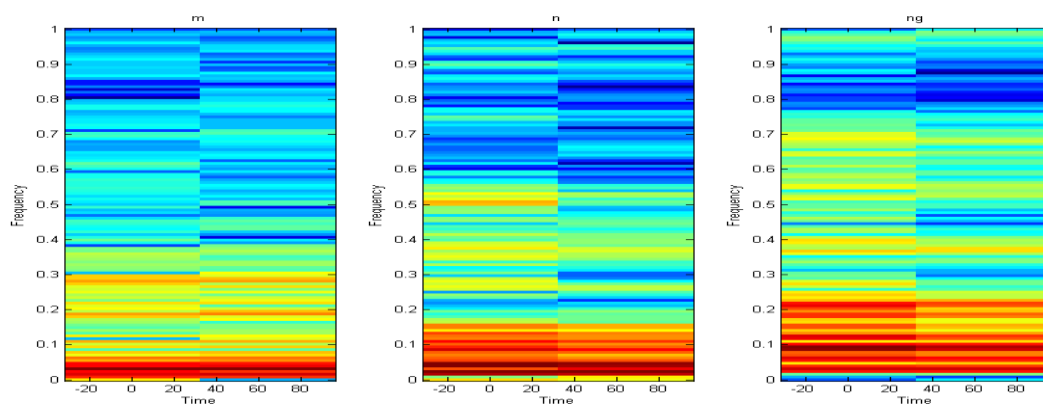
^{۴۲}.Nasal Consonants

^{۴۳}.glottal

شکل ۵-۹ نمودار زمانی چند حرف بیصدای تو دماغی را نشان می‌دهد. همانطور که در این شکل نشان داده شده است، شکل موج /m/ و /n/ بسیار شبیه هم است. شکل ۵-۱۰ اسپکتروگرام حروف بیصدای تو دماغی را نشان می‌دهد. با مشاهده این شکل می‌فهمیم که تراکم انرژی در فرکانسهای پایین زیاد است و در فرکانس میانی، هیچگونه پیک قابل توجهی وجود ندارد.

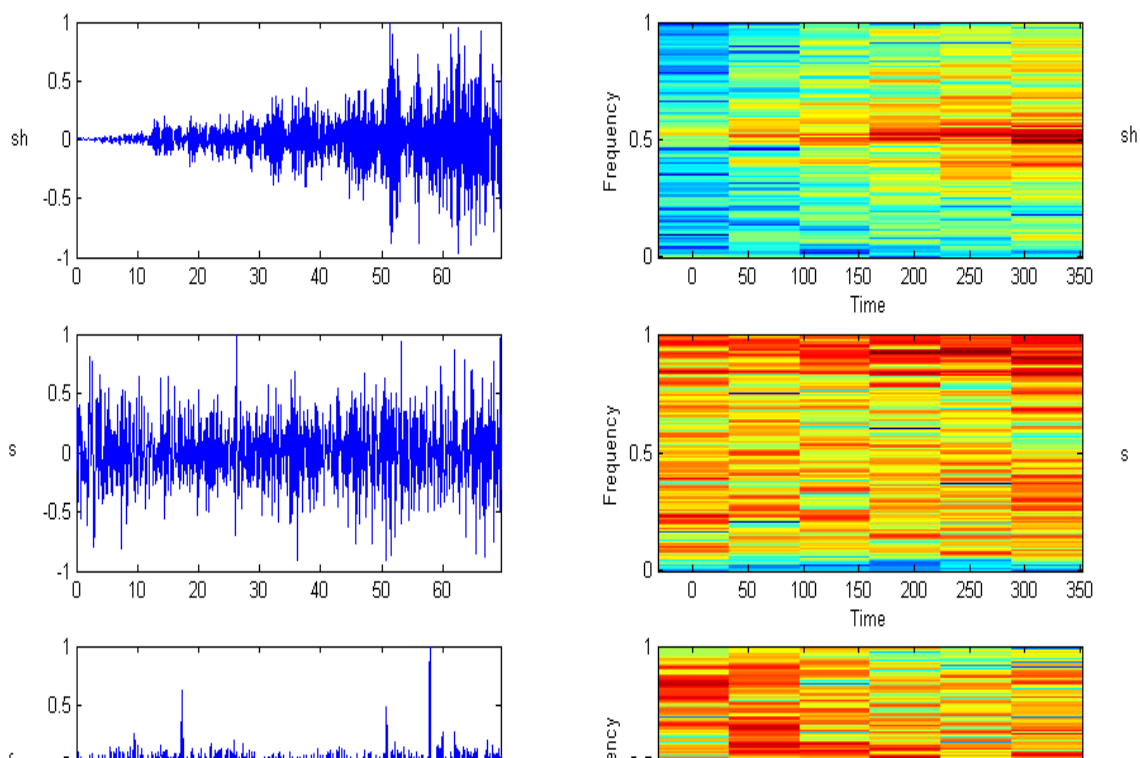


شکل ۵-۹ نمودار زمانی مربوط به اصوات /m/، /n/ و /ŋ/



۵-۲-۵- اصوات سایشی بیصدا

اصوات سایشی بیصدای /f/، /θ/، /s/ و /sh/ توسط تحریک حفره صوتی با جریان هوای ثابت ایجاد می‌گردد که این امر باعث انقباض قسمتهایی از حفره صوتی می‌گردد. بعبارت دیگر، اصوات سایشی اصواتی هستند که بر اساس اصطکاک نفس و وقفه تنفس ادا می‌گردند. قسمتهایی از حفره صوتی که انقباض صورت می‌گیرد، تعیین کننده نوع صوت سایشی تولید شده است. برای صوت سایشی /f/ انقباض نزدیک لبهاست؛ برای /θ/ نزدیک دندانها؛ برای /s/ این انقباض نزدیک وسط حفره دهان است؛ و برای /sh/ نزدیک پشت حفره دهان است. بنابراین سیستمی که برای تولید اصوات سایشی بیصدا استفاده می‌گردد، شامل یک منبع نویز در قسمتی که حفره صوتی به دو حفره جدا تقسیم می‌شود، است. عامل ایجاد صوت، لبها می‌باشند که تشکیل دهنده حفره جلویی است. حفره عقبی، که در اینحالت بینی است، وظیفه حبس نمودن انرژی و بدنبال آن ایجاد فرکانسهای صفر در داخل خروجی صوتی است. شکل ۵-۱۱ نمودار زمانی و اسپکتروگرام اصوات سایشی /f/، /s/ و /sh/ را نشان می‌دهد. خاصیت غیر پریودیک اصوات سایشی در شکل موجهای رسم شده کاملاً مشهود است. اختلاف طیفی در بین اصوات سایشی را میتوان با مقایسه این سه اسپکتروگرام براحتی مشاهده نمود.

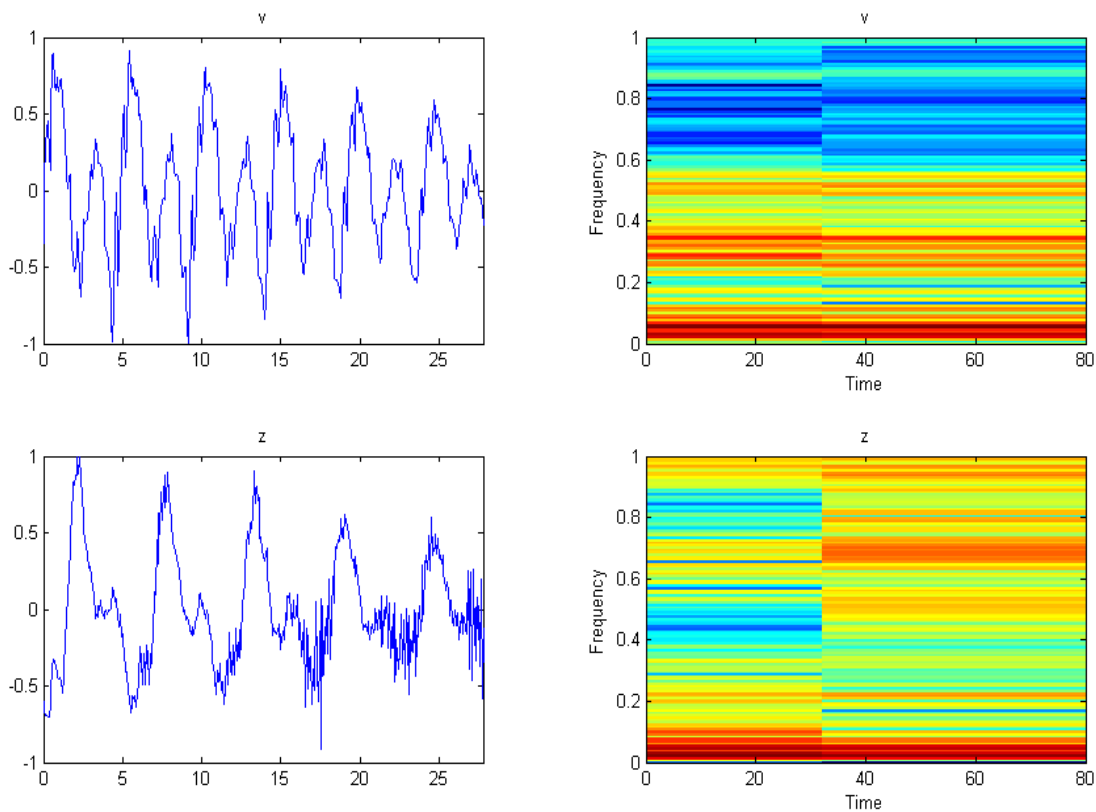


شکل ۵-۱۱ نمودار زمانی (سمت چپ) و اسپکتروگرام (سمت راست) اصوات سایشی بیصدای /f/، /s/ و /sh/

۵-۲-۶- اصوات سایشی صدا دار⁴⁵

اصوات سایشی صدا دار /v/، /th/، /z/ و /zh/ بترتیب نقطه مقابل اصوات سایشی /f/، /θ/، /s/ و /sh/ می‌باشند. بهر حال، اصوات سایشی صدا دار بطور قابل ملاحظه‌ای با اصوات سایشی بیصدا تفاوت دارند، چون در اصوات سایشی صدا دار دو منبع تحریک در تولید صوت دخالت دارند، در حالیکه برای اصوات سایشی بیصدا یک منبع تحریک وجود دارد. برای اصوات سایشی صدا دار تارهای صوتی بصورت نوسانی عمل می‌کنند و در نتیجه یکی از منابع تحریک در دهانه حنجره است. بهر حال با توجه به اینکه حفره صوتی در بعضی از نقاط جلویی دهانه حنجره منقبض می‌گردد، شارش هوا باعث اغتشاش در نزدیکی انقباضات می‌گردد. بنابراین طیف اصوات سایشی صدا دار را می‌توان بر حسب دو مولفه مستقل از هم رسم نمود. این ویژگیهای تحریک را به آسانی می‌توان در شکل ۵-۱۲ مشاهده نمود که نمودار زمانی و اسپکتروگرام دو صوت سایشی صدا دار /v/ و /z/ نشان داده شده است.

⁴⁵ Voiced fricative



شکل ۵-۱۲ نمودار زمانی و اسپکتروگرام اصوات سایشی صدا دار /v/ و /z/

۵-۲-۷ - مکث‌های بیصدا و صدا دار⁴⁶

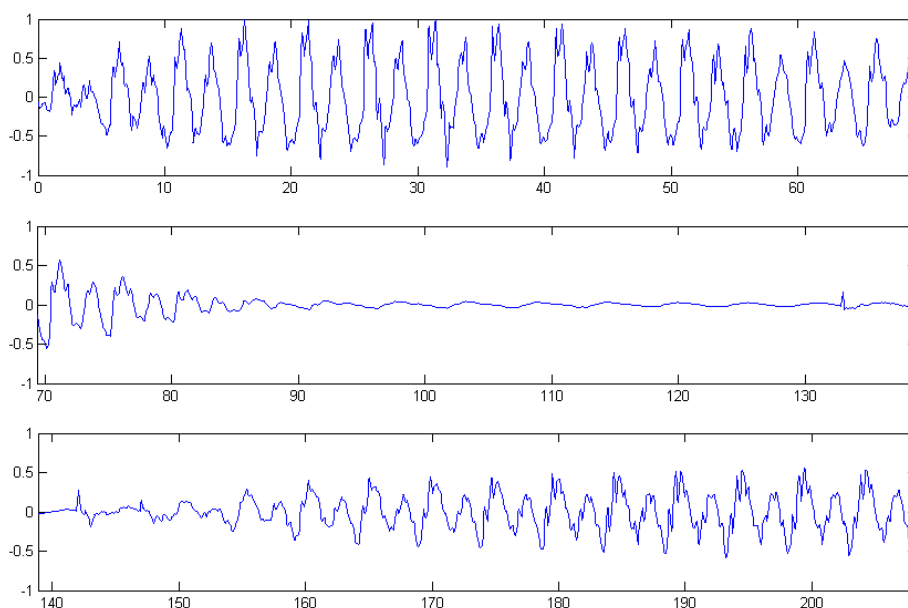
اصوات صامت مکثی صدا دار /b/، /d/ و /g/ گذرا می‌باشند، این اصوات با افزایش فشار در حفره دهانی و سپس آزاد نمودن با شتاب این فشار تولید می‌گردند. برای /b/ انقباض در لبها صورت می‌گیرد؛ برای /d/ انقباض در پشت دندانها و برای /g/ نزدیک لبهاست.

با توجه به اینکه اصوات مکثی دارای ماهیتی متغیر با زمان می‌باشند، ویژگیهای آنها بسیار تاثیرپذیر از آوایی است که بدنبال اصوات صامت مکثی می‌آیند. همچنین، شکل موج اصوات صامت مکثی اطلاعات

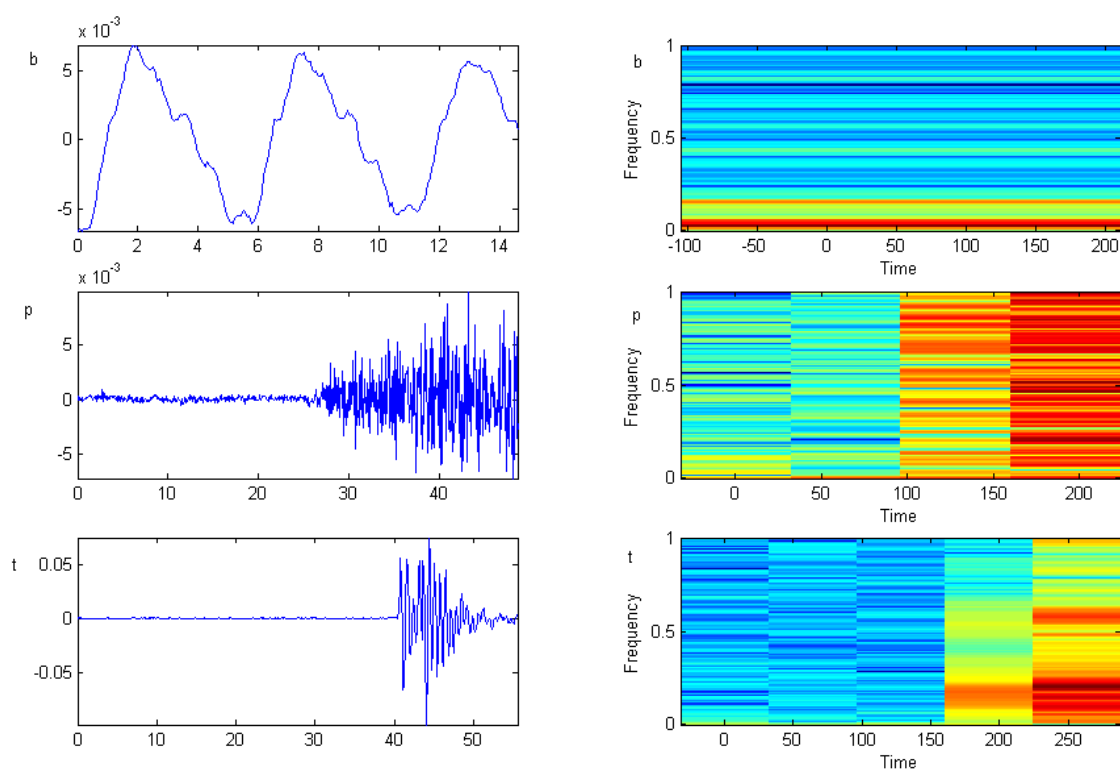
⁴⁶ .voiced and unvoiced stops

کمی در رابطه با این اصوات در اختیار ما قرار می‌دهد. شکل ۵-۱۳ شکل موجهای هجاهای /i-b-u/ را نشان می‌دهد. شکل موج /b/ ویژگیهایی با وجه تمایز نسبتا کمی دارد.

مکثهای بیصدای /p/، /t/ و /k/ نقطه مقابل مکثهای صدادار /b/، /d/ و /g/ می‌باشند. شکل ۵-۱۴ شکل موجهای اسپکتروگرام صوت مکثی صدا دار /b/ و صوت مکثی بیصدای /p/ و /t/ را نشان می‌دهد. ناحیه مکث⁴⁷ یا بازه زمانی که در طی آن فشار حبس شده است به آسانی در این شکل قابل مشاهده است.



شکل ۵-۱۳ نمودار زمانی شکل موج رشته ادا شده /i-b-u/

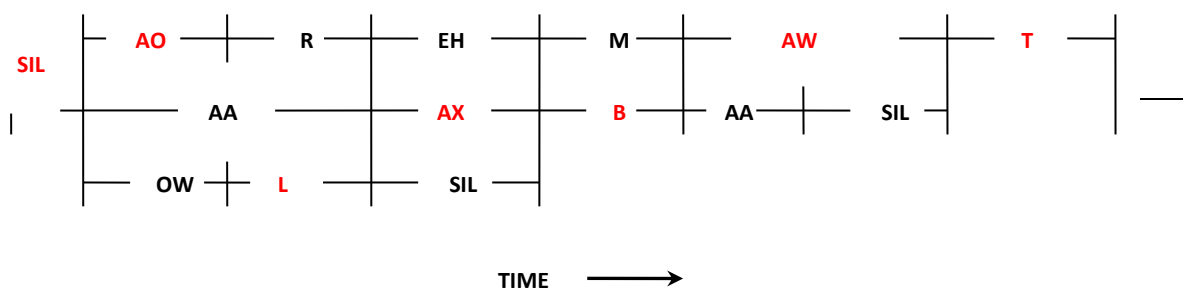


شکل ۵-۱۴ مقایسه نمودار زمانی و اسپکتروگرام اصوات مکثی صدا دار و بیصدای /b/، /p/ و /t/

۵-۳- روشهای تشخیص اتوماتیک گفتار با استفاده از ماشین

موارد ذکر شده در قسمت قبل، باعث ایجاد یک روش مناسب و کارآمد برای اجرا و راه‌اندازی تشخیص گفتار با استفاده از ماشین می‌گردد که بوسیله آن ماشین سعی در کد گذاری سیگنالهای صوتی بر اساس ویژگیهای صوتی مشاهده شده سیگنال و روابط مشخصی که بین ویژگیهای صوتی و نمادهای مصوت⁴⁸ برقرار است، دارد. بر روی این روش که آنرا صوت شناسی⁴⁹ می‌نامند، در طی چهل سال گذشته تحقیقات وسیعی انجام شده است. بهر حال، به دلایل زیادی، روش واج شناسی شنیداری نتوانست نتایجی مشابه با روشهای دیگر در سیستمهای واقعی بدست آورد. بنابراین، در این بخش، قصد داریم تا در مورد یک سری از روشهای ارائه شده در سیستمهای اتوماتیک تشخیص گفتار با استفاده از ماشین، با هدف یادگیری مطالب اساسی و نقاط ضعف و قوت هر روش، صحبت کنیم.

با توجه به مطالب ذکر شده در بالا، سه روش برای شناسایی گفتار وجود دارد که عبارتند از:



شکل ۵-۱۵ شبکه واجی برای یک سری کلمه

۱- روش واج شناسی شنیداری

۲- روش تشخیص الگو

۳- روش هوش مصنوعی

روش واج شناسی شنیداری بر اساس تئوری واجهای صوتی است که لازمه آن محدود و متمایز بودن واجها در زبان گفتگو است و دیگر اینکه بتوان هر واج را بوسیله مجموعه‌ای از خصوصیات که تفسیر کننده سیگنال صوتی، یا طیف فرکانسی آن، بر حسب زمان باشد، نشان داد. حتی اگر ویژگیهای صوتی واجها، با گوینده‌ها و واجهای نزدیک بهم (اصطلاحاً آنرا طرز گفتار گویند) تفاوت زیادی داشته باشند، فرض بر این است که قواعد حاکم بر تغییرات مفید بوده و می‌تواند بسهولت یاد گرفته شود و در شرایط واقعی بکار رود. بنابراین، اولین قدم در روش واج شناسی شنیداری برای تشخیص گفتار، مرحله جداسازی و برچسب زنی نامیده می‌شود، با انجام این مرحله، تشخیص گفتار به سیگنالهای صوتی گسسته (در زمان) تقسیم می‌شود تا بدین وسیله ویژگیهای صوتی هر سیگنال، بیانگر یک (یا حدالمقدور چندین) واج (یا کلاس) باشد. سپس براساس ویژگیهای صوتی، یک یا چند برچسب واج به هر ناحیه جدا شده نسبت داده شود. برای آنکه تشخیص گفتار عملی شود، مرحله دوم نیاز است. مرحله دوم سعی دارد تا کلمه (یا رشته‌ای از کلمات) معتبر را از رشته برچسبهای واج تولید شده در مرحله اول مشخص نماید.

برای شرح و توضیح مراحل روش واج شناسی شنیداری جهت تشخیص گفتار، شبکه واجهای شکل ۵-۱۵ را در نظر بگیرید. (شبکه واج نتیجه مرحله جداسازی و برچسب زنی پروسه‌های شناسایی است و مجموعه‌ای از واجها را نشان می‌دهد که بیشترین شباهت را با اصوات ورودی ادا شده دارند). مشکل کد

گزارى مجموعه واجها به رشته کلمه (یک یا چند کلمه) است، بگونه‌ایکه هر بازه زمانى در برگيرنده یکى از واجهاى مجموعه گردد و همچنين هر کلمه (یا رشته کلمه) بر اساس قواعد گرامرى انگليسى، موجود باشد. (نماد SIL نشان دهنده سکوت یا وقفه بين اصوات یا کلمات است؛ موقعيت عمودى مجموعه، در هر زمان، اندازه‌گیرى بهترين تطبيق صوتى با واجها است، با بالاترين مقدار بيشترين تطبيق را داريم). با تحقيق اندک مى‌توان شبکه واجهاى SIL-AO-L-AX-B-AW-T را معادل با رشته کلمه "all about" در نظر گرفت. واجهاى B,AX و L انتخابهاى دوم یا سوم رشته و بقيه واجها جزء اولين انتخابها مى‌باشند. اين مثال ساده بخوبى سختى کد گزارى واجها به رشته کلمات را نشان مى‌دهد. اين مرحله را اصطلاحاً دست‌يابی واژه‌ای⁵⁰ مى‌نامند. بطور قابل ملاحظه‌اى، همانطور که در بخش بعدى خواهيدديد، مشکل اصلى روش واج شناسى شنيدارى براى تشخيص گفتار، يافتن یک رشته واج معتبر براى مرحله دست‌يابی واژه‌اى است.

با بکارگیرى روش تشخيص الگو براى تشخيص گفتار، الگوهاى گفتار مستقيماً و بدون استخراج و جداسازى ويژگيهاى مشخص (در تشخيص صوت شناسى) مورد استفاده قرار مى‌گيرند. همانند بسيارى از روشهاى تشخيص الگو، اين روش شامل دو بخش است که بطور خلاصه، آموزش الگوهاى گفتار⁵¹ و شناسايى الگوها از طريق مقايسه الگو است. شناسايى گفتار توسط مرحله آموزش به سيستم انتقال داده مى‌شود. اين عبارت بدین معنى است که اگر حالتهاى مختلف الگوهاى گفتار (صوت، کلمه، عبارت و غيره)، که در مجموعه آموزشى و توسط الگوريتم تهيه مى‌شوند، در نظر گرفته شود، مرحله آموزش بايد قادر باشد تا بخوبى خصوصيات صوتى الگو (بدون توجه به اطلاعات هر یک از الگوهاى قرار گرفته در مرحله آموزش) را توصيف نمايد. اينگونه توصيف گفتار با استفاده از آموزش را طبقه‌بندي الگو⁵² مى‌نامند، زيرا ماشين ياد مى‌گيرد که کدامیک از ويژگيهاى صوتى کلاس گفتار در بين تمام داده‌هاى آموزشى معتبر و تکرار پذير است. مزيت اين روش مرحله مقايسه الگوهاست که مستقيماً گفتار مجهول (گفتارى که بايد شناسايى گردد) را با هر یک از الگوهاى ممکن که در مرحله آموزش به سيستم آموزش داده شده است، مقايسه نموده و گفتار مجهول بر اساس بهترين تطبيق ممکن، دسته بندي مى‌شود.

در اينجا بايد به اين نکته اشاره نمود که روش تشخيص الگو به سه دليل روش منتخب تشخيص گفتار است:

⁵⁰ .lexical access

⁵¹ . training of speech patterns

⁵² .pattern classification

۱- سادگی از لحاظ استفاده. این روش براحتی قابل فهم است، در ریاضیات و تئوری مخابرات پیشرفت زیادی نموده است و کاربرد بسیار زیادی دارد.

۲- خاصیت پایداری و نامتغیر بودن در مقابل واژه‌های گفتاری مختلف، کاربرها، مجموعه ویژگیها، الگویتماهای مقایسه الگو و قواعد تصمیم‌گیری. این خاصیت باعث کارایی الگوریتم برای محدوده وسیعی از واحدهای گفتار (محدوده‌ای شامل شبه واجها⁵³، جملات، عبارات)، واژگان کلمات، تعداد افراد گوینده، محیط‌های زمینه⁵⁴، شرایط انتقال و غیره می‌گردد.

۳- کارایی بسیار بالا. می‌توان نشان داد که بکارگیری روش تشخیص الگو در تشخیص گفتار کارایی بسیار بالایی را ایجاد می‌کند و مسیر مشخصی برای گسترش تکنولوژی در تمام جهات ایجاد می‌کند، بگونه‌ایکه کارایی، بدون در نظر گرفتن افزایش پیچیدگی مسئله، بیشتر و بیشتر می‌شود.

روش هوش مصنوعی بکار رفته در تشخیص گفتار ترکیبی از روش واج شناسی شنیداری و تشخیص الگو است. این روش سعی دارد تا پروسه تشخیص را بر اساس روشی که افراد از هوش خود در مشاهدات، آنالیز و در نهایت تصمیم‌گیری ویژگیهای صوتی اندازه‌گیری شده بهره می‌برند، مکانیزه کند.

۵-۳-۱- روش واج شناسی شنیداری تشخیص گفتار

شکل ۵-۱۶ بلوک دیاگرام روش واج شناسی شنیداری را که در تشخیص صوت بکار می‌رود را نشان می‌دهد. اولین قدم (که معمولا در تمام سیستمهای تشخیص صوت بکار می‌رود) تحلیل سیستم صوتی است (که انرا اصطلاحا اندازه‌گیری ویژگی⁵⁵ می‌نامند)، که توصیف مناسبی از مشخصه‌های صوتی سیگنال نسبت به تغییرات زمان ایجاد می‌کند. روش مرسوم که برای اجرای آنالیز طیفی استفاده می‌گردد، مجموعه‌ای از بانکهای فیلتر⁵⁶ و روشهای کد گذاری محاسبات خطی⁵⁷ است. بطور کلی، این روشها توصیف طیفی مناسبی از تغییرات سیگنال نسبت به زمان ایجاد می‌کنند.

⁵³ . phonemelike

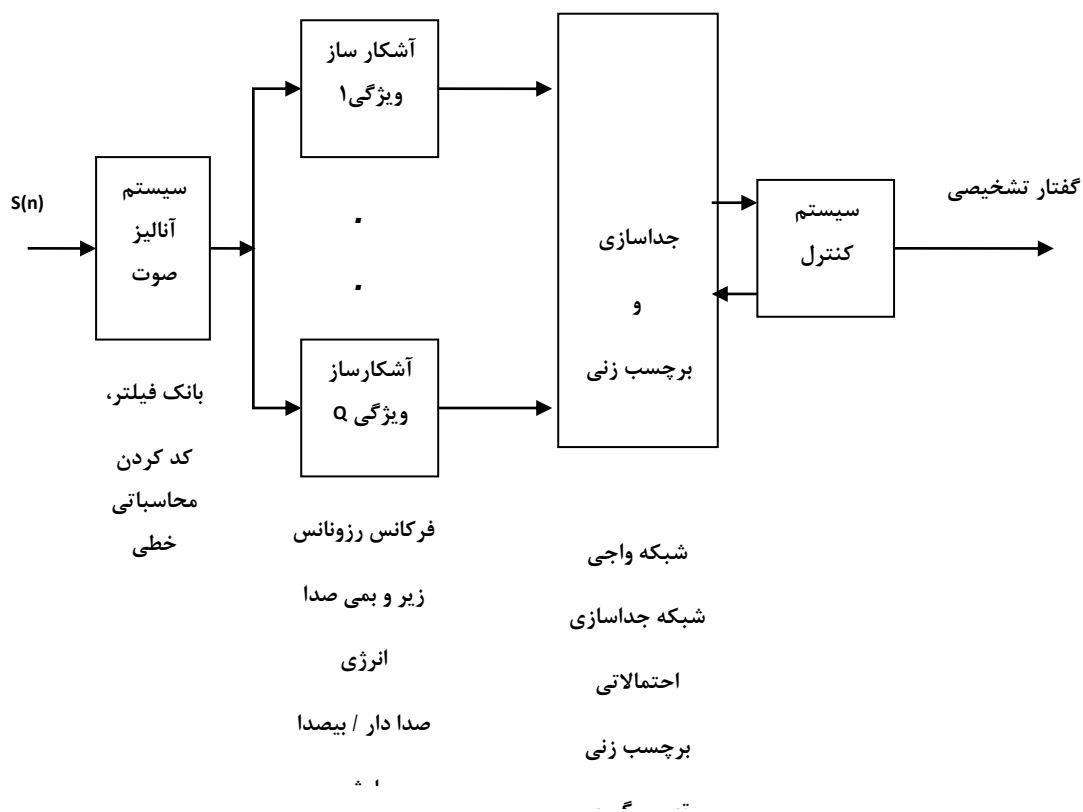
⁵⁴ . background environments

⁵⁵ . feature measurement method

⁵⁶ . filter banks

⁵⁷ . Linear Predictive Coding (LPC)

قدم یا مرحله بعدی تحلیل، مرحله شناسایی ویژگی⁵⁸ است. این مرحله اندازه‌گیریهای طیفی را به مجموعه‌ای از ویژگیها که توصیف کننده مشخصه‌های صوتی واجهای مختلف می‌باشند، تبدیل می‌نماید. یکی سری از این ویژگیها عبارتند از : وجود یا عدم وجود رزونانسهای بینی، تحریکهای تصادفی در صوت، مکان سه فرکانس اول فرکانسهای رزونانس، دسته بندی صدا دار و بیصدا (تحریک پرئودیک یا غیر پرئودیک) و نسبتهای انرژی فرکانسی بالا و پایین. بسیاری از ویژگیهای ارائه شده ذاتا باینری می‌باشند و بقیه پیوسته‌اند. مرحله تشخیص ویژگی معمولا شامل مجموعه‌ای از آشکارسازها



⁵⁸ . feature detection

شکل ۵-۱۶ بلوک دیاگرام سیستم تشخیص صوت با استفاده از روش واج شناسی شنیداری

است که بصورت موازی عمل می‌کنند و توسط تحلیل‌های مناسب و منطقی، در مورد وجود یا عدم وجود ویژگی تصمیم‌گیری می‌کند.

سومین مرحله تحلیل، مرحله جداسازی و برچسب زنی است که توسط آن سیستم سعی در یافتن نواحی پایدار می‌نماید (که تغییرات ویژگیها در آنجا خیلی کم باشد) و سپس بر اساس بیشترین تطبیق صورت گرفته بین نواحی پایدار و واجهای موجود، برچسب زنی انجام می‌گیرد. این مرحله قلب تشخیص دهنده واج شناسی شنیداری است و داشتن درک و فهم صحیحی از آن، بسیار دشوار است. بهمین علت از پارامترهای کنترلی بسیاری برای محدود کردن محدوده نقاط تصمیم‌گیری و حالت‌های برچسب زنی استفاده می‌گردد. بعنوان نمونه، برای یک سیستم تشخیص کلمه، با محدودیت‌های حداقل دو واج و حداکثر شش واج، بدین معنی است که مرحله کنترل باید شامل روشهایی با نقاط جداسازی بین ۱ تا ۵ باشد. این محدودیتها بسیاری از اوقات می‌تواند باعث افزایش کارایی سیستم گردد و فضای جستجو را کاهش دهد.

نتیجه جداسازی و برچسب زنی، معمولا مجموعه‌ای از واجهاست. کیفیت تطبیق ویژگیها با واجها می‌تواند با تعلق احتمال به هر یک از برچسبها صورت پذیرد. نتیجه نهایی تشخیص دهنده، یک کلمه یا شبکه‌ای از کلمات (مشابه شکل ۵-۱۵) است که بیشترین تطبیق را دارا می‌باشد.

۵-۳-۱-۱- دسته کننده آوایی واجهای صوتی

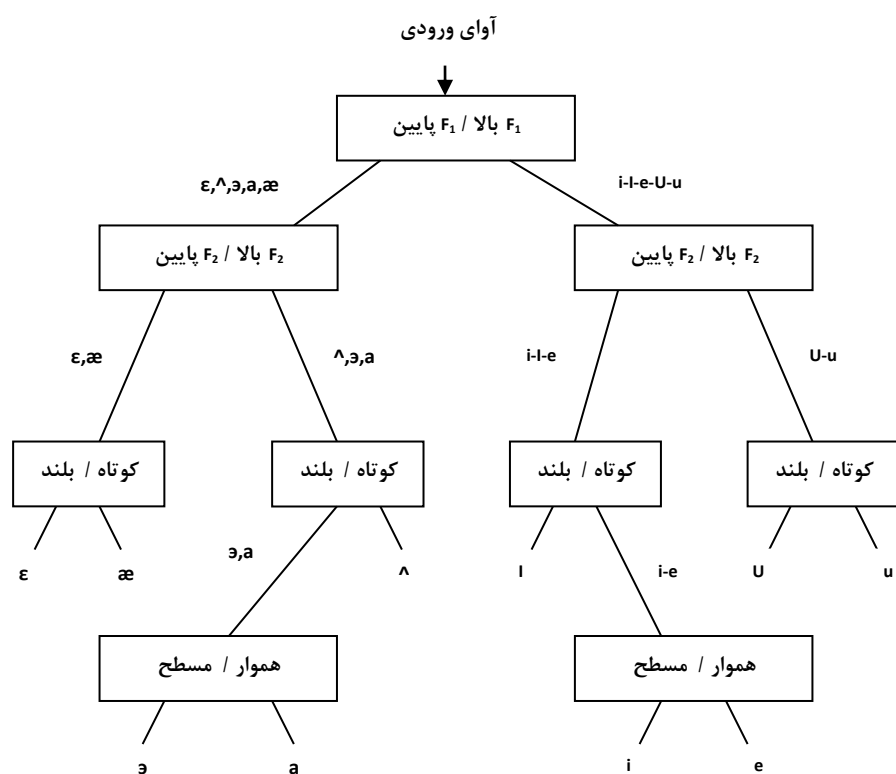
برای مشخص نمودن پروسه برچسب زنی در یک بخش طبقه بندی شده مانند آواها، فلوجارت شکل ۵-۱۷ را در نظر بگیرید. فرض کرده‌ایم که با سه ویژگی می‌توان تمام بخشها را از یکدیگر تشخیص داد که عبارتند از: فرکانس رزونانس اول، F_1 ، فرکانس رزونانس دوم، F_2 ، و طول بخش، D . مجموعه‌ای از آوای یکنواخت⁵⁹ (به استثنای اصوات ترکیبی) را در نظر بگیرید. برای طبقه بندی آوا از میان ۱۰ آوای یکنواخت، چندین آزمایش برای جداسازی گروهی آواها صورت می‌گیرد. همانگونه که در شکل ۵-۱۷ نشان داده شده

⁵⁹ . steady vowels

است، اولین آزمایش آواها را به فرکانسهای F_1 پایین (که آنها را آواهای پخشی⁶⁰ می‌نامند که شامل /ɛ/,/æ/,/ə/,/ʌ/,/a/ می‌شوند) به F_1 بالا (که آنها را آواهای فشرده⁶¹ می‌نامند و شامل /i/,/ɪ/,/U/,/u/,/ə/ می‌گردد) تقسیم می‌کند. هر یک از این مجموعه‌ها می‌تواند بر اساس اندازه‌گیریهای F_2 به زیر مجموعه‌های کوچکتر تقسیم گردد، که با انجام اینکار، آواها به دو بخش آواهایی که F_2 بالایی دارند و آواهایی که F_2 پایینی دارند، تقسیم‌بندی می‌شوند. سومین آزمایش بر اساس طول بخش است، که آواهای کشیده⁶² (دارای D بالا) را از آواهای کوتاه⁶³ (دارای D کم) جدا می‌سازد. در نهایت، با انجام یک آزمایش مناسب بر روی مقادیر فرکانسهای رزونانس، آواهای جدا نشده باقیمانده را جدا می‌سازیم، در اینحالت آواها به آواهای هموار⁶⁴ (که F_1+F_2 آنها از سطح آستانه بیشتر شود) و آواهای مسطح⁶⁵ (که F_1+F_2 آنها از سطح آستانه کمتر شود) تقسیم می‌شوند.

واضح است که چندین سطح آستانه را می‌توان در داخل الگوریتم دسته بندی وارد نمود. چنین آستانه‌هایی می‌تواند دقت نهایی سیستم تشخیص را تحت تاثیر قرار دهند.

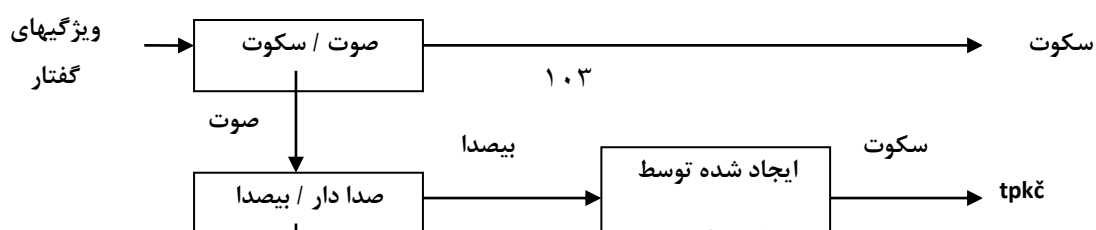
60 . diffuse vowels
 61 . compact vowels
 62 . tense vowels
 63 . lax vowels
 64 . flat vowels
 65 . plain vowels



شکل ۵-۱۷ دسته بندی آوای صوت شناسی

۵-۳-۱-۲- دسته بندی اصوات گفتار

طبقه بندی آواها فقط قسمت کوچکی از شیوه برچسب زنی تشخیص دهنده واج شناسی شنیداری است. در عمل، ما بدنبال روشی برای دسته بندی اصوات به یکی (یا بیشتر) از ۴۰ واج بحث شده قبلی هستیم. قبل از اینکه درباره چگونگی حل این مسئله صحبت کنیم، قدری مسئله را ساده تر در نظر بگیرید و فرض کنید می‌خواهیم تعلق آنها را به یکی از چند کلاس صوتی موجود بعنوان نمونه مکتهای بیصدا، مکتهای صدا دار و اصوات سایشی بیصدا بدست آوریم. در اینجا نیز هیچ روش کلی و ساده‌ای برای انجام اینکار وجود ندارد، با این وجود، در شکل ۵-۱۸ یک روش ساده و مفید برای انجام چنین کاری آورده شده است.



شکل ۵-۱۸ دسته بندی اصوات گفتار بوسیله درخت دودویی

در این روش از درخت دودویی⁶⁶ برای تصمیم گیری استفاده شده است. اولین تصمیم گیری جدا نمودن صوت/سکوت⁶⁷ است که در آن ویژگیهای صوت (در این مثال انرژی) با سطوح آستانه مقایسه می گردد، و در صورتیکه مقدار ویژگی کمتر از سطح آستانه باشد، سکوت می شود. دومین تصمیم گیری تصمیم گیری صدا دار/ بیصدا است (بر اساس پریودیک بودن قسمتهاست) که در آن اصوات صدا دار از اصوات بیصدا جدا می شود. آزمایشی برای صداهای صامت مکثی بیصدا انجام می گیرد و با اینکار حروف مکثی بیصدا (/t/, /p/, /k/, /ç/ از حروف سایشی بیصدا (/f/, /θ/, /s/, /š/) جدا می شود. آزمایش (انرژی) فرکانس پایین/ فرکانس

⁶⁶ . binary tree

⁶⁷ . sound/silence

بالا، حروف سایشی صدا دار (/v/, /ð/, /z/, /ž/) را از اصوات صدا دار جدا می‌سازد. مکثهای صدا دار با چک کردن اینکه صوت سکوت (یا شبه سکوت) است، انجام می‌گیرد. در نهایت، آزمایش طیفی آوا/ پر صدا⁶⁸ (یافتن فضای خالی طیفی) آواها را از اصوات پر صدا (حروف صامت دماغی و /w/, /l/, /r/ و /y/) جدا می‌سازد. طبقه‌بندی کننده آوای شکل ۵- ۱۷ را می‌توان برای تشخیص آوا استفاده نمود.

آزمایشات نشان داده شده در شکل ۵- ۱۸ نسبتاً خام بوده و بهمین علت خطای بالایی دارند. بعنوان نمونه، بعضی از اصوات صامت صدا دار بعنوان سکوت یا شبه سکوت شناخته نمی‌شوند. مشکل دیگر این است که هیچ روشی برای جداسازی اصوات ترکیبی از آواها وجود ندارد.

۵-۳-۱-۳-۵- مثالهای برجسب زنی واج شناسی صوتی

برای روشنتر شدن بعضی از مشکلات روش واج شناسی شنیداری در تشخیص صوت، مثال زیر را در نظر بگیرید. در این مثال (گرفته شده از یک تشخیص دهنده واج شناسی شنیداری واقعی) می‌بینیم که واجهای اضافه شده (Y در "MAY" و AH در "BY")، واجهای حذف شده (N در "EARN" و N در "MONEY") و جایگزینیهای واجی (Z بجای K در "WORKING" و N بجای NG در "WORKING") وجود دارد. پیچیدگی کد گزاری درست و صحیح واجها در داخل جملات و کلمات بصورت چشمگیری با افزایش نرخ واجهای اضافه شده، حذف شده و جایگزین شده افزایش می‌یابد.

Phonemes: /sil/ -/j/-/e/-/n/ /m/-/e/-/y/ /ʒ/-/m/-/ə/-/r/ /m/-/ʌ/-/sil/-/e/

Arpabet: SIL -JH-EY-N + M-EY-Y + ER-M-AO-R + M-AH-SIL-EY

Words: JANE MAY EARN MORE MONEY

Phonemes: /b/-/aʰ/-/ə/ /w/-/ʒ/-/sil/-/j/-/l/-/n/ /h/-/a/-/r/-/sil/-/d/

Arpabet: B-AY-AX + W-ER-SIL-J-IH-N + HH-AA-R-SIL-D

Words: BY WORKING HARD

⁶⁸ . vowel/sonorant test

دو مثال دیگر جداسازی و برچسب زنی واج شناسی شنیداریدر شکل‌های ۵-۱۹ و ۵-۲۰ آورده شده است. در این شکل‌ها محدوده انرژی سیگنال صوت، طبقه بندی صدا دار- بیصدا- سکوت در طول زمان، نقاط جداسازی و شبکه‌های واجی نشان داده شده است. رمزگشای مناسب شبکه برای کلمات ادا شده بصورت واجهایی ضمیمه شده به خطوط کاملاً پر نشان داده شده‌اند. برای مثال در شکل ۵-۱۹ (رشته حرفی "seven-six")، می‌بینیم که خطای سیستم نسبتاً زیاد است. برای رمزگشایی به اعداد، خطای بین کلاسها معمولاً کم و قابل صرف‌نظر است.

برای مثال، در شکل ۵-۲۰ (رشته کلمه "did you")، فقط رمز گشایی به واجهای اولین مرحله، مشکل است، چونکه اصوات صوتی پایه کلمات "did" و "you" در متن از D-IH-D-Y-UW به D-IH-J-UH تغییر یافته است.

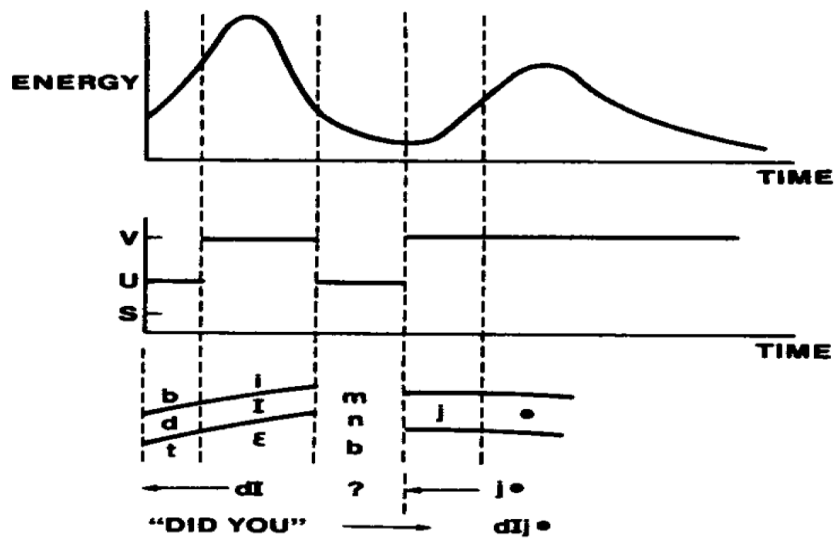
۵-۳-۱-۴- مشکلات روش واج شناسی شنیداری

مشکلات زیادی در رابطه با استفاده از روش واج شناسی شنیداری در تشخیص گفتار وجود دارد. این مسایل، در بسیاری روشها، بعنوان علل عدم موفقیت یک سیستم تشخیص گفتار شناخته می‌شوند. این عوامل عبارتند از:

- ۱- روش نیازمند دانش گسترده‌ای در رابطه با ویژگیهای صوتی واجها است.
- ۲- انتخاب ویژگیها بر اساس غیر قابل تعمیم بودن آن است. برای بسیاری از سیستمها انتخاب ویژگیها بر اساس شواهد است.



شکل ۵-۱۹ جداسازی و برچسب زنی جمله "seven-six"



شکل ۵- ۲۰ جداسازی و برچسب زنی جمله "did you".

۳- طراحی طبقه‌بندی کننده گفتار عملی نیست. روش‌های تک منظوره معمولاً برای محدود نمودن درخت‌های تصمیم‌گیری باینری بکار می‌روند. اما اخیراً، روش‌های طبقه‌بندی و برگشتی درختی برای تصمیم‌گیری بکار می‌روند که دارای کارایی بالایی نیز هستند.

۴- روش مشخصی برای پروسه اتوماتیک نمودن و تنظیم پارامترهای کنترلی وجود ندارد. در واقع، حتی یک روش ایده‌آل برای برچسب زنی صوت‌های آموزشی در یک شبکه محدود وجود ندارد.

بعلت مشکلات فوق، روش واج شناسی شنیداری در تشخیص صوت نمی‌تواند بعنوان یک روش مفید و قابل تعمیم مورد بررسی قرار گیرد و باید بدنبال یک روش بهتر با درصد موفقیت بالاتر باشیم.

۵-۳-۲- روش تشخیص الگوی آماری برای تشخیص گفتار

بلوک دیاگرام روش تشخیص الگوی که در تشخیص گفتار بکار می‌رود، در شکل ۵- ۲۱ نشان داده شده است. مدل نشان داده شده دارای چهار قدم است که عبارتند از:

۱- اندازه‌گیری ویژگی، که در آن مجموعه‌ای از اندازه‌گیریها بعنوان سیگنال ورودی قرار داده می‌شوند، که آنرا الگوی آزمایشی⁶⁹ می‌نامند. برای سیگنال‌های گفتار، ویژگیها معمولاً خروجی بعضی از روش‌های تحلیل طیفی است، مانند آنالیزور بانک فیلتر، آنالیز رمز گشایی محاسبه خطی، یا آنالیز تبدیل فوریه گسسته.

۲- الگوهای آموزش، که در آن از یک یا چند الگوی آزمایشی اصوات گفتار یکسان، برای تولید الگوی نماینده ویژگیهای آن کلاس استفاده می‌شود. الگوی حاصل، که معمولاً آنرا الگوی مرجع⁷⁰ می‌نامند،

⁶⁹ . test pattern

⁷⁰ . reference pattern

می‌تواند یک نمونه یا الگوی بدست آمده از بعضی از روشهای میانگین‌گیری، یا مدلی که توصیف‌کننده ویژگیهای آماری الگوهای مرجع است، باشد.

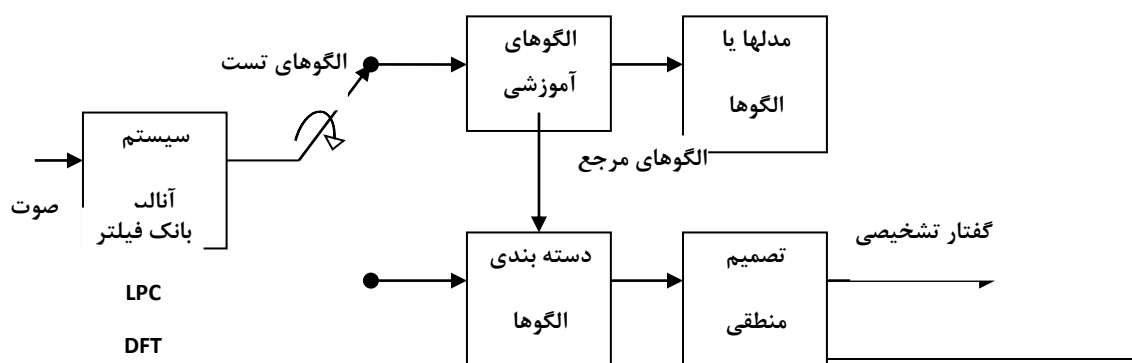
۳- طبقه‌بندی الگوها، که در آن الگوهای آزمایشی مجهول با هر یک از الگوهای مرجع مقایسه می‌گردند و شباهت بین الگوی تست با هر یک از الگوهای مرجع اندازه‌گیری می‌گردد. برای مقایسه الگوهای گفتار (که شامل مجموعه‌ای از بردارهای طیفی است)، ما نیاز به اندازه‌گیری فاصله محلی، فاصله طیفی بین دو بردار طیفی، و پروسه تطبیق زمانی (که آنرا الگوریتم انتقال زمانی دینامیکی⁷¹ می‌نامند) داریم، که دومی برای تصحیح سرعت‌های مختلف صحبت (مقیاس‌های زمانی) دو الگو بکار می‌رود.

۴- تصمیم منطقی، که در آن با استفاده از میزان شباهت الگوی تست با هر یک از الگوهای مرجع، بهترین کلاس برای الگوی آزمایشی انتخاب می‌گردد.
نقاط قوت و ضعف مدل تشخیص الگو عبارتند از:

۱- کارایی سیستم وابسته به میزان داده‌های آموزشی است که برای تولید الگوهای مرجع کلاس صوت استفاده می‌گردند؛ معمولاً، هر چقدر میزان داده‌های آموزشی سیستم زیاد باشد کارایی سیستم بیشتر خواهد شد.

۲- الگوهای مرجع به مشخصه‌های انتقال و محیط صحبت حساس هستند؛ این امر بعثت اثری است که نویز زمینه و انتقال بر روی مشخصه‌های طیفی صوت دارد.

۳- هیچگونه دانش مشخص و معینی در سیستم بکار نمی‌رود؛ از اینرو، روش فوق به انتخاب کلمات واژگان، عبارات و کار معنایی غیر حساس است.



⁷¹ . Dynamic Time Warping (DTW)

شکل ۵-۲۱ بلوک دیاگرام تشخیص دهنده گفتار با استفاده از روش تشخیص الگو

- ۴- میزان محاسبات صورت گرفته برای آموزش سیستم و طبقه بندی الگوها معمولاً متناسب با تعداد الگوهای آموزش داده شده یا تشخیص داده شده است؛ بنابراین، سعی بر این است که از تعداد نمونه های زیادی در آموزش سیستم استفاده نگردهد.
- ۵- با توجه به اینکه سیستم نسبت به کلاسهای صوتی غیر حساس است، روش کلی مورد استفاده برای محدوده وسیعی از اصوات صوتی، شامل اصطلاحات، تمام کلمات و واحدهای کوچکتر از کلمه قابل استفاده است. بنابراین خواهیم دید که چگونه مجموعه ای از دستورات ساده و تکنیک های گسترش یافته برای یک کلاس صوتی (بعنوان نمونه کلمه) می تواند بطور مستقیم و بدون انجام اصلاحات در الگوریتم، در کلاسهای صوتی مختلف (بعنوان نمونه واحدهای کوچکتر از کلمه) بکار رود.
- ۶- وارد نمودن محدودیتهای قواعد صرفی و نحوی (و حتی معنایی) در داخل ساختار تشخیص الگو امری مهم بنظر می رسد، چون با انجام اینکار دقت تشخیص افزایش و پیچیدگی محاسباتی الگوریتم کاهش می یابد.

۵-۳-۳- روشهای هوش مصنوعی برای تشخیص گفتار

ایده اولیه استفاده از روش هوش مصنوعی در تشخیص صوت، جمع آوری و ثبت اطلاعات از منابع اطلاعاتی گوناگون و تاثیر آن در حل مسئله است. بنابراین، بعنوان مثال، ممکن است روش هوش مصنوعی برای کامل نمودن پروسه جداسازی و برچسب زنی، از اطلاعات صوتی، دانش واژگانی، دانش صرفی و نحوی و دانش معنایی بهره ببرد. برای روشنتر شدن مسئله، در ابتدا این دانشهای مختلف را تعریف می کنیم:

- دانش صوتی⁷² - بررسی اینکه کدام اصوات (که قبلا آنرا واج نامیدیم) در اندازه‌گیریهای طیفی گفته شده است و بررسی وجود یا عدم وجود ویژگیها.
- دانش لغوی⁷³ - ترکیب نمودن شواهد صوتی بگونه‌ای که اصوات را به کلمات (یا بطور معادل تجزیه کردن کلمات به اصوات) نگاشت دهد.
- دانش صرفی و نحوی⁷⁴ - ترکیب نمودن کلمات برای ایجاد جملات صحیح از لحاظ گرامری (بر اساس مدل‌های زبانی) مانند: جملات یا عبارات.
- دانش معنایی⁷⁵ - درک حوزه کاری بگونه‌ای که قادر به تصدیق جملاتی (یا عبارات) که سازگار با کارهای انجام شده یا سازگار با جملات رمزگشایی شده قبلی است، باشد.
- دانش عمل گرایانه⁷⁶ - توانایی استنتاج در حل نمودن معانی گنگ و نامفهوم بر اساس روشهایی که در آن کلمات استفاده می‌شوند، لازم است.

برای درک بهتر توانایی تصحیح این منابع آموزش جملات زیر را در نظر بگیرید:

- ۱- به یخچال برو و برایم کتاب بیاور.
- ۲- خرسها گوسفندان را نابود کردند.
- ۳- موتورهای برقی رنگ پریده کاملا قدیمی.
- ۴- خیلی اوقات ایده‌های خوب هنگامی بکار می‌افتند که کمترین چیز مد نظر ماست.

اولین جمله از لحاظ قواعد صرفی و نحوی درست است، اما از لحاظ معنایی غلط است. دومین جمله را می‌توان از لحاظ گرامری به دو صورت مختلف بیان نمود، بسته به اینکه مکان وقوع در حادثه جنگل یا توصیفی از مسابقه فوتبال بین دو تیم با نامهای خرسها و گوسفندان باشد. سومین جمله از لحاظ قواعد صرفی و نحوی غیر قابل قبول است و از لحاظ معنایی بی‌معنی است. چهارمین جمله از لحاظ معنایی

⁷² . acoustic knowledge

⁷³ . lexical knowledge

⁷⁴ . syntactic knowledge

⁷⁵ . semantic knowledge

⁷⁶ . pragmatic knowledge

ناهماهنگ است و می‌توان آنرا با یک جایگزینی ساده مثل تغییر کلمه "بکار افتادن" با "رخ دادن" تصحیح نمود.

چندین روش برای جمع آوری منابع اطلاعاتی در داخل یک تشخیص دهنده صوتی وجود دارد. شاید یکی از روشهای بسیار معروف پردازنده "پایین - بالا"⁷⁷ باشد (شکل ۵-۲۲)، که در آن پردازشهایی با مرتبه پایین (بعنوان نمونه شناسایی ویژگی، رمزگشایی صوتی) تقدم بیشتری نسبت به پردازشهای مرتبه بالا (مدل زبانی، رمز گشایی لغوی) در حالت ترتیبی دارند و پردازش هر طبقه را تا حد ممکن محدود می‌کنند. در روش دیگری که آنرا پردازنده "بالا- پایین"⁷⁸ می‌نامند، مدل زبانی کلمات فرضی تولید می‌کند که با سیگنالهای صوتی بیشترین تطبیق را داشته باشند و از لحاظ معنایی و قواعد صرفی و نحوی، جملات معناداری بر اساس میزان تطبیق کلمات ایجاد می‌گردد. شکل ۵-۲۳ سیستمی را نشان می‌دهد که در حالت بالا - پایین و با جمع آوری واحدهای تطبیق، رمزگشای لغوی و آنالیزکننده‌های قواعد صرفی و نحوی در داخل یک چهارچوب پایدار مدلسازی شده است. (برای توضیحات بیشتر می‌توانید به مرجع [۳] فصل شناسایی اصوات صوتی پیوسته با واژگان وسیع مراجعه نمایید).

سومین روش پرکاربرد روش تخته سیاه⁷⁹ نام دارد، که در شکل ۵-۲۴ نشان داده شده است. در این روش، تمام منابع اطلاعاتی مستقل از یکدیگر در نظر گرفته می‌شوند. هر منبع اطلاعاتی بر اساس وقوع الگوها در تخته سیاهی که نمونه‌ها را با استفاده از منبع اطلاعاتی مشخص می‌کند، داده گرا است.

۵-۳-۴ - شبکه‌های عصبی و کاربرد آنها در تشخیص گفتار

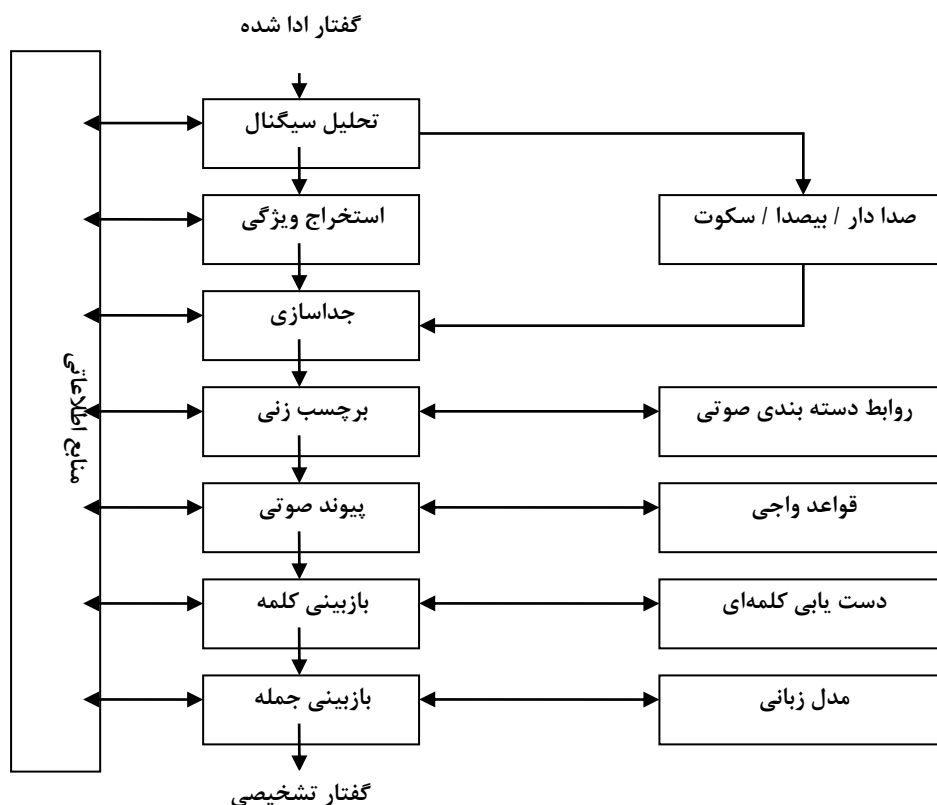
برای ایجاد یک سیستم تشخیص صوت با استفاده از هوش مصنوعی نیاز به منابع اطلاعاتی گوناگونی داریم. از اینرو، دو مبحث بسیار مهم در هوش مصنوعی، جمع‌آوری اطلاعات بصورت خودکار (یادگیری) و تطبیق‌دهی آن است. روشی که در آن این مباحث اجرا شده است، روش شبکه‌های عصبی است. در این بخش، در مورد اینکه چرا شبکه‌های عصبی بوجود آمد و دیگر اینکه چطور می‌توان از این روش در تشخیص گفتار استفاده نمود، صحبت می‌کنیم. شکل ۵-۲۴ بلوک دیاگرام یک سیستم ادراکی صوتی که بر اساس مدل ادراکی

77 . bottom-up

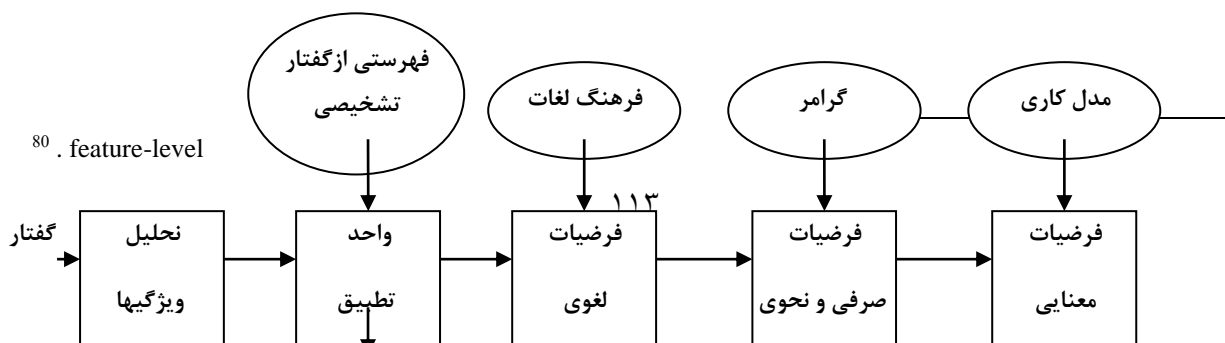
78 . top-down

79 . black board

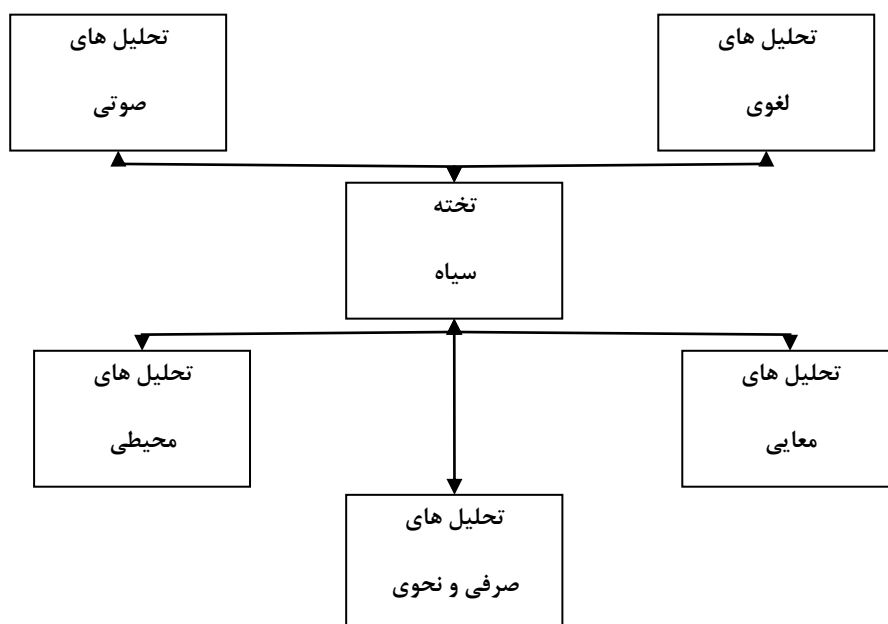
انسان است، را نشان می‌دهد. سیگنال صوتی ورودی بوسیله مدل گوش انسان آنالیز می‌گردد، که در آن اطلاعات طیفی یک سیگنال محاسبه می‌شود و در یک بانک اطلاعاتی ذخیره می‌گردد. اطلاعات حسی دیگر (نظیر بینایی و لامسه) در بانک اطلاعاتی قرار می‌گیرد و برای تولید چندین توصیف "مرتبه - ویژگی" ⁸⁰ استفاده می‌گردد. در نهایت بعد از چند طبقه تشخیص دهنده ویژگی، خروجی نهایی سیستم، تفسیر کننده اطلاعات ورودی صوتی است.



شکل ۵- ۲۲ روش پایین- بالا برای جمع آوری اطلاعات سیستم تشخیص گفتار



شکل ۵- ۲۳ روش بالا- پایین برای جمع آوری اطلاعات سیستم تشخیص گفتار



شکل ۵-۲۳ روش تخته سیاه برای جمع آوری اطلاعات سیستم تشخیص صوت [۶].

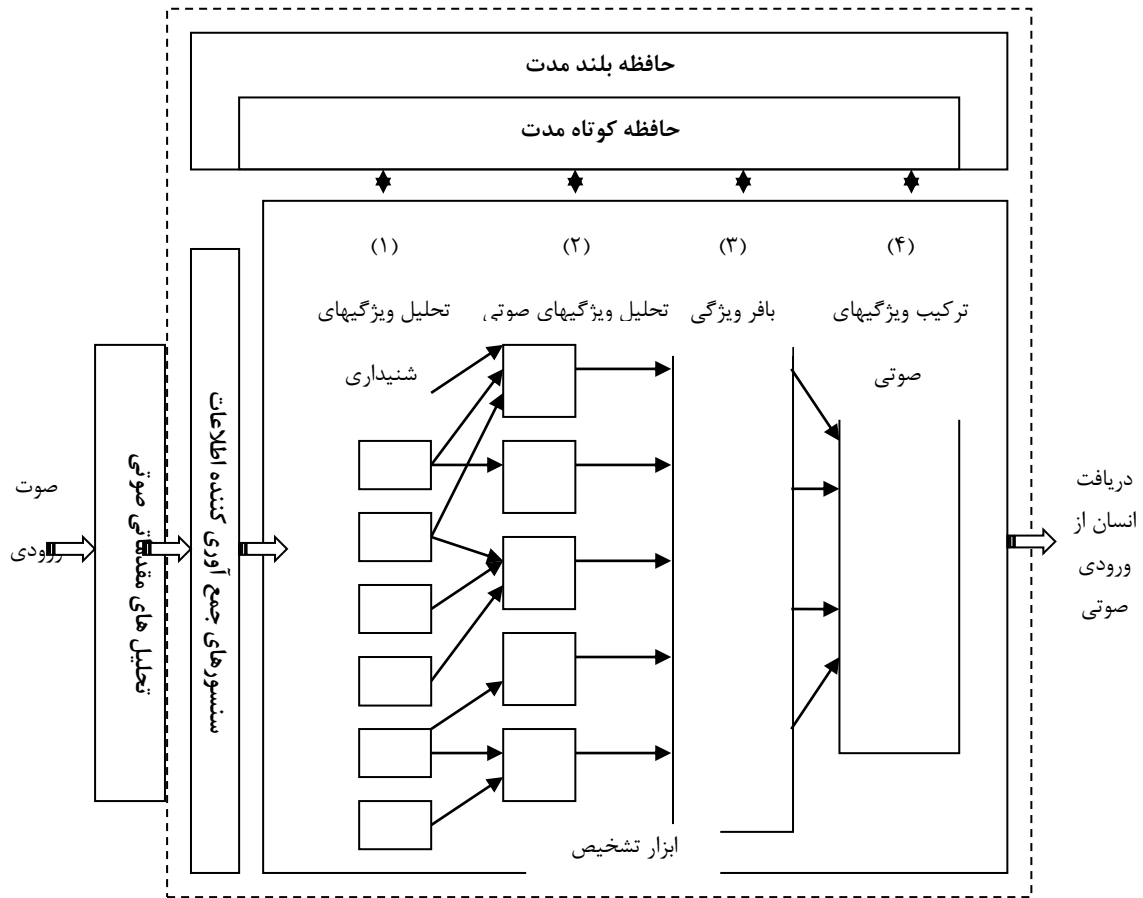
سیستم شکل ۵-۲۴ معدلی از سیستم ادراک صوتی انسان است. تحلیلگرهای صوتی بر اساس اطلاعات بدست آمده از گوش انسان عمل می‌کند. تحلیلگرهای ویژگی مختلف نشان دهنده تحلیل‌ها با سطوح مختلف در مسیر عصبی به مغز می‌باشند. فرم کلی این مدل که شبکه پیوندی جلو برنده⁸¹ است، یک شبکه عصبی است.

شبکه‌های عصبی مرسوم بر اساس الگوهای استاتیکی سازمان دهی شده‌اند. همانطور که در این فصل بیان گردید، صوت ذاتا دارای ماهیتی دینامیکی است. بهمین علت نیاز است تا یک سری اصلاحات در مدل‌های شبکه عصبی که در سیستم‌های تشخیص صوت بکار می‌روند، صورت پذیرد. شاید یکی از ساده‌ترین انواع شبکه‌های عصبی که الگوهای صوتی دینامیکی را در خود جای داده‌اند، شبکه عصبی تاخیر زمانی⁸² باشد که در شکل ۵-۲۵ نشان داده شده است. این ساختار هر ورودی صوتی را به N فریم صوتی تبدیل می‌کند (بعنوان نمونه بردارهای طیفی که مدت زمان $N\Delta$ را تحت پوشش قرار می‌دهد که در آن Δ زمان جدا شده بین طیف صوتی کنار هم است). با باز نمودن ورودی به N فریم (که N از مرتبه ۱۵ است)، بسیاری از انواع آشکارسازهای واج شناسی شنیداری در این شبکه‌ها قابل استفاده خواهند بود. بعنوان نمونه، شکل ۵-۲۶ یک شبکه TDNN را با دو لایه مخفی که برای شناسایی /b/ از /d/ و از /g/ استفاده می‌شود، را نشان می‌دهد.

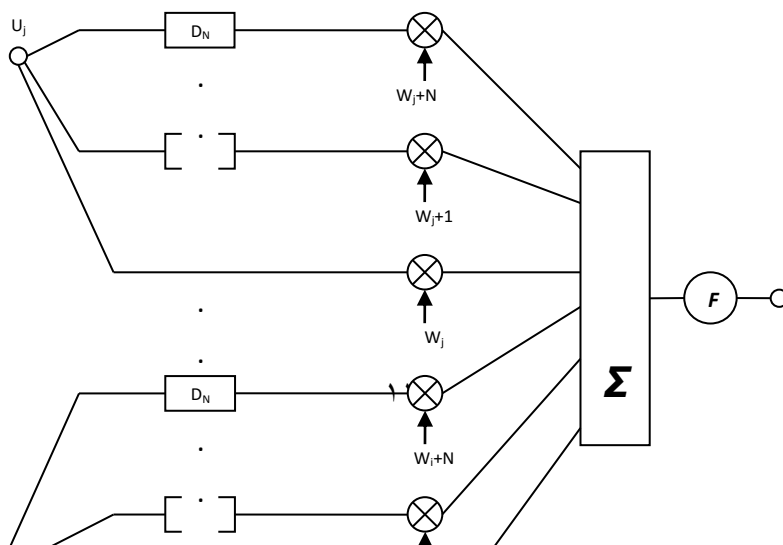
اختلاف کم ساختار شبکه‌های عصبی که برای تشخیص گفتار بکار می‌رود، ترکیب روش اجرای فیلتر تطبیقی با شبکه عصبی مرسوم برای محاسبه اصوات دینامیکی است، که در شکل ۵-۲۷ نشان داده شده است. دسته کننده الگو، بردار ویژگی صوتی بدست آمده را می‌گیرد و آنرا با فیلترهای تطبیقی کانالو می‌کند تا ویژگیهای صوتی و نتایج در طول زمان جمع و محاسبه گردد.

⁸¹ . feed forward connectionist network

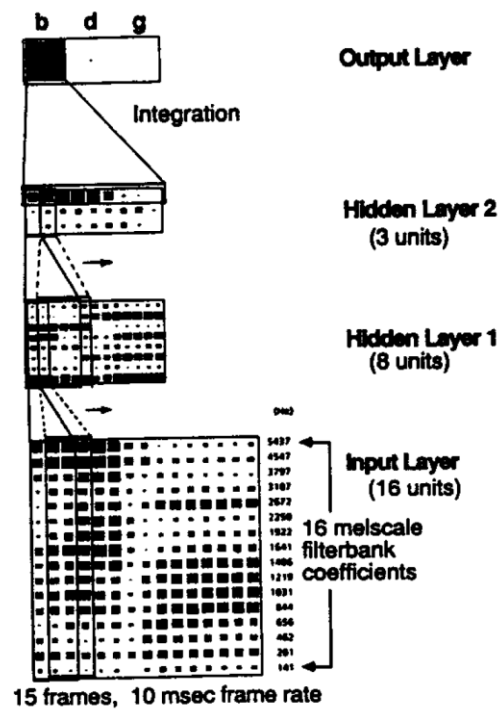
⁸² . Time Delay Neural Network (TDNN)



شکل ۵-۲۴ بلوک دیاگرام سیستم شنوایی گوش انسان



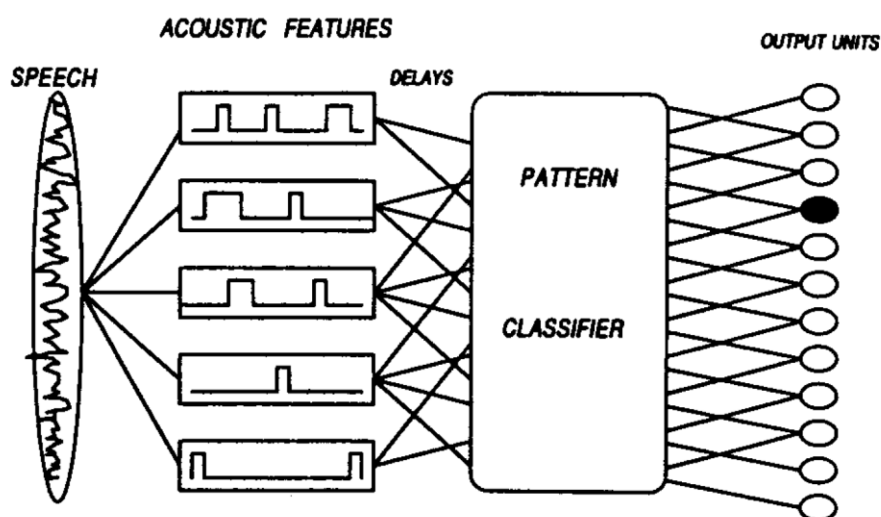
شکل ۵-۲۵ مولفه‌های محاسباتی شبکه عصبی تاخیر زمانی [۷].



شکل ۵-۲۶ ساختار داخلی TDNN برای تشخیص /g/ و /b/،/d/ [۷].

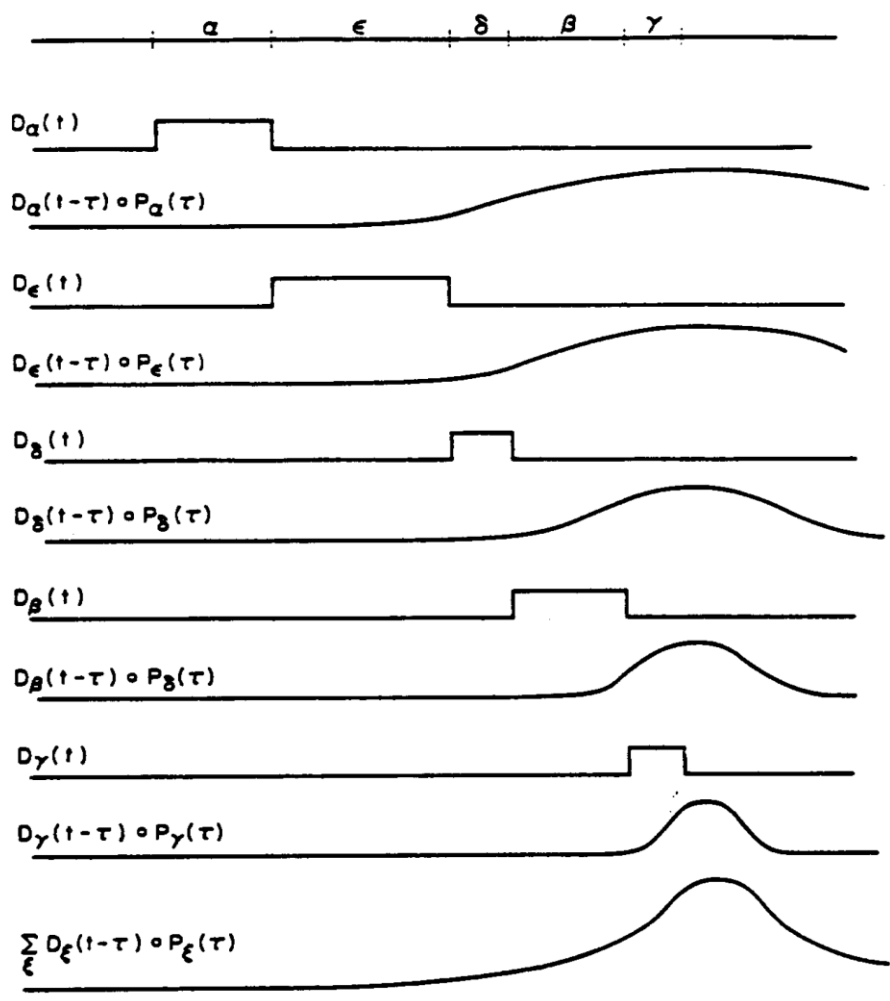
برای درک این مطلب که چطور شبکه شکل ۵-۲۷ برای تشخیص گفتار بکار می‌رود، همانطور که در شکل ۵-۲۸ نشان داده شده است، یک صوت تشخیص داده شده که دارای بردار ویژگی صوتی $(\alpha, \epsilon, \delta, \beta, \gamma)$ را در نظر بگیرید. فرض کنید که این صوت ورودی شبکه شکل ۵-۲۷ است و ورودی، همانند خط اول، شکل ۵-۲۸ است. هنگامیکه ویژگی صوتی α شناسایی گردید، تاخیر داده می‌شود و با فیلتر تطبیقی کانوالو می‌گردد. بهمین ترتیب ویژگیهای صوتی ϵ, δ, β و γ شناسایی می‌گردد. در نهایت، در انتهای دنباله، خروجیهای کانوالو شده جمع می‌گردد و بیشترین مقدار محاسبه می‌گردد، که نشان دهنده تشخیص سیستم به ازای سیگنال صوتی ورودی است.

در پایان، روش سوم جمع آوری اطلاعات زمانی در داخل یک شبکه عصبی در شکل ۵-۲۹ نشان داده شده است. این شبکه را شبکه عصبی کنترلی مخفی⁸³ می‌نامند و پارامتر متغیر با زمان، c ، بعنوان مکمل ورودی استاندارد، x ، امکان تغییر خصوصیات شبکه را با تغییرات زمان می‌دهد.

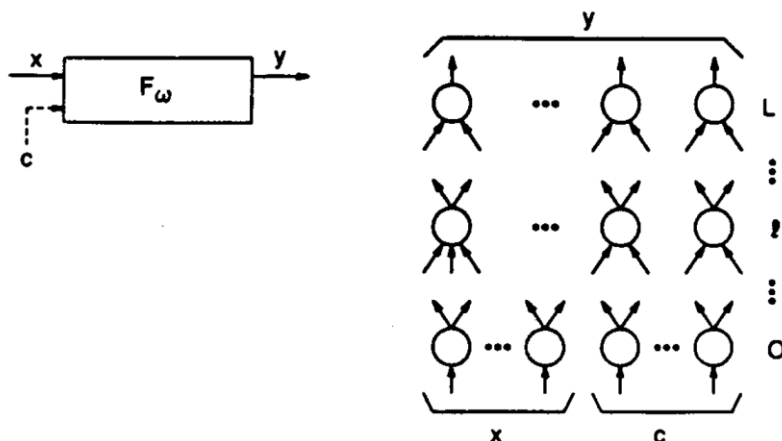


⁸³ . Hidden Control Neural Network (HCNN)

شکل ۵-۲۷ ترکیب شبکه عصبی و فیلتر تطبیقی برای تشخیص گفتار [۸].



شکل ۵-۲۸ مثالی برای درک بهتر ترکیبی از شبکه عصبی با مجموعه‌ای از فیلترهای تطبیقی



شکل ۵-۲۹ شبکه عصبی کنترل پنهانی [۹].

۵-۴- مراجع

- [1] L.R. Rabiner and R.W. Schafer, "Digital Processing of Speech Signals," Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- [2] Hector Perez Meana, "Advances in Audio and Speech Signal Processing," Chapter 11, pages 325-349, IDEA Group Publishing.
- [3] L.R. Rabiner and B. H. Jang, "Fundamental of Speech Recognition," chapter 2, pages 11-65. Published by PTR Prentice-Hall, Inc.
- [4] G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," J. Acoust. Soc. Am., Vol. 24, No. 2, pp. 175-184, March 1952.

- [5] A. Holbrook and G. Fairbanks, "Diphthong Formants and Their Movements," J. Speech Hearing Res., 5(1): 38-58, March 1962.
- [6] V.R. Lesser, R.D. Fennell, L.D. Erman and D.R. Reddy, "Organization of the Hearsay-II speech understanding system," IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-23(1); 11-23, 1975.
- [7] A. Weibel, T. Hanazawa, G. Hinton, K. Shikano, and K.J. Lang, "Phoneme Recognition Using Time Delay Neural Networks," IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-37:328-339, 1989.
- [8] D.W. Tank and J.J. Hopfield, "Neural Computation by Concentrating Information in Time," Proc. Nat. Academy Sciences, 84:1896-1900, April 1987.
- [9] E. Levin, "Word Recognition Using Hidden Control Neural Architecture," Proc. ICASSP 90, 433-436, Albuquerque, NM, April 1990.

فصل ششم

الگوریتم پیشنهادی برای

شناسایی آواها و نتایج

عملی

۶-۱- مقدمه

همانطور که در فصل‌های پیشین بیان گردید، هدف از طراحی سیستم‌های تشخیص اتوماتیک گفتار ایجاد رابطه از طریق زبانی بین کامپیوتر و انسان است تا از این طریق بتوان دستوراتی را به آن داد. بعنوان نمونه بجای آنکه یک متن را از طریق صفحه کلید به کامپیوتر انتقال دهیم، براحتی می‌توان با طراحی سیستم تشخیص گفتار، متن را خواند و سپس بوسیله آن به کامپیوتر انتقال داد.

در این فصل قصد داریم تا با ارائه یک الگوریتم مناسب و جدید جهت استخراج ویژگی از آواها بتوانیم آنها را دسته‌بندی کنیم. آواهای یکسان تولید شده بوسیله افراد مختلف با سن یا جنس متفاوت یا توسط همان فرد با فرکانس اصلی⁸⁴ متفاوت باید بوسیله سیستم شنیداری ما قابل تشخیص باشد. در واقع سیستم باید بتواند آواها را بدون توجه به مرد، زن یا کودک ادا کننده آواها تشخیص دهد. هدف ما بررسی فرضیه بکار بستن تبدیل میلیون برای استخراج ویژگی از طیف فرکانسی آواها است که این ویژگیها باید مستقل از گوینده باشند. بدون شک می‌توان ادعا نمود که مهمترین بخش یک سیستم تشخیص الگو، همین بخش استخراج ویژگی از سیگنال است. تاکنون روشهای متنوعی برای استخراج ویژگی از سیگنالهای صوتی ارائه شده است که معروفترین آنها ضرایب طیف فرکانسی مل⁸⁵ و محاسبه خطی ادراکی⁸⁶ است.

در این فصل در ابتدا رابطه بین آواها و سپس روشهای استخراج پوش طیفی از یک سیگنال مورد بررسی قرار می‌گیرد. پس از آن روش کپستروم بهبود یافته معرفی می‌گردد که با استفاده از آن پوش طیفی آواها را با دقت و کنترل بیشتری بدست می‌آوریم. بدنبال آن با استفاده از الگوریتم تبدیل میلیون سریع که

⁸⁴ . pitch

⁸⁵ . Mel frequency Cepstral Coefficient (MFCC)

⁸⁶ . Perceptual Linear Prediction (PLP)

در فصل سوم درباره آن صحبت نمودیم، دامنه تبدیل مقیاس پوش طیفی را بدست می‌آوریم و در پایان با ترکیب مراحل فوق با یکی از روشهای مهم استخراج ویژگی، بردار ویژگی هر آوا را استخراج می‌کنیم و آنها را بر اساس بردارهای مرجع دسته بندی می‌کنیم.

نتایج و آزمایشات انجام شده نشان می‌دهد که روش فوق نه تنها دارای دقت بالایی می‌باشد، بلکه با توجه به اینکه به تعداد ویژگی خیلی کمی نیاز دارد، سرعت بالاتری نسبت به الگوریتمهای ارائه شده قبلی در کاربردهای تشخیص آوا از خود نشان می‌دهد.

۶-۲- رابطه میان آواها و افراد مختلف

هنگامیکه از رادیو و یا تلفن صدای ناشناسی را می‌شنویم، سریعاً با یک روش تشخیص بسیار پیچیده و قدرتمند، در ابتدا تشخیص می‌دهیم که گوینده بزرگسال است یا خردسال و اگر بزرگسال است زن است یا مرد. هدف ما در این رساله، شناسایی سن و جنسیت گوینده نیست، بلکه بدنبال آن هستیم که آواهای یکسان ادا شده توسط افراد مختلف را به توزیعهای یکسان نگاشت دهیم و برای انجام اینکار بایستی شناخت کاملی از تفاوت‌های بین آواهای یکسان ادا شده توسط افراد مختلف داشته باشیم.

در شکل ۶-۱ چند آوای مختلف نشان داده شده است. همانگونه که در این شکل می‌بینید، آواها، اصواتی با ماهیت پریودیک دارند. عکس دوره تناوب یک سیگنال آوایی را فرکانس اصلی نامند، از طرفی فرکانسهای متناظر با پیکهای اندازه تبدیل فوریه سیگنال، برابر با فرکانسهای رزونانس می‌باشند که در اکثر آزمایشات انجام شده سه فرکانس اول از درجه اهمیت بالاتری نسبت به بقیه فرکانسهای رزونانس برخوردار می‌باشند. ساختار آواهای ادا شده توسط افراد مختلف با جنسیتها و سنین متفاوت از دو لحاظ با یکدیگر تفاوت دارد. مردها دارای فرکانس اصلی پایینی می‌باشند، در حالیکه بچه‌ها دارای فرکانس اصلی بالا و در زنها بین این دو قرار دارد [۱]. فرکانس اصلی بوسیله سرعت باز و بسته شدن تارهای صوتی (سرعت پالس حنجره⁸⁷) تعیین می‌گردد. تفاوت دیگر، اختلاف طول حفره صوتی در افراد مختلف است. طول حفره صوتی در مردان بیشترین، در بچه‌ها کمترین و در زنها متوسط است (Fitch & Giedd, 1999). همانگونه که یک کودک بزرگ می‌شود (بین ۴ تا ۱۲ سالگی)، طول حفره صوتی بسته به جنسیت، با سرعت ثابتی افزایش یا

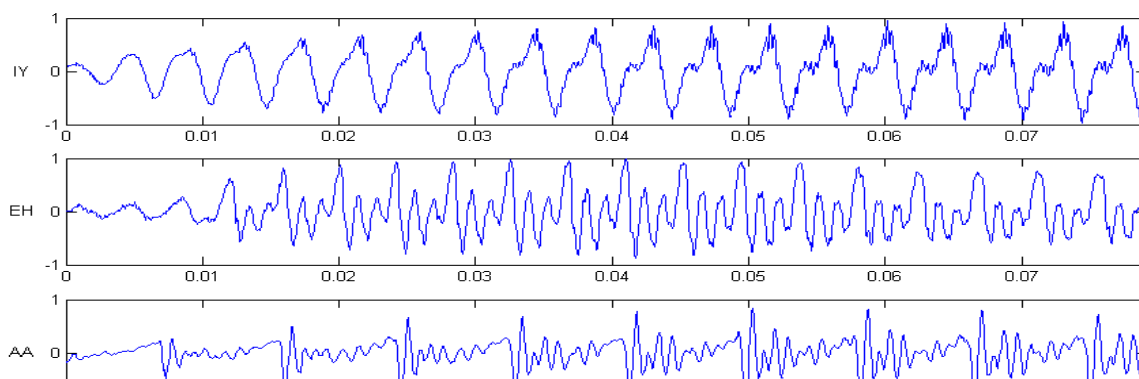
⁸⁷ . Glottal Pulse Rate (GPR)

کاهش می‌یابد که در نتیجه آن، اندازه فرکانسهای رزونانس کاهش یا افزایش می‌یابد، در نتیجه اختلاف در طول حفره صوتی منجر به شیفت فرکانسهای رزونانس سیگنال صوتی می‌گردد، بعبارت دیگر، هر چقدر طول حفره صوتی بیشتر می‌گردد، فرکانسهای رزونانس بیشتر به سمت فرکانسهای پایین انتقال می‌یابند (Fant, 1970).

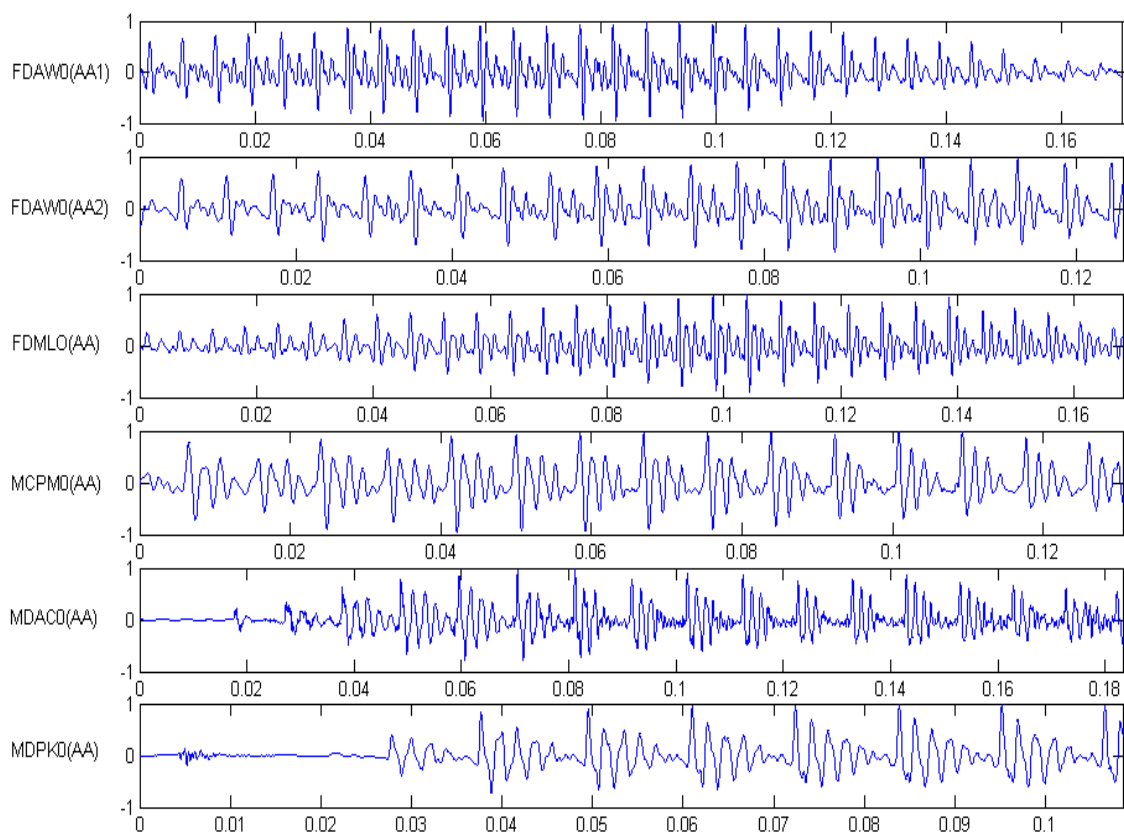
در آواها شکل حفره صوتی گوینده تعیین کننده نوع آوا است [۲]. شکل حفره صوتی معمولا با استفاده از محل قرار گرفتن زبان در دهان تعیین می‌گردد و تغییرات آن می‌تواند بر روی فرکانسهای رزونانس تاثیر گذار باشد..

از لحاظ زمانی، سیگنالهای آواهای ادا شده توسط یک فرد در زمانهای مختلف بدین علت که دارای فرکانس اصلی متفاوتی می‌باشند، با هم یکسان نمی‌باشند. از طرفی چون طول حفره صوتی در افراد مختلف با یکدیگر متفاوت است، آواهای یکسان ادا شده توسط افراد مختلف نیز با هم برابر نمی‌باشند. بعنوان نمونه، در شکل ۶- ۲ آوای 'aa' که توسط سه گوینده زن و سه گوینده مرد گفته شده است، نشان داده شده است، دو آوای اول متعلق به یک فرد است. این آواها از مجموعه داده TIMIT استخراج شده است. اساسا، حفره صوتی هر فرد را می‌توان با یک لوله صوتی بدون تلفات مدلسازی نمود که این لوله بین افراد مختلف فقط از لحاظ طول متفاوت می‌باشد. پاسخ ضربه این لوله‌های صوتی حالت مقیاس‌بندی شده یکدیگر است. مقیاس‌بندی در واقع انبساط یا انقباض محور زمانی تابع اصلی است. بنابراین، اگر یک پریود از سیگنال هر آوا استخراج گردد و آنرا با یک پریود از سیگنال همان آوا ولی ادا شده توسط شخص دیگری (با طول حفره صوتی متفاوت) مقایسه کنیم، دو نوع مقیاس‌بندی شده از یک سیگنال یکسان را داریم (شکل ۶- ۲ را ببینید).

حال، مشکل ما با توجه به ایده فوق، استخراج یک پریود از سیگنال آوا، یافتن نقاط شروع و پایانی و مقایسه (مستقیم در حالت اصلاح فرکانس اصلی و با نرمالیزه نمودن مقیاس برای حفره‌های صوتی مختلف) سیگنالها است. یافتن یک پریود کاری بسیار دشوار است (مسئله یافتن فرکانس اصلی)



شکل ۶-۱ نمودار زمانی چند آوای مختلف



شکل ۶-۲ مقایسه‌ای بین نمودار زمانی آواهای یکسان ادا شده توسط یک فرد در زمانهای مختلف و افراد مختلف.

و نتایج بدست آمده ممکن است با توجه به کوتاه بودن یک پریود (تعداد نمونه‌های بسیار کم) مناسب نباشد و همچنین می‌تواند نتایج متفاوتی با توجه به روش ارائه شده داشته باشد. با توجه به نمودارهای ارائه شده در شکل ۶-۲ براحتی می‌توان فهمید که کار در حوزه زمان بسیار دشوار است و از طرف دیگر اطلاعات کمی را برای شناسایی آواها در اختیار ما قرار می‌دهد. ولی روش دیگری وجود دارد که می‌توان با استفاده از آن این مشکلات را تا حدودی رفع نمود. همانطور که می‌دانید استفاده از حوزه‌های مختلف جهت بررسی سیگنالها می‌تواند برای استخراج ویژگیهای بیشتر، مفید باشد. یکی از حوزه‌های بسیار معروف در تحلیل سیگنالها حوزه فرکانس است که معمولاً با محاسبه تبدیل فوریه سیگنال صورت می‌گیرد. با اینکار سیگنال از حوزه زمان به حوزه فرکانس انتقال می‌یابد. استفاده از حوزه فرکانس دو فایده مهم دارد: اول اینکه، ما زمان را از دست می‌دهیم، با اینکار شیفت زمانی بین سیگنالها بی‌اهمیت می‌شود و مشکل همزمان نمودن سیگنالها از بین می‌رود و دوم اینکه در صورت کار کردن در حوزه فرکانس نیاز به یک سری پردازشهای اولیه زمانی مانند: یافتن یک پریود، نقطه شروع و نقطه پایانی و ... از بین می‌رود.

با توجه به روشهای موجود، برای انتقال سیگنال از حوزه زمان به حوزه فرکانس روشهای متفاوتی وجود دارد. ایده ما در این قسمت محاسبه طیف فرکانسی سیگنال (فقط فرکانسهای مثبت) است. الگوریتمهای مختلفی برای استخراج پوش طیفی یک سیگنال وجود دارد که از آن جمله می‌توان به روشهای بانک فیلتر، کانالهای کد کننده صوتی⁸⁸، کپستروم⁸⁹ و کد کننده‌های خطی اشاره نمود. با توجه به آزمایشات انجام شده و با مقایسه‌ای که بین پوش طیفی بدست آمده از روشهای فوق انجام گرفت، بدین نتیجه رسیدیم که روش کپستروم، با اعمال تغییراتی، می‌تواند معیار مناسبی برای نمایش پوش طیفی یک سیگنال، با توجه به پارامترهای کنترلی آن، باشد.

⁸⁸ . Channel Vocoder

⁸⁹ . Cepstrum

۳-۶- مراحل اجرای الگوریتم پیشنهادی

همانطور که در ابتدای این فصل بیان نمودیم، هدف از این فصل ارائه الگوریتمی جدید جهت استخراج ویژگی از آواها برای تشخیص آنها می‌باشد. الگوریتم پیشنهادی ما شامل مراحل زیر است:

۱- با استفاده از روش کپستروم بهبود یافته (پیشنهاد شده در این رساله) سیگنال صوت از حوزه زمان به حوزه فرکانس انتقال می‌یابد و پوش طیفی آن محاسبه می‌گردد.

۲- محاسبه دامنه تبدیل مقیاس پوش طیفی با استفاده از الگوریتم تبدیل میلیون سریع.

۳- محاسبه ضرایب MFCC یا PLP.

۴- دسته‌بندی آواها بر اساس بردارهای ویژگی استخراج شده.

در ادامه راجع به هر یک از مراحل فوق بطور کامل توضیح داده می‌شود.

۳-۶-۱- تحلیل طیفی

نمایش زمان - فرکانس، تغییر شکل زمانی طیف محاسبه شده از فریم‌های زمانی را در اختیار ما قرار می‌دهد. مفهوم پوش طیفی استخراج شده از اینگونه نمایشات، بیشتر از سیستم‌های تشخیص و تولید صوت آمده است: برای تولید صوت تارهای صوتی بصورت یک تحریک و دهان و دماغ بعنوان سیستم تحریک کننده یا غیر تحریک کننده عمل می‌کنند. آواها دارای طیف فرکانسی هارمونیک می‌باشند که پوش طیفی بر روی آن قرار می‌گیرد. همین ویژگی باعث می‌شود بتوان براحتی آواها را از بقیه اصوات جدا نمود [۳]: ما بدنبال پوش طیفی یک سیگنال بعنوان نشانه‌ای برای شناسایی و طبقه‌بندی اصوات هستیم.

بهرحال، معمولاً تحلیل‌های سیگنالی برای استخراج پوش طیفی از یک طیف دارای شرایط خاصی می‌باشند. اگر صوت کاملاً هارمونیک باشد، می‌توان گفت که پوش طیفی، منحنی است که از میان تمام نقاط متناظر با این هارمونیکها عبور می‌کند. به بیان دیگر، پوش طیفی هموار شده طیف فرکانسی است که از هر

گونه تغییرات شدید جلوگیری نموده و در عین حال سعی می‌کند شکل کلی طیف را حفظ کند. برای انجام اینکار از روشهای مختلفی می‌توان استفاده نمود که از آن جمله به روشهای زیر می‌توان اشاره نمود:

- ۱- **کانالهای کد کننده صوتی** از باندهای فرکانسی استفاده نموده و تخمینی از دامنه سیگنال را در داخل این باند بدست می‌آورد که در نتیجه آن پوش طیفی حاصل می‌گردد.
- ۲- **محاسبه خطی** یک فیلتر تمام قطب را تخمین می‌زند که منطبق بر مولفه‌های طیفی صوت است. هنگامیکه مرتبه این فیلتر کم باشد، فقط فرکانسهای رزونانس بدست می‌آید که همان پوش طیفی است.
- ۳- **روشهای کپستروم** هموار نمودن طیف لگاریتمی شده بدست آمده با استفاده از تبدیل فوریه را بر عهده دارد تا بدین وسیله این منحنی را به قسمتی با تغییرات کند (پوش طیفی) و قسمتی با تغییرات شدید (سیگنال مبدا⁹⁰) تقسیم بندی نماید.

۶-۳-۱-۱- کپستروم

روش کپستروم امکان محاسبه پوش طیفی یک سیگنال را با استفاده از مقادیر تبدیل فوریه سیگنال پنجره‌بندی شده می‌دهد. بدین منظور می‌توان از روشهای قرار دادن صفر و پنجره‌های گوسین یا همینگ⁹¹ بسته به تعداد نقاطی که برای محاسبه پوش طیفی بکار می‌روند، استفاده نمود. کپستروم با استفاده از تبدیل فوریه گسسته محاسبه می‌گردد. طبق تعریف تبدیل فوریه سیگنال پنجره‌بندی شده $x(n)$ بصورت زیر بیان می‌شود:

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn} = |X(k)| e^{j\varphi_x(k)}, \quad k = 0, 1, \dots, N-1 \quad (۶-۱)$$

با محاسبه لگاریتم رابطه فوق

$$\hat{X}(k) = \log X(k) = \log |X(k)| + j\varphi_x(k) \quad (۶-۲)$$

⁹⁰ . Source signal

⁹¹ . Gaussian or Hamming windows

و محاسبه عکس تبدیل فوریه سیگنال $\hat{X}(k)$ ، که منجر به ایجاد کپستروم مختلط⁹² می‌گردد:

$$\hat{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{X}(k) W_N^{-kn} \quad (۳-۶)$$

کپستروم حقیقی از قسمت حقیقی رابطه (۶-۲) بدست می‌آید که در آن

$$c(n) = \frac{1}{N} \sum_{k=0}^{N-1} \hat{X}_R(k) W_N^{-kn}, \quad \hat{X}_R(k) = \log|X(k)| \quad (۴-۶)$$

با توجه به اینکه $\hat{X}_R(k)$ تابعی زوج است، تبدیل فوریه معکوس گسسته آن تابعی زوج است که می‌توان آنرا بر حسب کپستروم مختلط $\hat{x}(n)$ و بصورت زیر بیان نمود:

$$c(n) = \frac{\hat{x}(n) + \hat{x}(-n)}{2} \quad (۵-۶)$$

شکل ۶-۳ مراحل لازم برای محاسبه پوش طیفی را با استفاده از روش کپستروم حقیقی نشان می‌دهد. کپستروم حقیقی $c(n)$ در واقع عکس تبدیل فوریه لگاریتم دامنه تبدیل فوریه سیگنال پنجره‌بندی شده $x(n)$ است. پنجره پایین گذر برای وزندار نمودن کپستروم $c(n)$ استفاده می‌گردد و از رابطه زیر بدست می‌آید:

$$\omega_{lp}(n) = \begin{cases} 1 & n = 1, N_1 \\ 2 & 1 \leq n < N_1 \\ 0 & N_1 < n \leq N-1 \end{cases} \quad (۶-۶)$$

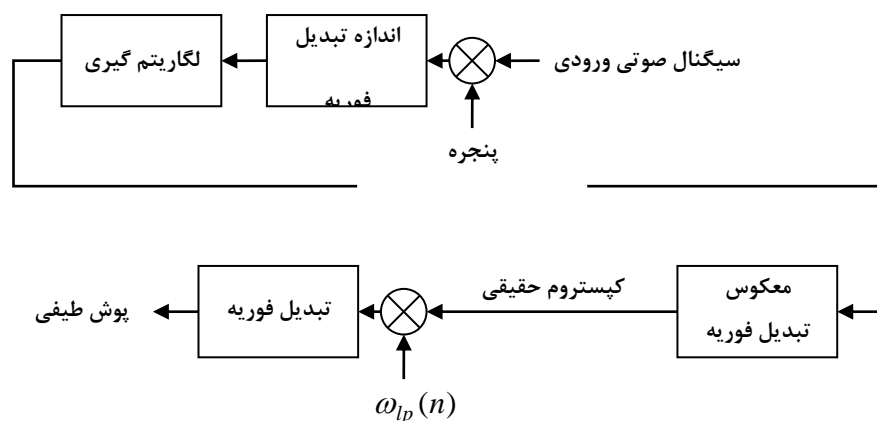
که در رابطه فوق N_1 پارامتر قطع است و بایستی $N_1 \leq N/2$. تبدیل فوریه کپستروم پنجره‌بندی شده $c_{lp}(n)$ پوش طیفی را در اختیار ما قرار می‌دهد

$$C_{lp}(k) = FFT[c_{lp}(n)] \quad (۷-۶)$$

حاصل رابطه فوق، هموار شده طیف $X(k)$ بر حسب dB است. بعنوان نمونه، در شکل ۶-۴ مثالی برای درک بهتر کپستروم حقیقی یک سیگنال آورده شده است.

⁹² . complex cepstrum

همانگونه که در این شکل نشان داده شده است، پارامتر قطع یا همان N_1 برابر با ۲۰ انتخاب شده است، در نتیجه قسمت اول کپستروم $c(n)$ که در آن $0 \leq n \leq 20$ است، توسط پنجره پایین گذر وزندار می‌گردد و حاصل آن $c_{lp}(n)$ است. عکس تبدیل فوریه $c_{lp}(n)$ پوش طیفی را بر حسب دیبی



شکل ۳-۶ بلوک دیاگرام مراحل محاسبه روش کپستروم حقیقی

می‌دهد، همانگونه که در سمت راست شکل ۳-۶ نشان داده شده است. قسمت بالاگذر کپستروم $c(n)$ که در آن $20 < n \leq 512$ است، سیگنال مبدا را نشان می‌دهد، که اولین پیک فرکانسی در $n = 218$ نشان دهنده دوره تناوب فرکانس اصلی^{۹۳} T_0 است که بر حسب فرکانس نمونه‌برداری و تعداد نمونه‌ها بدست می‌آید [۳]. در نتیجه $f_0 = 24000 \text{ Hz} / 218 \approx 110 \text{ Hz}$. به بیان دیگر با استفاده از روش کپستروم می‌توان دوره تناوب فرکانس اصلی را با استفاده از اولین پیک موجود در طیف فرکانسی بدست آورد.

نکات مورد توجه در محاسبه کپستروم عبارتند از:

- فیلتر نمودن پایین گذر با پنجره بندی (صفر نمودن مقادیر بالاتر از پارامتر قطع) انجام می‌گیرد. این عملیات مشابه با فیلتر نمودن در حوزه فرکانسی با $\sin(f)/f$ است. از حالت‌های دیگری نیز برای هموارتر نمودن انتقال، بجای قطع نمودن ناگهانی در حوزه کپستروم استفاده نمود.
- روش کپستروم تخمینی از پوش طیفی را در اختیار ما قرار می‌دهد. اما، مقادیر لگاریتمی می‌تواند به ازای مقادیر صفر به سمت منفی بینهایت میل نماید. بدین منظور مقادیر لگاریتمی را محدود

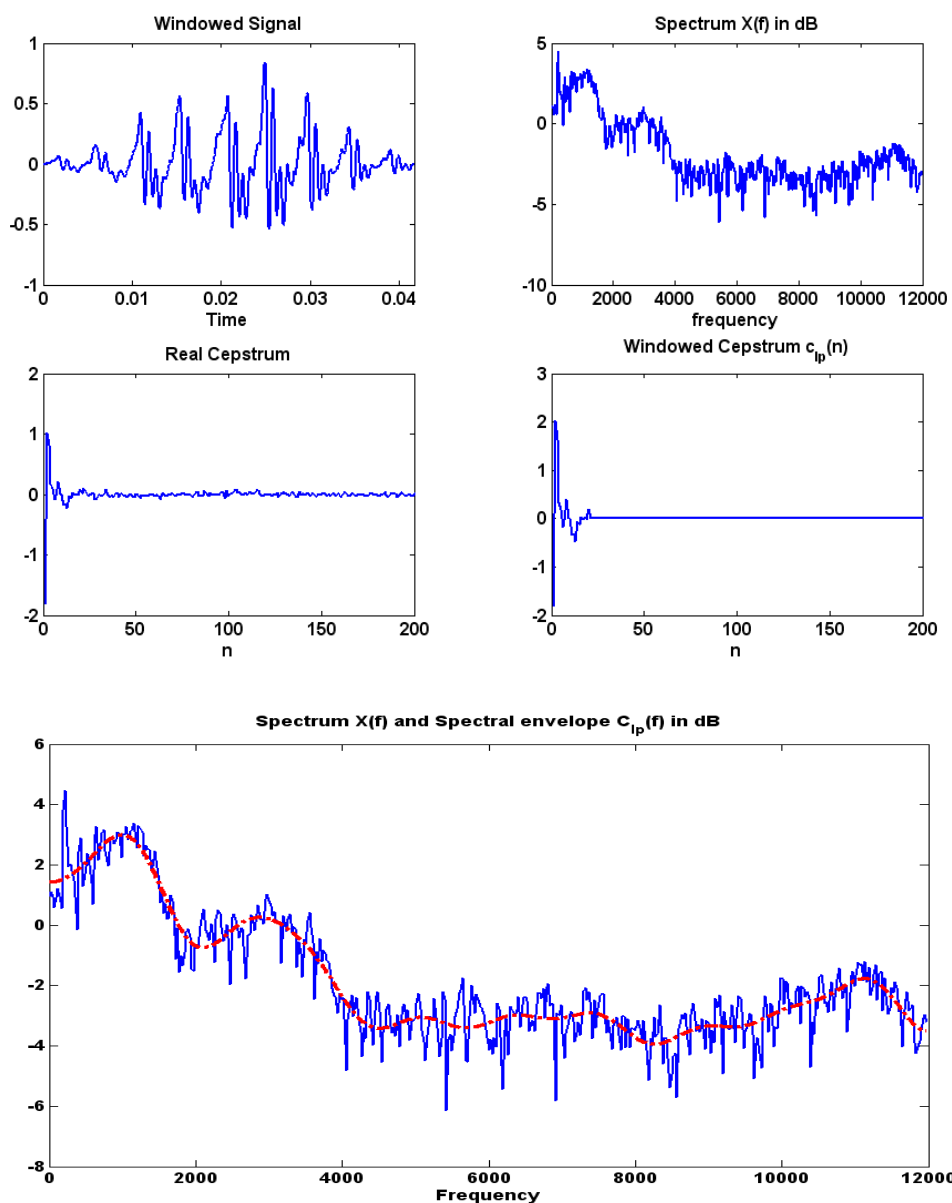
^{۹۳} . pitch period

می‌سازیم. بعنوان مثال مقدار تبدیل فوریه را می‌توان با یک عدد مثبت نزدیک صفر جمع نمود تا هیچگاه حاصل لگاریتم بینهایت نشود.

- برای بهبود پوش طیفی محاسبه شده، می‌توان از الگوریتمهای تکرار پذیر استفاده نمود تا از این طریق فقط اختلافهای مثبت بین طیف لحظه‌ای و پوش طیفی تخمین زده شده در هر قدم محاسبه گردد.

- با وجود اینکه کپستروم حقیقی دارای کاربرد وسیعی است، اما می‌توان از کپستروم مختلط نیز برای محاسبه پوش طیفی یک سیگنال استفاده نمود.

در شکل ۶-۵ پوش طیفی حاصل از روش کپستروم معمولی به ازای آواهای یکسان شکل ۲-۶ آورده شده است.



شکل ۴-۶ نمودار سیگنال پنجره بندی شده، طیف فرکانسی بر حسب دیبی (طول تبدیل فوریه ۱۰۲۴)، کپستروم حقیقی، کپستروم پنجره بندی شده (پارامتر قطع برابر ۲۰ است) و پوش طیفی

همانطور که در این شکل نشان داده شده است، در حوزه فرکانس شباهت بین منحنی‌های آواهای یکسان ادا شده بوسیله افراد مختلف نسبت به حوزه زمان بیشتر شده است اما هنوز کاملاً شبیه یکدیگر نمی‌باشند و در اینحالت ویژگیهای بهتری را می‌توان برای دسته بندی آواها استخراج نمود. اما نتیجه حاصل نمی‌تواند بعنوان ورودی مرحله بعد سیستم (محاسبه دامنه تبدیل مقیاس) استفاده گردد. علت آن نیز

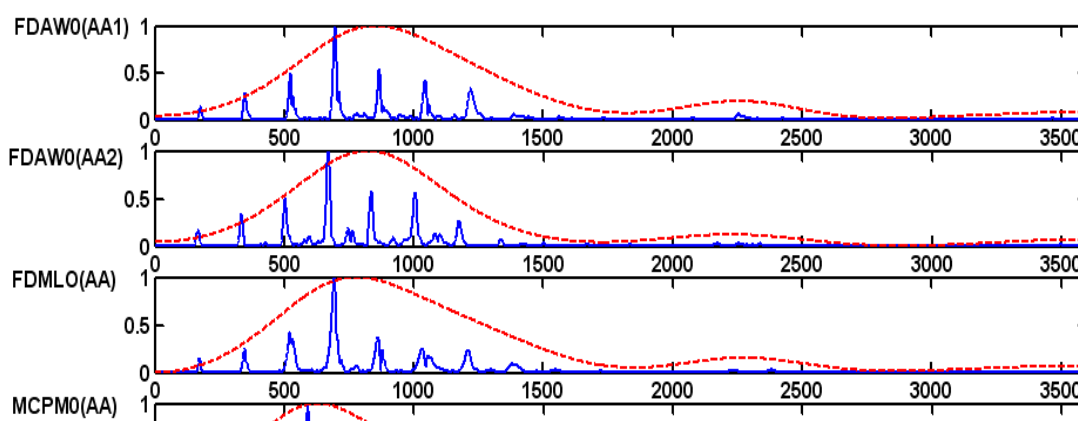
بخاطر این است که در اینحالت پوشهای طیفی حاصل، همانگونه که در شکل ۶-۶ نشان داده شده است، حالت مقیاس بندی شده یکدیگر (فشرده شده یا انبساط یافته محور فرکانسی) نمی‌باشند. از اینرو دامنه تبدیل مقیاس منحنی‌های پوش طیفی آواهای یکسان با هم برابر نخواهد بود.

در روش کپستروم دو پارامتر کنترلی برای هموارتر نمودن و دقیقتر نمودن پوش طیفی وجود دارد که پارامتر قطع (N_1) و میزان وزندار نمودن کپستروم ($\omega_{ip}(n)$) است. در شکل ۶-۷ میزان تغییرات پوش طیفی به ازای تغییرات پارامتر قطع و ثابت ماندن پنجره وزندار و در شکل ۶-۸ به ازای ثابت بودن پارامتر قطع و تغییرات پارامتر پنجره وزندار نشان داده شده است. هدف ما از محاسبه پوش طیفی سیگنال آوا، بدست آوردن یک رابطه منطقی بین آواهای یکسان ادا شده توسط افراد مختلف است. از اینرو با اعمال تغییراتی در روش کپستروم و اعمال نتایج آن در الگوریتم کلی به نتایج بهتری برای افزایش دقت دسته‌بندی آواهای مختلف دست یافتیم. روشی را در اینجا آنرا معرفی خواهیم نمود که آنرا روش کپستروم بهبود یافته^{۹۴} نامیده‌ایم.

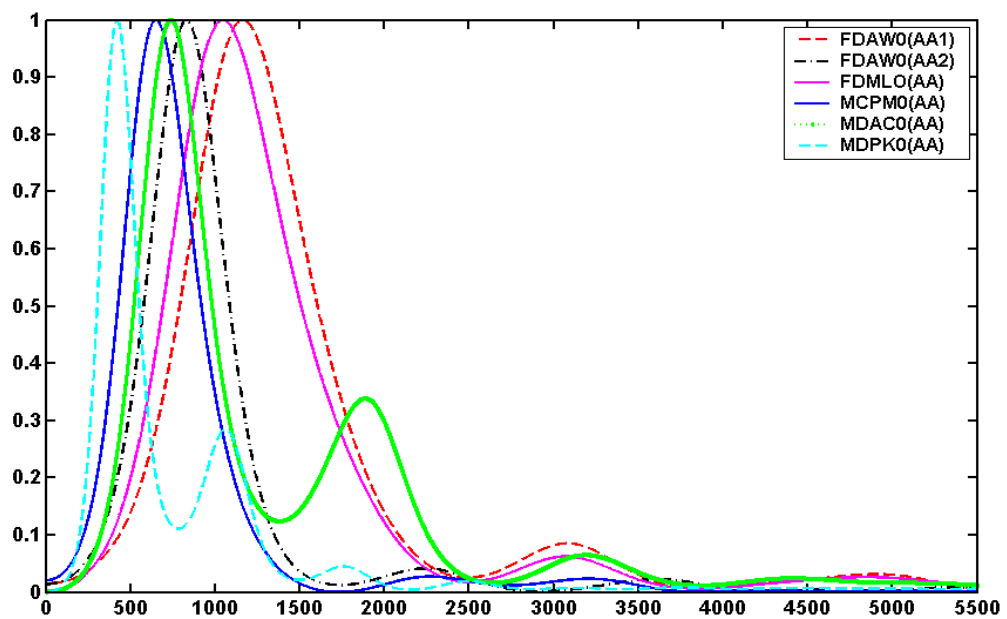
۶-۳-۱-۲- روش ابداعی کپستروم بهبود یافته

یکی از ویژگیهای تابع لگاریتم، فشرده‌سازی اطلاعات است. لازم به ذکر است که در تابع لگاریتم مورد استفاده در روش کپستروم تغییر اندیس پایه تاثیری بر روی پوش طیفی خروجی ندارد و یا عبارت دیگر نسبت به تغییرات آن ثابت است. از اینرو ما بجای استفاده از این تابع از تابع توانی استفاده نمودیم. این تابع نسبت به تابع لگاریتمی دارای این مزیت می‌باشد که دارای یک پارامتر کنترلی برای افزایش دقت سیستم تشخیص آواها است که در ادامه بیشتر در مورد آن صحبت خواهیم نمود.

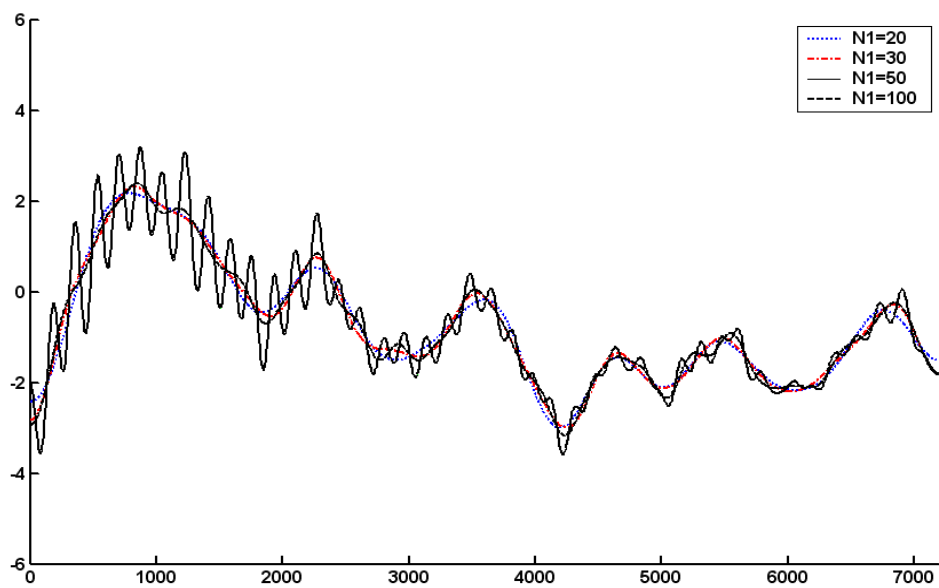
بلوک دیاگرام الگوریتم پیشنهادی کپستروم بهبود یافته در شکل ۶-۹ نشان داده شده است. همانطور که در این شکل نشان داده شده است، در ابتدا سیگنال پنجره بندی شده و سپس به انتهای



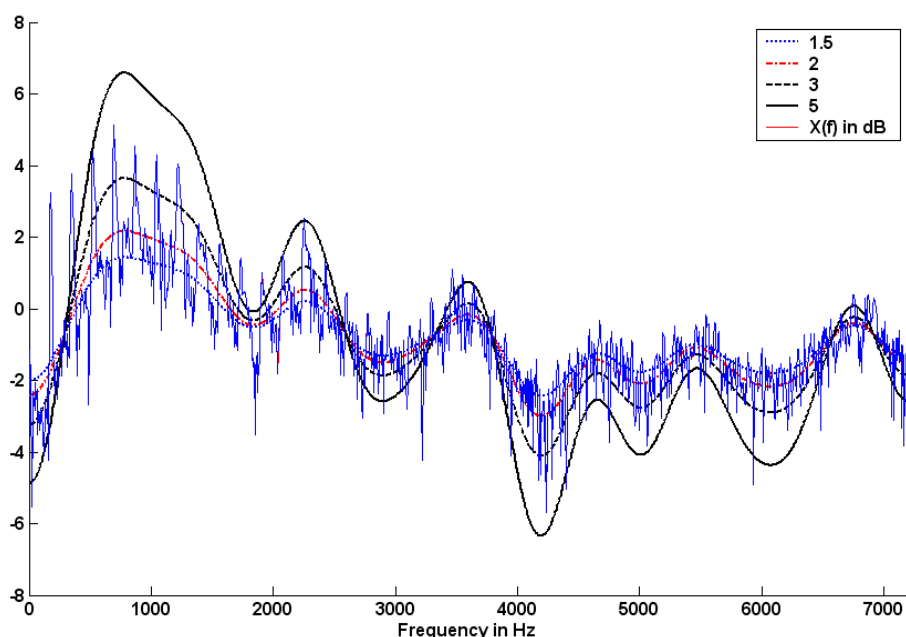
شکل ۶-۵ پوش طیفی آواهای "aa" شکل ۶-۲ با استفاده از روش کپستروم به ازای پارامتر قطع برابر با ۲۰ و استفاده از رابطه (۶-۶).



شکل ۶-۶ مقایسه پوش طیفی آوای "aa" ادا شده توسط افراد مختلف با استفاده از روش کپستروم



شکل ۶-۷ مقایسه پوش طیفی به ازای تغییرات پارامتر قطع با وزندار نمودن یکسان



شکل ۶-۸ مقایسه پوش طیفی به ازای تغییرات پارامتر روزندار نمودن با پارامتر یکسان ($N_1=12$)

آن صفر اضافه می‌شود که اصطلاحاً این عمل را اضافه کردن صفر⁹⁵ می‌نامند. استفاده از این دو روش باعث افزایش رزولوشن فرکانسی می‌گردد که برای تخمین پوش طیفی بهتر موثر خواهند بود. در مرحله بعد دامنه تبدیل فوریه سیگنال محاسبه می‌گردد و پس از نرمالیزه کردن انرژی، بعنوان ورودی تابع توانی $\gamma^{fft(x)}$ قرار داده می‌شود، در اینجا پارامتر γ پارامتر کنترلی سیستم است که بسته به تغییرات آن، دقت سیستم تغییر می‌کند. در ادامه از حاصل فوق عکس تبدیل فوریه گرفته می‌شود و در پنجره پایین گذر $\omega_{lp}(n)$ رابطه (۶-۲) ضرب می‌گردد. در نهایت پس از محاسبه تبدیل فوریه و استفاده از تابع ریشه⁹⁶ حاصل، پوش طیفی سیگنال صوتی ورودی خواهد بود.

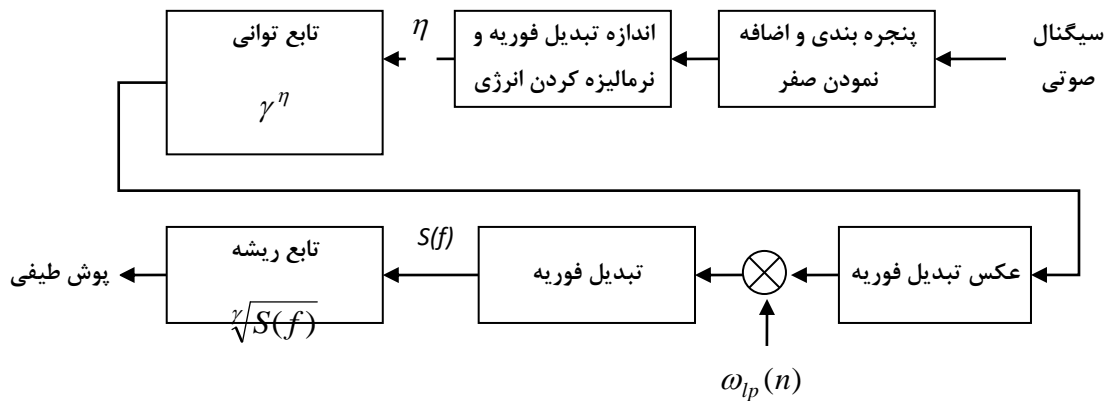
⁹⁵ . zero padding
⁹⁶ . Root function

بطور خلاصه، پوش طیفی حاصل از روش کپستروم بهبود یافته از رابطه زیر بدست می آید:

$$Sp = \sqrt[3]{FFT(\omega_{lp}(n).IFFT(\gamma^{fft(x(n))}))} \quad (۸-۶)$$

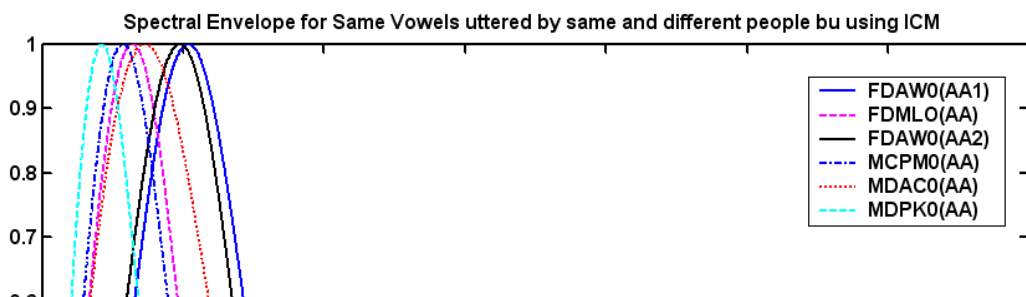
که در رابطه فوق $x(n)$ ، سیگنال پنجره بندی شده است که در انتهای آن صفر اضافه شده است، و Sp پوش طیفی سیگنال ورودی است.

همانطور که قبلا به آن اشاره کردیم، آواهای یکسان ادا شده توسط یک فرد دارای فرکانس اصلی متفاوتی می باشند، این در حالی است که با توجه به یکسان بودن طول حفره صوتی مکان فرکانسهای رزونانس جابجا نمی گردد، در نتیجه با توجه به اینکه فرکانس اصلی تغییر می کند، اما با استفاده از روش کپستروم بهبود یافته، پوش طیفی حاصل وابستگی بسیار کمی نسبت به تغییرات فرکانس اصلی



شکل ۶-۹ بلوک دیاگرام الگوریتم محاسبه پوش طیفی با استفاده از کپستروم بهبود یافته

دارد. از طرف دیگر در افراد مختلف با توجه به تغییرات همزمان فرکانس اصلی و طول حفره صوتی، فقط تغییرات طول حفره صوتی، که باعث جابجایی فرکانسهای رزونانس می گردد، اثر خود را بصورت فشرده گی یا انبساط پوش طیفی نشان می دهد. بمنظور روشنتر شدن مطلب، پوش طیفی حاصل از روش کپستروم بهبود یافته آواهای شکل ۶-۲ در شکل ۶-۱۰ آورده شده است.



شکل ۶-۱۰ پوش طیفی شکل ۲-۶ با استفاده از روش کپستروم بهبود یافته به ازای $\gamma=10$.

۶-۳-۲- تبدیل میلین و تبدیل مقیاس

در قسمت قبل ما توانستیم، تغییرات طول حفره صوتی را بصورت انبساط یا انقباض پوش طیفی آن نشان دهیم، عبارت دیگر:

$$Sp(f) \xrightarrow{\text{تغییرات طول حفره صوتی}} Sp(af) \quad (۶-۹)$$

که در رابطه فوق $Sp(f)$ پوش طیفی سیگنال آوای ورودی و a ضریب مقیاس بندی است که رابطه ای غیر خطی بر حسب فرکانس دارد. این مطلب را براحتی می توان با مقایسه پوش طیفی آواهای یکسان شکل ۶-۱۰ فهمید. در اینجا باید بدنبال تبدیلی باشیم که بتواند منحنی هایی با ضریب مقیاس متفاوت را به منحنی های یکسان نگاشت دهد.

همانطور که در فصل ۲ بیان گردید، یکی از ویژگیهای کلیدی تبدیل میلین و نوع محدود شده آن یعنی تبدیل مقیاس، ثابت ماندن اندازه تبدیل نسبت به تغییرات مقیاس بندی است. این عبارت بدین معنی است که اگر f یک تابع و g نوع مقیاس بندی شده f باشد، دامنه تبدیل مقیاس دو تابع با هم برابر خواهد بود. تغییرات مقیاس، انبساط یا انقباض محور زمانی سیگنال اصلی است که انرژی آن ثابت باقی مانده است. بنابراین، تابع $g(t)$ را می توان با تغییرات مقیاس تابع $f(t)$ بدست آورد، اگر $g(t) = \sqrt{a}f(at)$ باشد، که در این رابطه a مقدار حقیقی و مثبت است. هنگامیکه $a < 1$ باشد، انبساط مقیاس و هنگامیکه $a > 1$ باشد، فشردگی مقیاس را داریم. با تغییر مقیاس بوسیله پارامتر a تبدیلات مقیاس سیگنالهای اصلی و مقیاس بندی شده توسط رابطه زیر بهم مربوط می شوند:

$$D_g(c) = a^{jc} D_f(c) \quad (۱۰-۶)$$

رابطه فوق در واقع برگرفته از روابط تبدیل میلین است. در واقع اگر $h(t) = f(at)$ باشد در اینصورت

$$M_h(p) = a^{-p} M_f(p) \quad (۱۱-۶)$$

در روابط (۱۰-۶ و ۱۱) اثر مقیاس بندی بصورت ضریب نشان داده است و در رابطه (۱۰-۶) چنین ضریبی باعث شیفت فاز اصلی می گردد. بنابراین، دامنه تبدیل مقیاس سیگنال اصلی و سیگنال مقیاس بندی شده با هم برابر است. به بیان دیگر:

$$|D_g(c)| = |D_f(c)| \quad (۱۲-۶)$$

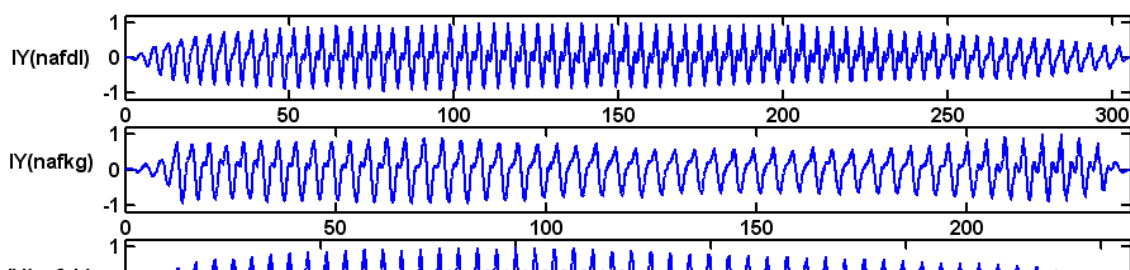
برای محاسبه تبدیل مقیاس پوشهای طیفی، با توجه به اینکه گسسته می باشند، از تبدیل میلین سریع ارائه شده در بخش ۳-۳ استفاده می کنیم که در آن قسمت حقیقی پارامتر میلین برابر با ۵، ۰ انتخاب شده است. یکی از نکاتی که باید در مورد آن صحبت نمود، مجموعه داده هایی است که برای آموزش و تست سیستم استفاده می شود. مجموعه داده TIMIT در برگیرنده جملات بیان شده توسط افراد مختلف است که از کنار هم قرار دادن یک سری واج بوجود آمده اند و برای اجرای الگوریتمهای فوق، ما در ابتدا آواها را از داخل این جملات استخراج نمودیم. اما بنا به دلایلی از جمله مشخص نبودن ابتدا و انتهای بازه هر آوا و وابستگی شکل آوا بسته به واج بیان شده قبل و یا بعد از آن، ما را بر آن داشت تا از مجموعه داده ای که مختص آواها باشد، استفاده نماییم. این مجموعه داده از [۴] گرفته شده است که در آن ۹ کلاس آوایی مختلف وجود دارد که هر کلاس شامل ۳۰ گوینده مختلف اعم از زن، مرد و بچه می باشد که هر گوینده هر آوا را چهار بار

ادا نموده است. بنابراین، در مجموع ۱۰۸۰ آوا داریم. مزیت این مجموعه داده نسبت به آواهای استخراج شده از مجموعه داده TIMIT در این است که هر گوینده فقط آوا را بیان کرده است و نیازی به استخراج آن از داخل جملات وجود ندارد.

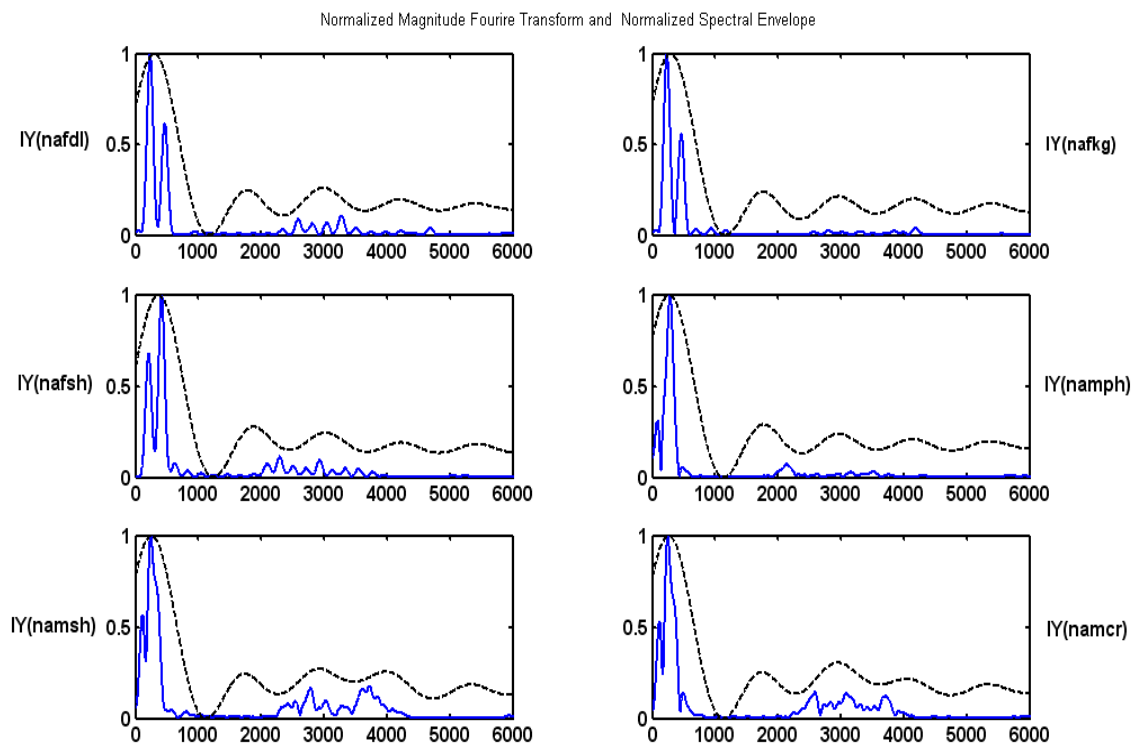
همانگونه که در شکل ۶-۱۱ می‌بینید، مدل زمانی آوای /IY/ که توسط سه گوینده مرد و سه گوینده زن مختلف ادا شده است، آورده شده است. با استفاده از مدل زمانی براحتی می‌توانید اختلاف آواهای یکسان بیان شده توسط افراد مختلف را مشاهده نمایید. در شکل ۶-۱۲ پوش طیفی مربوط به هر یک از آواها، با استفاده از روش کپستروم بهبود یافته، آورده شده است. مسلم است که در اینحالت ویژگیهای بیشتری را می‌توان برای نگاشت آواهای یکسان به فضای یکسان استخراج نمود. در نهایت با توجه به توضیحات ابتدایی این بخش، از تبدیل میلین سریع برای نگاشت پوشهای طیفی به منحنی‌های تقریباً یکسان استفاده نمودیم که نتایج آن در شکل ۶-۱۳ آورده شده است. اما همانطور که در این شکل نشان داده شده است، تبدیل میلین آواهای افراد مختلف با هم کاملاً برابر نیست. علت این امر هم نبودن یک رابطه خطی مناسب بین پوش طیفی افراد مختلف است. برای حل این مشکل از مشتق مراتب بالای پوش طیفی استفاده نمودیم که نتایج آن در شکل ۶-۱۴ آورده شده است. با مقایسه این شکل، که تبدیل مقیاس مشتق درجه سوم پوش طیفی است، و شکل ۶-۱۳ براحتی می‌توان دید که در اینحالت منحنی‌های یکسانی بدست آمده است.

همچنین در شکل‌های ۶-۱۵ تا ۶-۱۷ بترتیب، مدل زمانی، پوش طیفی و دامنه تبدیل مقیاس مشتق مرتبه سوم پوش طیفی آوای /ao(oo)/ که بوسیله افراد مختلف بیان شده است، نشان داده شده است.

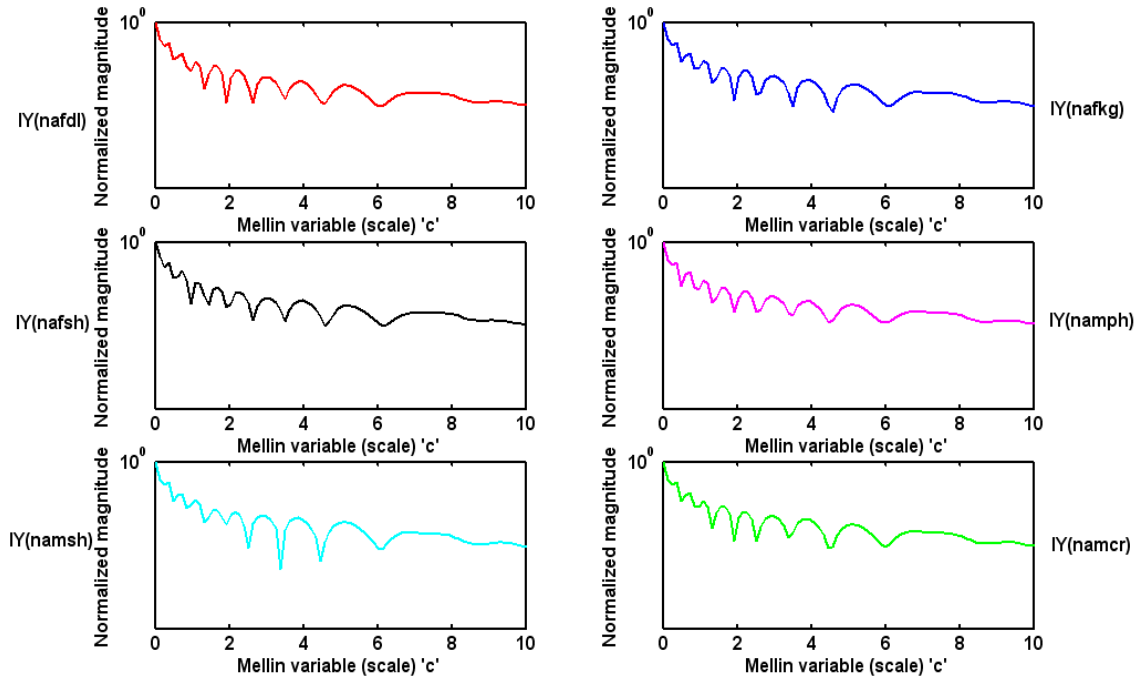
بنابراین با توجه به توضیحات فوق، ما توانستیم آواهای یکسان ادا شده توسط افراد مختلف را به منحنی‌های یکسان نگاشت دهیم. اما بایستی دید الگوریتم ما این توانایی را نیز دارد که آواهای غیر یکسان را به منحنی‌های متفاوت نگاشت دهد یا نه. بهمین منظور دامنه تبدیل مقیاس ۶ آوای مختلف با هم مقایسه گردید که نتایج آن در شکل‌های ۶-۱۸ تا ۶-۲۰ آورده شده است. همانطور که در شکل ۶-۲۰ نشان داده شده است، دامنه تبدیل مقیاس آواهای مختلف با یکدیگر فرق دارد. بنابراین ما با استفاده از مراحل فوق توانستیم آواهای یکسان ادا شده توسط افراد مختلف را که در حالت زمانی دارای سیگنال‌های مختلفی می‌باشند را به منحنی‌های یکسان و آواهای مختلف را که از



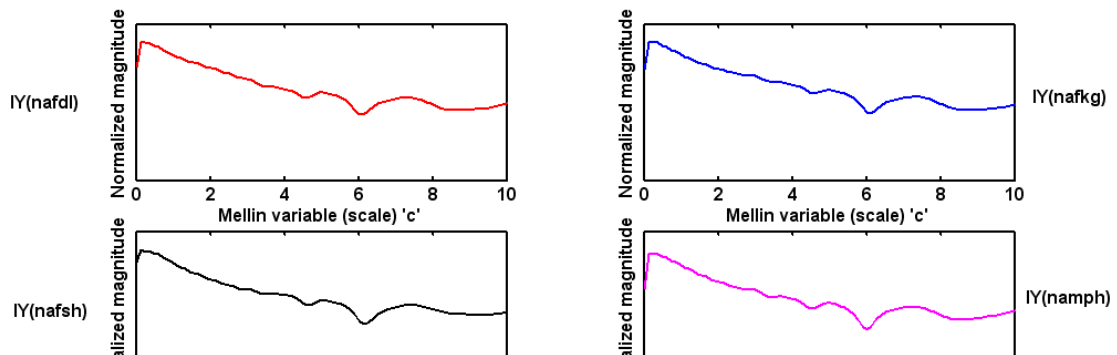
شکل ۶-۱۱ نمودار زمانی آوای /IY/ که بوسیله افراد مختلف ادا شده است. آواها برگرفته از [۴] می باشند



شکل ۶-۱۲ پوش طیفی استخراج شده با استفاده از روش کپستروم بهبود یافته برای سیگنالهای شکل ۶-۱۱



شکل ۶-۱۳ دامنه تبدیل مقیاس منحنیهای شکل ۶-۱۲



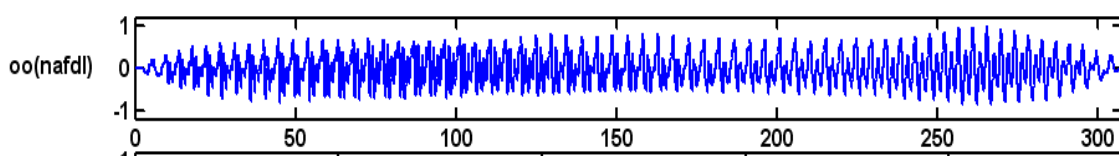
شکل ۶-۱۴ دامنه تبدیل مقیاس به ازای مشتق مرتبه سوم شکل ۶-۱۲

لحاظ زمانی دارای شکلهای متفاوتی می باشند را به منحنی های متفاوت نگاشت دهیم.

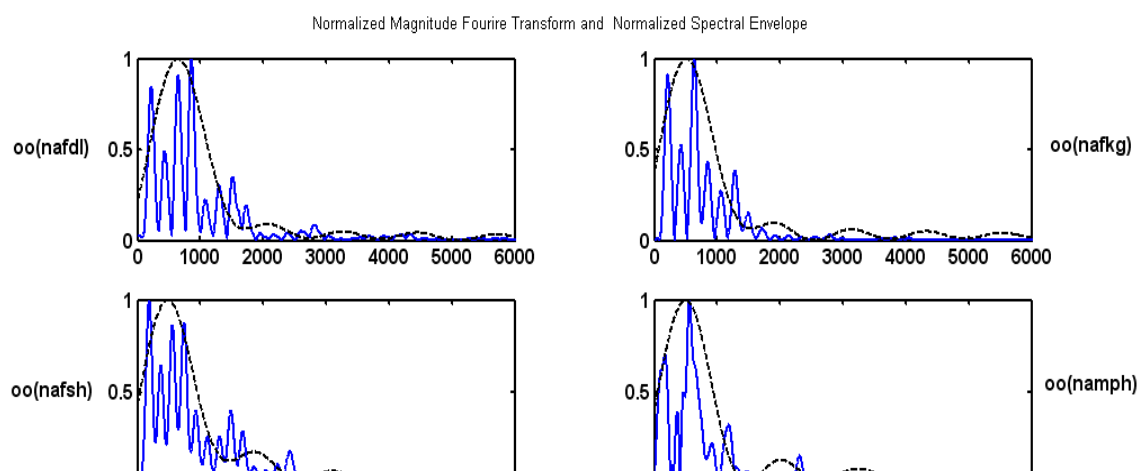
مرحله بعد، استفاده از روشهای استخراج ویژگی برای استخراج ویژگی مناسب از آواهاست. در این مرحله ما به دنبال استخراج ویژگیهایی هستیم که به کمک آنها بتوانیم، آوای ورودی را دسته بندی نماییم. بلوک دیاگرام خلاصه شده الگوریتم اجرایی ما در شکل ۶-۲۱ نشان داده شده است.

۶-۳-۳- محاسبه ضرایب MFCC و PLP

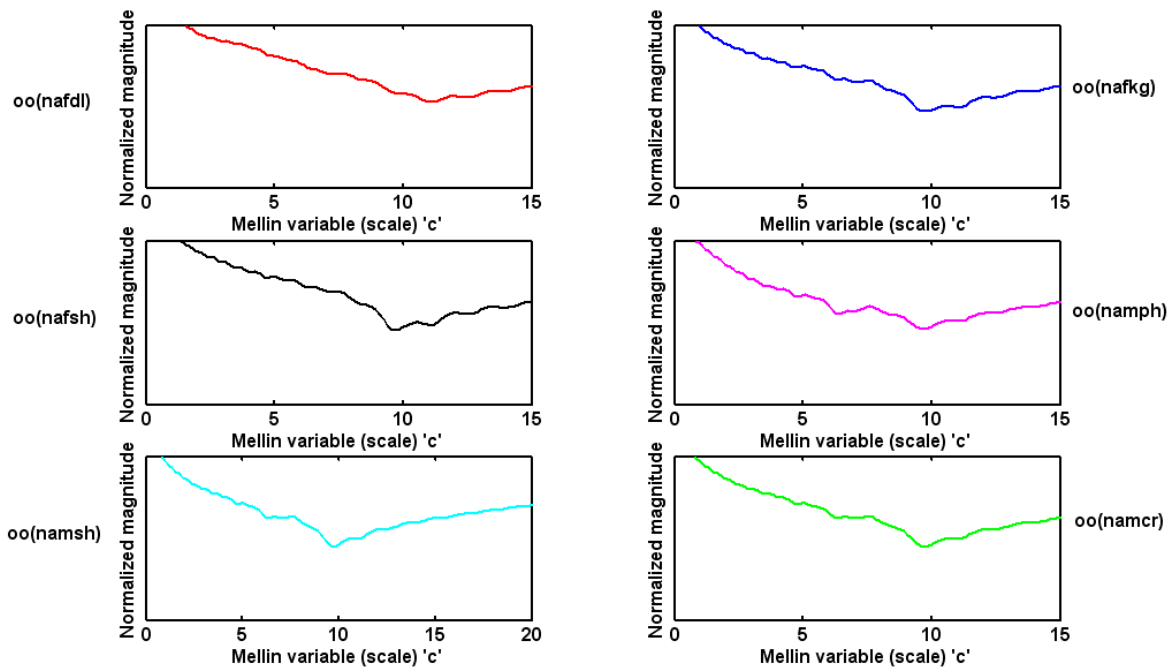
استخراج ویژگی صوتی یکی از مهمترین قسمتهای اکثر سیستمهای تشخیص گفتار اتوماتیک است. اگر این مرحله، دارای ویژگیهای استخراج شده با کیفیت کم باشد، در نتیجه سیستم تشخیص دارای کارایی بسیار کمی خواهد بود. به همین علت مبحث استخراج ویژگی یکی از موضوعات مهم و مورد علاقه هر محققی است.



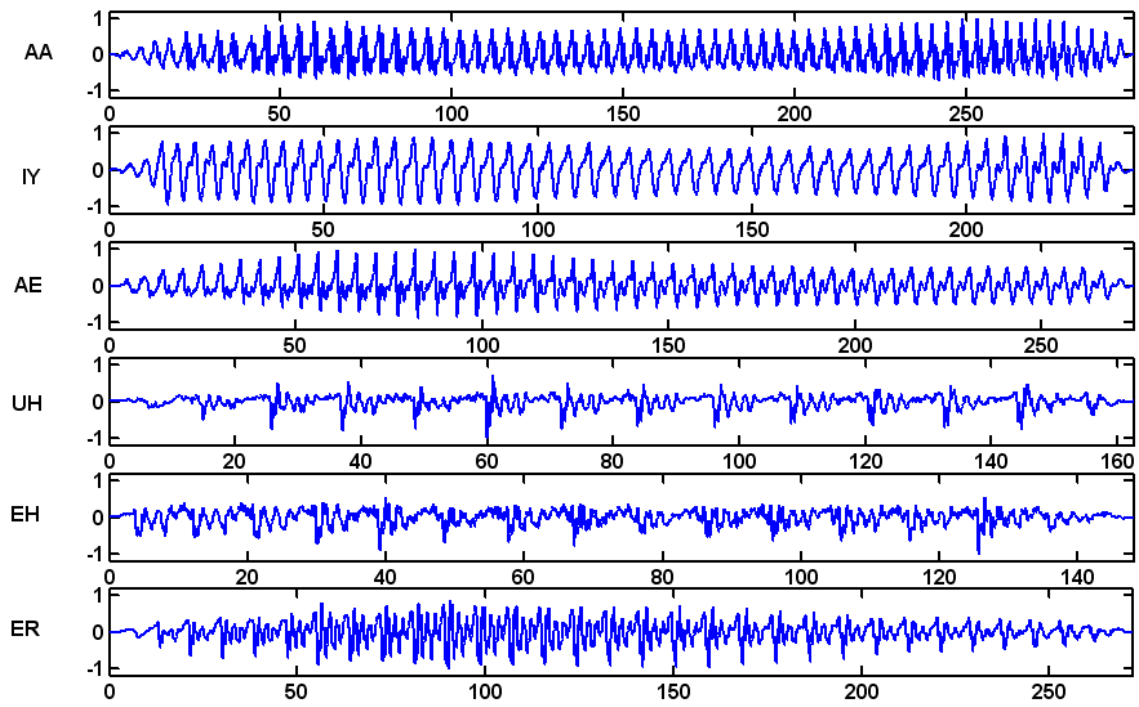
شکل ۶-۱۵ نمودار زمانی آوای /ao/ که توسط افراد مختلف ادا شده است.



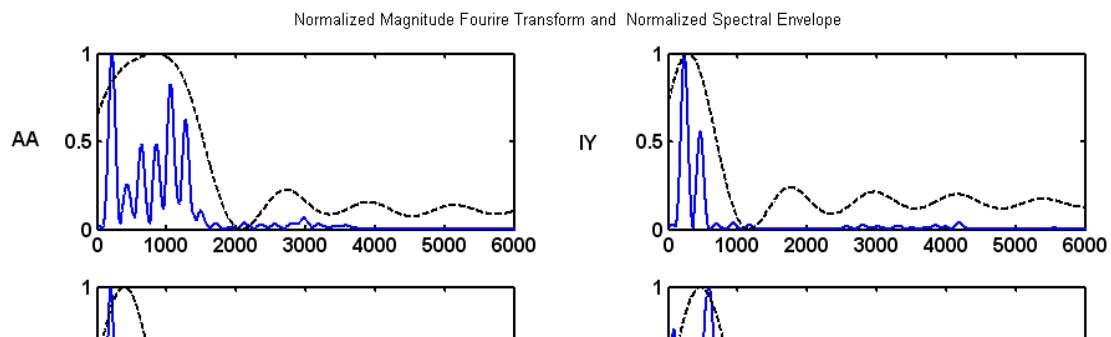
شکل ۶-۱۶ پوش طیفی استخراج شده از روش کپستروم بهبود یافته متناظر با آوای شکل ۶-۱۵



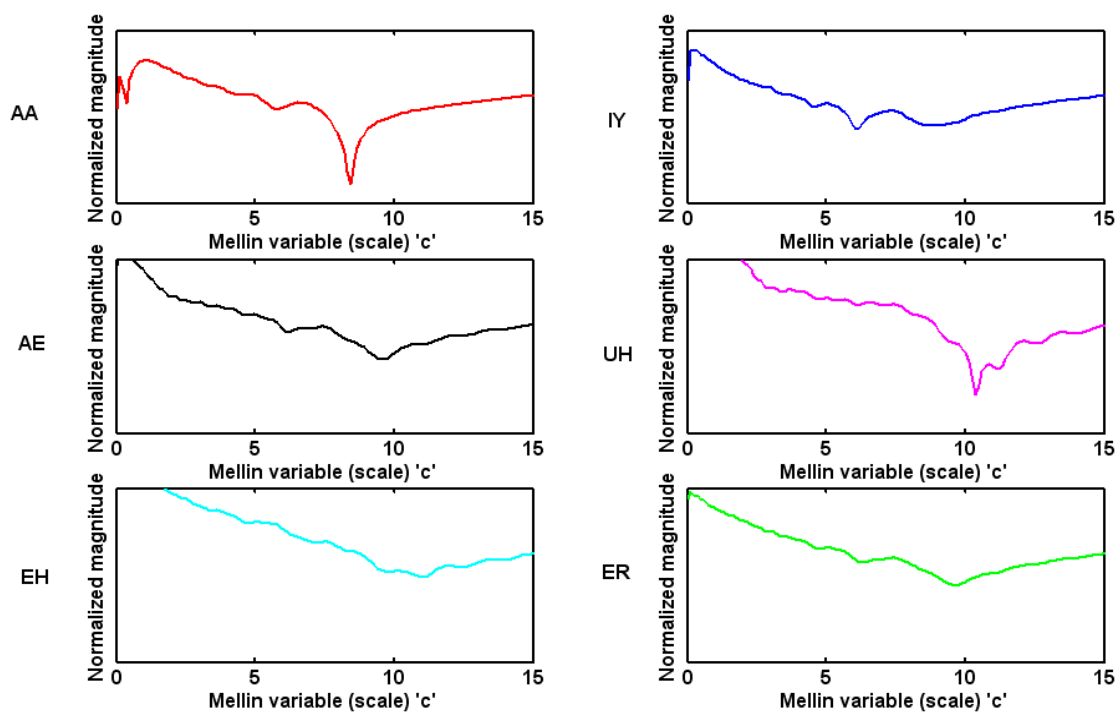
شکل ۶-۱۷ دامنه تبدیل مقیاس مشتق مرتبه سوم شکل ۶-۱۶



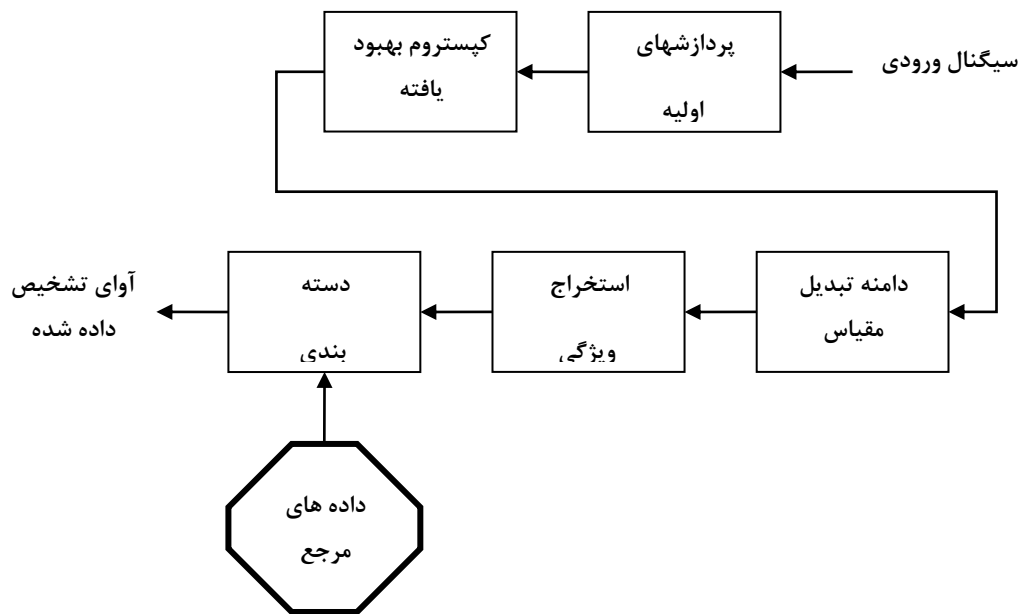
شکل ۶-۱۸ نمودار زمانی آواهای /er/, /eh/, /uh/, /ae/, /iy/, /aa/ و /er/



شکل ۶-۱۹ پوش طیفی آواهای مختلف بدست آمده از روش کپستروم بهبود یافته



شکل ۶-۲۰ دامنه تبدیل مقیاس مشتق مرتبه سوم پوش طیفی آواهای شکل ۶-۱۹



شکل ۶-۲۱ بلوک دیاگرام خلاصه شده الگوریتم پیشنهادی برای استخراج ویژگیها و سپس دسته بندی آواها

توجه به این نکته حائز اهمیت است که پارامترهای سیگنال صوتی پایستار و مقاوم باشند. امروزه، بسیاری از مدل‌های بکار رفته برای سیگنالها در محیط‌های نویزی قابل استفاده می‌باشند. برای اینکه پارامترهای استخراج شده قابل استفاده باشند، بایستی در مقابل تغییرات کانالهای ارتباطی، گوینده و مبدلها ثابت باقی بمانند. برای استخراج ویژگی از سیگنالهای صوتی، روشهای متعددی استفاده می‌گردد که ما در اینجا در مورد دو روش مهم آن صحبت می‌کنیم.

۶-۳-۱- ضرایب طیف فرکانسی مل 97

ضرایب طیف فرکانسی مل یکی از معروفترین روشهای استاندارد برای استخراج ویژگیهای یک سیستم تشخیص گفتار است. این روش توسط دیویس در سال ۱۹۸۰ [۵] ارائه گردید و از آن زمان به بعد بعنوان یک روش پرکاربرد باقی ماند. دیویس اعتقاد داشت که این ویژگیها نسبت به روشهای پارامتربندی دیگر، باعث افزایش فوق العاده کارایی می‌گردند که علت این امر اهمیت بیشتر دادن به اطلاعات دریافتی از طیف صوتی زمان کوتاه می‌باشد.

پروسه محاسبه MFCC را می‌توان بصورت یافتن تبدیل کسینوسی ترم حقیقی لگاریتم طیف توان زمان کوتاه در مقیاس فرکانس انتقال یافته مل توصیف نمود. جزئیات بیشتر در ادامه آورده شده است، همچنین بلوک دیاگرام این روش در شکل ۶-۲۲ نشان داده شده است.

- تحلیل طیفی تبدیل فوریه گسسته زمان کوتاه برای هر فریم صوتی از رابطه زیر محاسبه می‌گردد:

$$X(k) = \sum_{n=0}^{N-1} \omega(n) x(n) e^{-j2\pi nk/N} \quad k = 0, 1, \dots, N-1 \quad (۶-۱۳)$$

که در رابطه فوق k مولفه فرکانسی است و $f(k) = kf_s / N$ و $\omega(n)$ پنجره زمانی است. معمولا برای انتخاب پنجره از پنجره‌های معروفی همچون همینگ، هنینگ یا گوسی استفاده می‌شود.

97. Mel Frequency Cepstral Coefficients (MFCC)

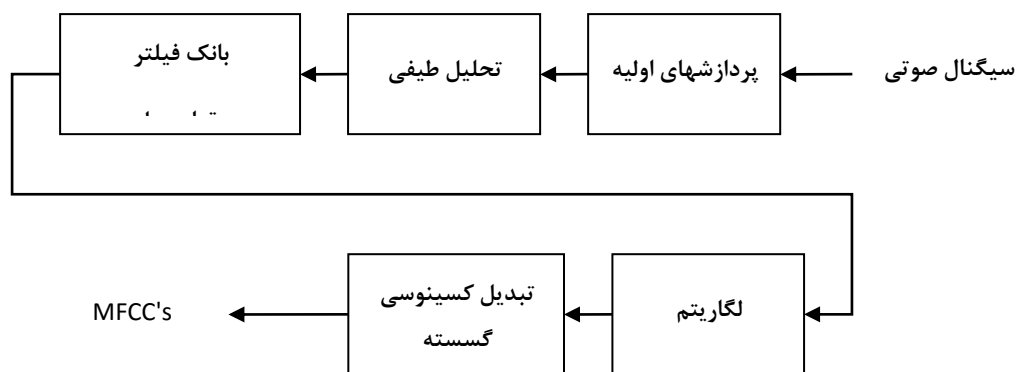
• **بانک فیلتر مقیاس مل** تحلیل بانک فیلتر مل بوسیله کوانتیزه نمودن غیر یکنواخت طیف فوریه بدست می‌آید. برای بدست آوردن کوانتیزه غیر یکنواخت، تابع پنجره بصورت غیر یکنواخت بر روی مقیاس مل جابجا می‌گردد و سپس به مقیاس هرتز انتقال می‌یابد (شکل ۶-۲۳). با انتقال توابع پنجره به مقیاس هرتز، در طیف توان فوریه ضرب می‌گردند و بعد از آن با یکدیگر جمع می‌گردند تا در نهایت ضرایب بانک فیلتر طیف مل را در اختیار ما قرار دهند. این ضرایب از رابطه زیر بدست می‌آید:

$$S(m) = \log \left[\sum_{k=0}^{N-1} |X(k)|^2 H(k, m) \right] \quad m = 1, 2, \dots, M \quad (6-14)$$

که در رابطه فوق M تعداد بانکهای فیلتر است و $M \ll N$. $H(k, m)$ بانک فیلتر مل است که مجموعه‌ای از فیلترهای مثلثی است که بوسیله فرکانسهای مرکزی $f_c(m)$ و بصورت زیر تعریف می‌گردد:

$$H(k, m) = \begin{cases} 0 & \text{for } f(k) < f_c(m-1) \\ \frac{f(k) - f_c(m-1)}{f_c(m) - f_c(m-1)} & \text{for } f_c(m-1) \leq f(k) \leq f_c(m) \\ \frac{f(k) - f_c(m+1)}{f_c(m) - f_c(m+1)} & \text{for } f_c(m) \leq f(k) < f_c(m+1) \\ 0 & \text{for } f(k) \geq f_c(m+1) \end{cases}$$

(6-15)



شکل ۶-۲۲ بلوک دیاگرام پروسه محاسبه پارامترهای MFCC

فرکانس مرکزی بانک فیلتر از رابطه زیر بدست می‌آید:

$$\phi = 2595 \log_{10} \left(\frac{f}{700} + 1 \right) \quad (۶-۱۶)$$

رابطه فوق به ازای تمام فرکانسها غیرخطی است. لگاریتم رابطه (۶-۱۴) چندین تابع را ایجاد می‌کند.

از مزایای الگوریتم فوق به موارد زیر می‌توان اشاره نمود:

- فشرده نمودن محدوده دینامیکی
- محدوده دینامیکی فضای ویژگیها فشرده می‌گردد. این امر باعث می‌گردد که کارایی آن در سیستمهای کامپیوتری بیشتر گردد.
- ویژگیهای دریافتی بسیار مهم
- سیستم شنوایی انسان از توابع لگاریتمی پیروی می‌کند مانند قاعده فشرده سازی
- هیستوگرام شبه گوسی
- هیستوگرامهای شبه گوسی را می‌توان با دقت بیشتری مدلسازی نمود و در صورتیکه از مدل مختلط گوسی⁹⁸ استفاده گردد پیچیدگی آن کمتر خواهد شد.

⁹⁸ . Gaussian Mixture Model (GMM)

همچنین این الگوریتم دارای معایبی نیز می‌باشد. لگاریتم مقادیر کوچک به سمت منفی بینهایت میل می‌کند. در یک طیف صوتی، کانال بین فرکانسهای رزونانس دارای توان کمی است و معمولاً دارای بدترین نسبت سیگنال به نویز می‌باشد. این مشکل را می‌توان با استفاده از روشهای مختلفی مانند اضافه نمودن یک ترم ثابت به تمام مقادیر و یا استفاده از تابع ریشه درجه سوم بجای لگاریتم حل نمود. با آزمایشات انجام شده به این نتیجه رسیدند که روش ریشه درجه سوم نتایج ضعیفتری را در اختیار ما قرار می‌دهد [۶].

بانک فیلتر مل فیلترهایی هستند که با فواصل خطی در فرکانسهای پایین و کمتر از ۱ کیلو هرتز و بصورت تقریباً لگاریتمی در فرکانسهای بالاتر قرار گرفته‌اند. این نوع نحوه قرارگیری اجازه می‌دهد بانک فیلتر ویژگیهای صدایی بسیار مهم را از سیگنال بیرون بکشد در عین اینکه تغییرات طیفی باندهای فرکانسی بالاتر را خنثی می‌سازد. تعداد فیلترهای موجود در بانک فیلتر بسته به نوع کاربرد متغیر است، اما معمولاً برای پهنای باند ۴ کیلو هرتز ۱۹ فیلتر و برای پهنای باند ۸ کیلو هرتز ۲۴ فیلتر استفاده می‌گردد. با توجه به روشی که برای محاسبه MFCC مورد استفاده قرار می‌گیرد، بانکهای فیلتری متفاوتی بکار می‌رود که نمونه‌هایی از آن در شکل ۶-۲۳ آورده شده است. به خصوصیتهای مختلف بانکهای فیلتری دقت گردد. در اجرای دیویس فاصله‌گذاری تا فرکانس ۱ کیلو هرتز خطی است و سپس فاصله‌گذاری لگاریتمی می‌گردد ولی در اینحالت دامنه فیلتر ثابت است. روش HTK⁹⁹ [۷] دارای فاصله‌گذاری لگاریتمی و دامنه ثابت است.

- **تبدیل کسینوسی گسسته¹⁰⁰ هنگامیکه لگاریتم انرژیهای بانک فیلتر محاسبه گردید، MFCC** های نهایی با محاسبه تبدیل کسینوسی گسسته لگاریتم انرژیهای بانک فیلتر بدست می‌آید. در سیستمهای تشخیص گفتار مرسوم است که فقط ضرایب $c(1)$ تا $c(12)$ را نگه دارند.

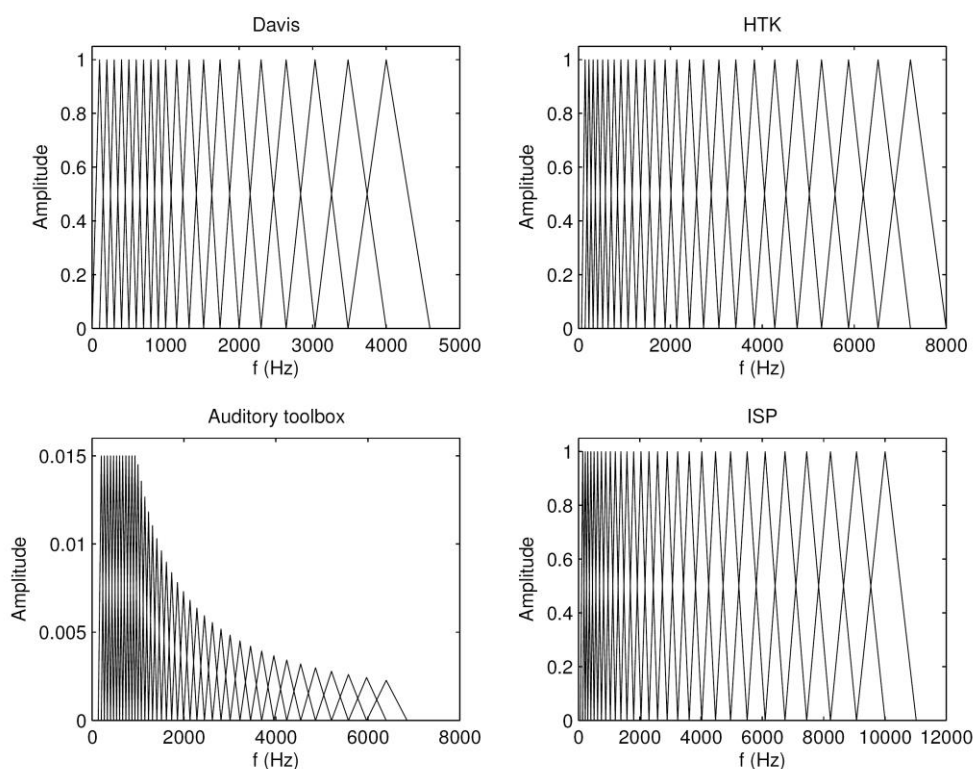
$$c(n) = \sum_{m=0}^{M-1} S(m) \cos\left(\frac{\pi n}{M} \left(\frac{m-1}{2}\right)\right), \quad 0 \leq n \leq M \quad (۶-۱۷)$$

⁹⁹ . hidden Markov model toolkit

¹⁰⁰ . Discrete Cosine Transform (DCT)

ضرایب طیف فرکانسی مل، طیف فشرده شده سیگنال صوتی را در اختیار ما قرار می‌دهد. دیویس نشان داد که تقریباً ۶ تا از ضرایب دارای بیشترین اطلاعات در مورد طیف صوتی می‌باشند. او بیان کرد که وارد نمودن ضرایب بیشتر بستگی به گوینده دارد [۶].

یکی از مشکلات MFCC مقاوم نبودن ضرایب در حضور نویز است، به همین علت عده‌ای از محققین اصلاحاتی را برای مقاوم سازی آن در برابر نویز ایجاد نموده‌اند. بعنوان نمونه، افزایش دامنه‌های لگاریتم - مل توسط مرتبه‌های مناسب (مانند ۲ یا ۳) که قبل از محاسبه تبدیل کسینوسی گسسته انجام می‌گیرد، اثر مولفه‌هایی با انرژی کمتر را کاهش می‌دهد [۸].



شکل ۶-۲۳ چهار حالت مختلف برای بانک فیلتر مل. به مقیاسهای مختلف محورهای فرکانسی هر نمودار دقت شود.

۶-۳-۳-۲- محاسبات خطی ادراکی¹⁰¹

روش PLP در ابتدا بوسیله هرمانسکی در سال ۱۹۸۹ ارائه گردید [۱۰]. این روش بر اساس روش محاسبات خطی ویژگیهایی را تولید می کند که بیشترین شباهت را با ویژگیهای شنیداری انسان دارند. برای آنکه ویژگیهای فوق با سیستم شنوایی انسان مطابقت داشته باشد، تحلیل طیفی باید در مقیاس فرکانسی انتقال یافته یا بصورتیکه ناحیه فرکانسی مورد نظر دارای حساسیت فرکانسی بیشتر نسبت به بقیه نواحی باشد، انجام گیرد. در محدوده ۱ کیلو هرتز، رزولوشن طیفی سیستم شنوایی انسان با فرکانس کاهش می یابد. همچنین، سیستم شنوایی انسان دارای حساسیت بیشتری در بخش میانی طیف فرکانسی است. مراحل اولیه روش پیشگویی خطی ادراکی سیگنال صوتی قبل از روش محاسبه خطی بکار می رود و سعی دارد خصوصیات ادراکی (دریافتی) سیستم شنوایی انسان را بدین وسیله تخمین بزند [۱۰]. یکی از مزایای روش PLP وارد کردن خصوصیت‌های فیزیولوژیکی سیستم شنوایی گوش انسان در تحلیل است. بلوک دیاگرام شکل ۶-۲۴ مراحل اساسی این الگوریتم را نشان می دهد.

در روش PLP با استفاده از ترکیب چندین روش در زمینه‌های مختلف مهندسی ساختار فیزیولوژیکی سیستم شنوایی گوش انسان تقریب زده می شود. الگوریتم PLP را می توان در ۵ بخش خلاصه نمود:

۱- آنالیز طیفی

بعد از یک سری پردازشهای اولیه (مانند نرمالیزه نمودن، پنجره بندی، عبور از یک فیلتر پایین گذر و حذف ترم DC آن) توان طیفی زمان کوتاه برای هر فریم محاسبه می شود.

۲- آنالیز طیفی باند بحرانی¹⁰²

¹⁰¹ . Perceptual Linear Prediction (PLP)

¹⁰² . Critical Band Analysis

آنالیزهای باند بحرانی با استفاده از انتقال فرکانس که بر اساس ساختارهای صوتی انجام می‌گیرد، باعث پر اهمیت شدن یک سری از فرکانسها می‌گردد. در این قسمت توان طیفی از حوزه فرکانسی انتقال می‌یابد تا با مقیاس بارک¹⁰³ هماهنگ گردد. تابع انتقال مقیاس بارک از رابطه زیر بدست می‌آید:

$$F_{Bark} = 6 \log \left[\frac{f}{600} + \sqrt{\left(\frac{f}{600}\right)^2 + 1} \right] \quad (۱۸-۶)$$

در آنالیز بانک فیلتر PLP شکل پنجره بر اساس شبیه‌سازی منحنی‌های ماسک‌گذاری قسمت بحرانی طراحی می‌گردد که نمونه‌ای از آن در شکل ۶-۲۵ نشان داده شده است. همانگونه که در این شکل نشان داده شده است، تعداد فیلترها در فرکانسهای پایینی بیشتر است. علت این امر حساس نمودن بیشتر الگوریتم نسبت به فرکانسهای پایینی و برگرفته از سیستم شنوایی انسان است. در ادامه منحنی ماسک‌گذاری باند بحرانی با توان طیفی کانالو می‌گردد تا اثر ماسک‌گذاری فرکانسی شبیه‌سازی گردد.

۳- بلندی برابر / پیش تاکید¹⁰⁴

برای جبران نمودن و یا عبارت دیگر خنثی نمودن حساسیت ناموزون سیستم شنوایی انسان در فرکانسهای مختلف، در این مرحله منحنی‌هایی با میزان بلندی یکسان شبیه‌سازی می‌گردد. این منحنی‌ها از رابطه زیر بدست می‌آیند:

$$E(\omega) = \frac{(\omega^2 + 56.8 \times 10^6) \omega^4}{(\omega^2 + 6.3 \times 10^6)^2 (\omega^2 + 0.38 \times 10^9)} \quad (۱۹-۶)$$

این تابع میزان حساسیت شنوایی گوش انسان را در ۴۰ دبی تخمین می‌زند.

۴- رابطه توانی شدت بلندی¹⁰⁵

در این مرحله رابطه غیرخطی بین شدت صدا و میزان بلندی شنیده شده مدلسازی می‌گردد. برای انجام اینکار ریشه سوم انرژیهای باند بحرانی با استفاده از رابطه زیر محاسبه می‌گردد:

¹⁰³ . Bark Scale

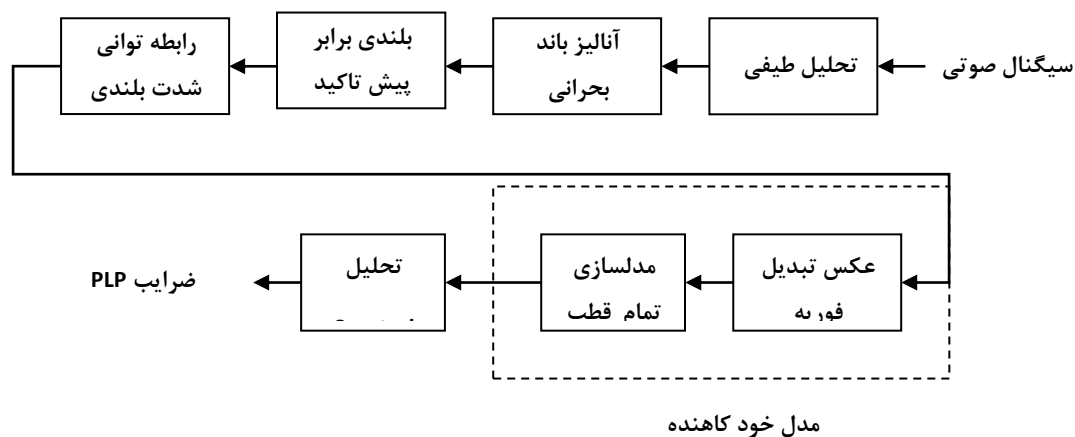
¹⁰⁴ . Equal Loudness / Pre-emphasis

¹⁰⁵ . Intensity-loudness power law

$$l(\omega) = \sqrt[3]{E(\omega)} \quad (۲۰-۶)$$

۵- مدل خود کاهنده¹⁰⁶

تبدیل معکوس فوریه برای تولید یک دنباله خود همبسته بکار می‌رود. پس از آن با جایگزاری مقادیر اتوکورلیشن مصنوعی در معادلات Yule - Walker مدل‌سازی تمام قطب صورت می‌پذیرد. وقتیکه ضرایب AR محاسبه گردید، آنها را به ضرایب طیفی تبدیل می‌شوند که



شکل ۶-۲۴ بلوک دیاگرام مراحل الگوریتم PLP

این عمل تبدیل را پروسه ضرایب طیفی محاسبات خطی¹⁰⁷ نامند. الگوریتم LPCC با محاسبه پارامترهای مدل خود کاهنده (a_k) برای هر فریم صوتی با در نظر گرفتن یک مدل تمام قطب شروع می‌گردد. فرض کنید مدل تمام قطب داده شده مشابه رابطه زیر باشد:

$$\hat{S}(z) = \frac{A}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (۲۱-۶)$$

¹⁰⁶ . Auto Regressive modeling (AR modeling)

¹⁰⁷ . Linear Prediction Cepstral Coefficient (LPCC)

با در نظر گرفتن اینکه در رابطه فوق هیچ پارامتر مجهولی وجود نداشته باشد، ضرایب طیفی با استفاده از تکرارهای زیر بدست می‌آید:

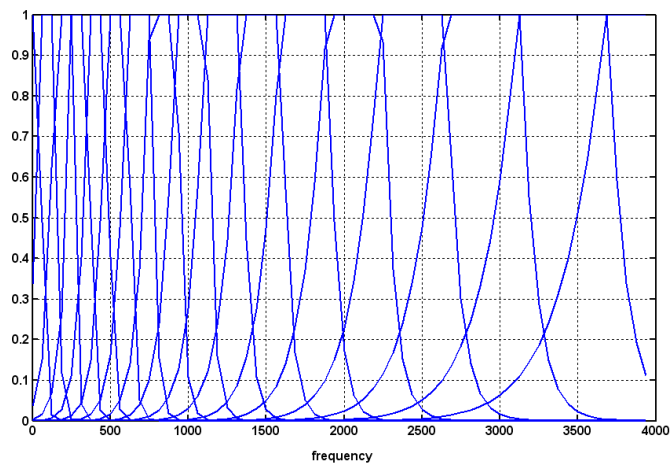
$$c(n) = \begin{cases} 0 & n < 0 \\ \log(A) & n = 0 \\ a_n + \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) c(k) a_{n-k} & 0 < n \leq p \\ \sum_{k=n-p}^{n-1} \left(\frac{k}{n}\right) c(k) a_{n-k} & n > p \end{cases} \quad (۲۲-۶)$$

با توجه به اینکه دستگاه فوق ممکن است بینهایت جواب داشته باشد، می‌توان از آن تعداد جوابهای مشخصی را بیرون کشید. معمولاً تعداد ضرایبی که کارایی مناسبی را ایجاد می‌کنند، بین ۱۲ تا ۲۰ ضریب می‌باشد.

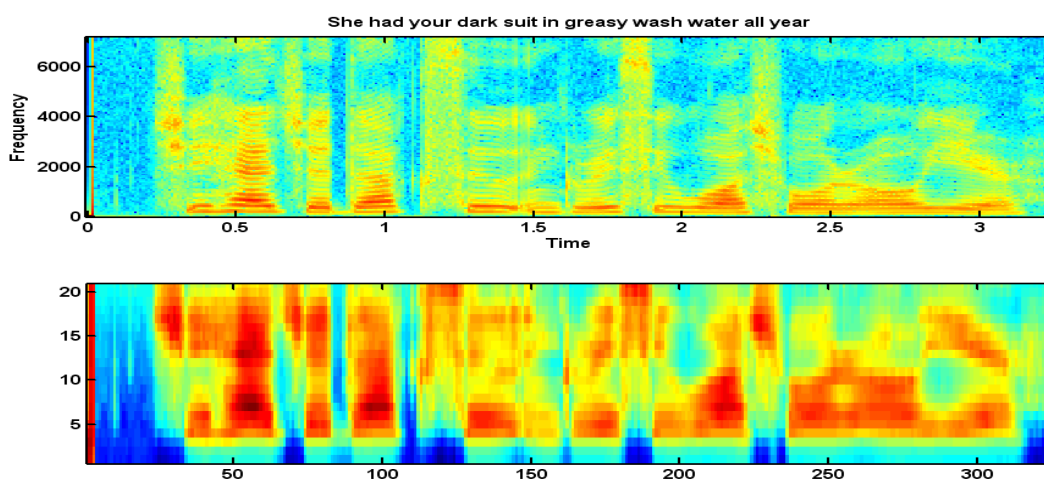
هرمانسکی اثبات نمود که مزیت اجرای الگوریتم PLP با استفاده از روش LPCC این است که با انتخاب مرتبه مناسب برای مدل می‌توان اطلاعات وابسته به گوینده را حذف نمود. با این وجود، در حالت‌های یکسان روش PLP حساسیت بیشتری به فرکانس اصلی نسبت به روش LP دارد. هرمانسکی معتقد است که این حساسیت با قوه شنوایی انسان مطابقت دارد، از اینرو او آزمایشاتی را انجام داد که مشخص می‌کرد، پیکهای متناظر با فرکانسهای رزونانس به سمت نزدیکترین پیک هارمونیک تمایل دارند [۱۰].

شکل ۶-۲۶ مقایسه بین دو اسپکتروگرام را نشان می‌دهد. شکل بالایی اسپکتروگرام جمله "She had your dark suit in greasy wash water all year" است که از مجموعه داده TIMIT استخراج شده و با استفاده از روشهای معمول بدست آمده است. شکل پایینی اسپکتروگرام محاسبه شده را با استفاده از بازیابی لگاریتم طیف از ضرایب PLP نشان می‌دهد. با مقایسه این دو شکل به این نتیجه می‌رسیم که الگوریتم PLP اطلاعات مربوط به فرکانسهای رزونانس را، بدون در نظر گرفتن ساختار هارمونیک آن، که بعلت سیگنال تحریک است، حفظ می‌کند.

عده‌ای بر این عقیده‌اند که کارایی روش PLP قابل مقایسه با روش MFCC است، ولی در صورت تغییر تعداد ضرایب و فیلترهای مورد استفاده در پارامتر بندی، این روش دارای نتایج پایدارتری می‌باشد [۱۱].



شکل ۶-۲۵ مجموعه‌ای از بارک‌ها به ازای فرکانس نمونه‌برداری ۸۰۰۰ و تعداد فیلتر ۱۷ با $NFFT = 128$.



شکل ۶-۲۶ اسپکتروگرام سیگنال اصلی (بالا) و اسپکتروگرام بازیابی شده با استفاده از ضرایب PLP

۶-۳-۴- نتایج

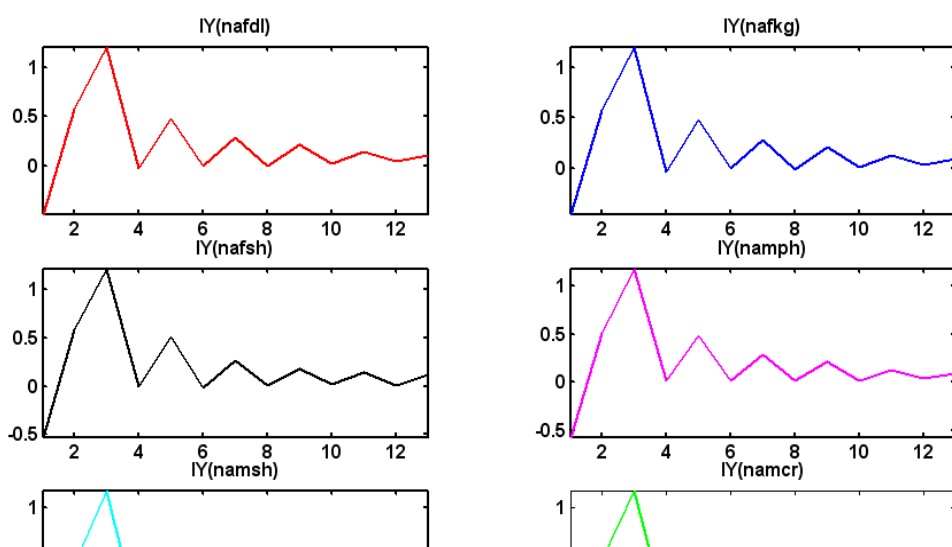
با توجه به توضیحات ارائه شده در بخش‌های قبلی و همچنین الگوریتم نشان داده شده در شکل ۶-۲۱ در قسمت استخراج ویژگی از روش‌های MFCC و PLP استفاده گردید. بعنوان نمونه و برای مقایسه کارایی این

دو روش، در روش MFCC تعداد ۱۳ ضریب انتخاب گردید و در روش PLP مرتبه مدل را با توجه به توضیحات بخش قبلی، ۱۲ در نظر گرفتیم. بنابراین در اینحالت ما توانستیم هر آوا را با یک بردار ویژگی ۱۳ تایی نمایش دهیم بگونه‌ای که آواهای یکسان ادا شده توسط افراد مختلف دارای بردار ویژگی یکسان و آواهای مختلف دارای بردار ویژگی متفاوتی باشند. برای درک بهتر موضوع در شکل ۶-۲۷، ۱۳ ضریب ابتدایی استخراج شده با استفاده از روش MFCC و PLP برای آوای /IY/ که بوسیله افراد مختلف ادا شده است و در شکل ۶-۲۸، همین آزمایش برای آواهای مختلف انجام شده است. این آواها از [۴] استخراج شده‌اند.

همانطور که در شکل‌های ۶-۲۷ و ۲۸ نشان داده شده است، آواهای یکسان ادا شده توسط افراد مختلف دارای بردار ویژگی یکسان و آواهای مختلف دارای بردار ویژگی متفاوتی می‌باشند. البته نکته‌ای که در اینجا لازم است به آن اشاره کنیم، این است که در مقایسه با ضرایب بدست آمده از روش PLP با روش MFCC، ضرایب PLP مقادیر کوچکتری می‌باشند و اختلاف ضرایب دوم تا سیزدهم در اینحالت نسبت به ضریب اول زیاد می‌باشد، بهمین علت برای درک و مشاهده بهتر اختلافات در حالت PLP از وارد نمودن ضریب اول خودداری نموده‌ایم، ولی در محاسبات خود اثر آنرا در نظر گرفته‌ایم.

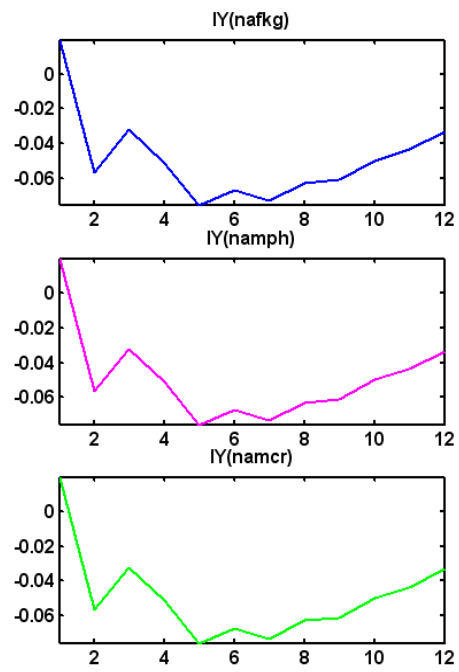
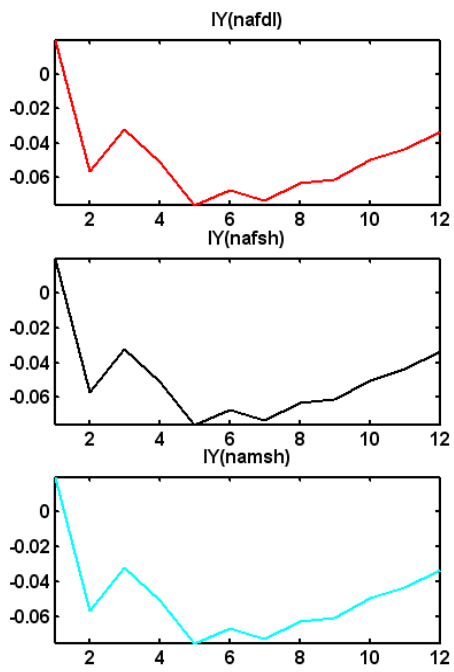
همانطور که می‌دانید، ما از روش‌های فوق برای استخراج ویژگی از منحنی‌های خروجی تبدیل مقیاس استفاده نموده‌ایم. در واقع ورودیهای بلوک دیاگرام بخش استخراج ویژگی، همانطور که در شکل ۶-۲۱ نشان داده شد، دامنه تبدیل مقیاس پوشهای طیفی حاصل از روش کپستروم بهبود یافته سیگنالهای صوتی است. در نتیجه با توجه به توضیحات فوق می‌توان از الگوریتم تبدیل مقیاس به همراه یکی از روشهای PLP یا MFCC جهت استخراج ویژگی از آواها استفاده کرد.

یکی از مزایای بسیار مهم این الگوریتم نسبت به الگوریتمهای مشابه دیگری که برای شناسایی آواها از یکدیگر استفاده می‌گردند، کوچکتر نمودن فضای ویژگیهاست که این امر باعث افزایش سرعت



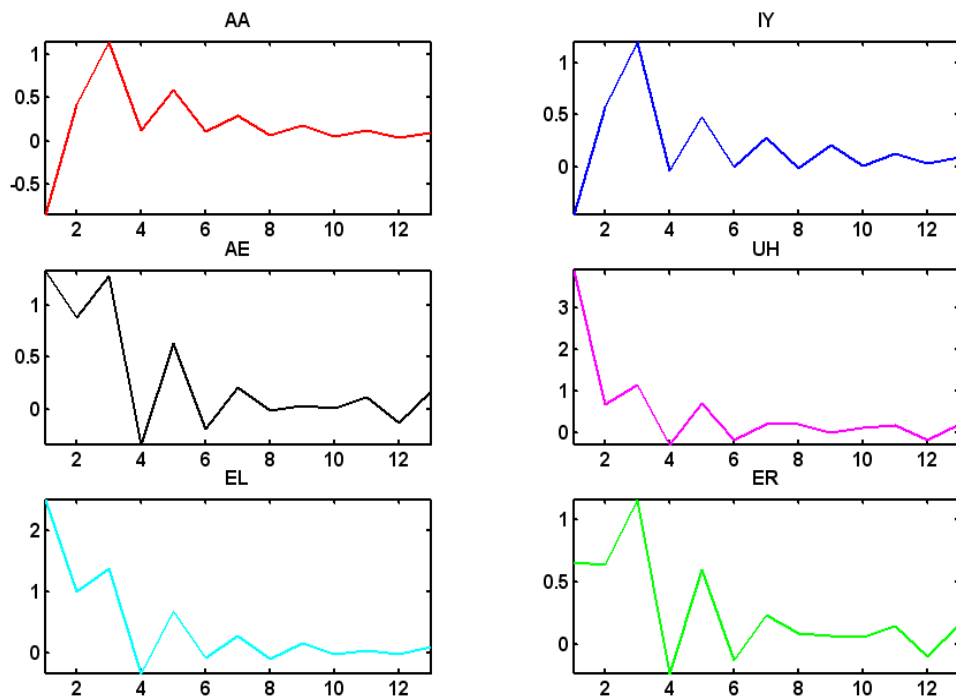
)

(الف)

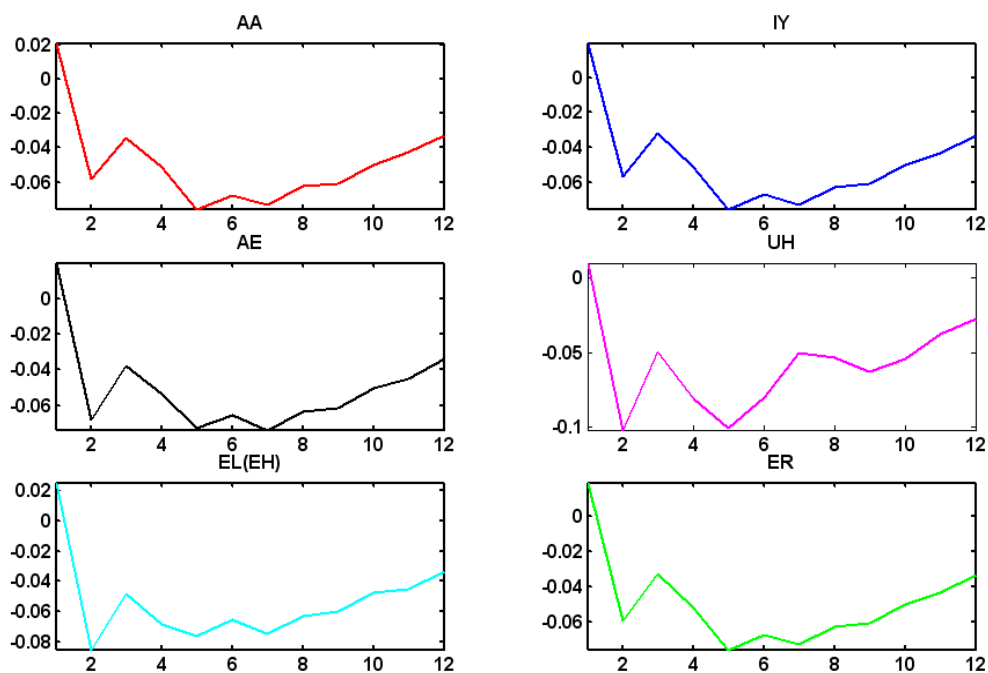


(ب)

شکل ۶- ۲۷ بردار ویژگی آوای /IY/ با استفاده از الگوریتم پیشنهادی به همراه (الف) روش MFCC (ب) روش PLP



(الف)



(ب)

شکل ۶- ۲۸ بردار ویژگی آواهای مختلف با استفاده از الگوریتم پیشنهادی به همراه (الف) روش MFCC (ب) روش PLP

الگوریتم برای جدا نمودن آواها از یکدیگر می‌گردد. بعنوان نمونه، در آزمایشات انجام گرفته سیگنالهای ورودی مورد استفاده دارای ۷۵۰۰ نمونه می‌باشند. با استفاده از روش کپستروم بهبود یافته تعداد نمونه‌ها یا ویژگیهای استخراج شده به ۳۵۰۰ نمونه کاهش می‌یابد. در ادامه با استفاده از تبدیل مقیاس تعداد ویژگیها به ۳۰۰ و در نهایت با استفاده از الگوریتم MFCC یا PLP به ۱۳ ویژگی کاهش می‌یابد. برای مقایسه بهتر، تعداد ویژگیهای استخراج شده در هر مرحله از الگوریتم پیشنهادی در جدول ۶- ۱ آورده شده است.

بنابراین با استفاده از این روش ما فشرده سازی اطلاعاتی (کوچکتر کردن فضای ویژگی) داشته‌ایم. یکی از کاربردهای بسیار مهمی که برای این روش می‌توان در نظر گرفت، کد نمودن اطلاعات است که با استفاده از روش فوق هم فضای ذخیره سازی یا عبارت دیگر جدول مراجعه کاهش می‌یابد و هم سرعت ارسال اطلاعات بیشتر خواهد شد.

در ادامه قصد داریم، کارایی ویژگیهای استخراج شده را برای دسته‌بندی آواها از یکدیگر مورد بررسی قرار دهیم.

جدول ۶- ۱ تعداد نمونه های استخراج شده در هر مرحله از الگوریتم پیشنهادی

سیگنال یا منحنی ورودی	تعداد ویژگیهای استخراج شده
پردازشهای اولیه	۷۵۰۰
کپستروم بهبود یافته	۳۵۰۰
تبدیل مقیاس	۳۰۰
استخراج ویژگی	۱۳

۶-۳-۵- طبقه‌بندی ویژگیهای استخراج شده

هدف از اجرای این قسمت آن است که با توجه به ویژگیهای استخراج شده مشخص شود، آوای ورودی به کدام کلاس تعلق دارد. برای انجام اینکار در ابتدا با استفاده از الگوریتم پیشنهادی یک سری ویژگیهای مشخص را از سیگنال ورودی استخراج می‌کنیم و سپس آنها را با ویژگیهای مرجع مقایسه می‌کنیم. در این قسمت باید دو کار را انجام دهیم. در ابتدا بایستی بردارهای مرجع ایجاد گردند و سپس نوع روشی که برای مقایسه ویژگیهای استخراج شده با ویژگیهای مرجع بکار می‌رود، مشخص گردد. برای دسته‌بندی داده‌ها روشهای متفاوتی وجود دارد که از آن جمله می‌توان به روشهایی ساده مانند نزدیکترین همسایه، فاصله اقلیدسی¹⁰⁸، فاصله مالهونوبیس¹⁰⁹، و روشهای پیچیده‌تر اما با دقت بسیار بیشتری مانند الگوریتم ژنتیک¹¹⁰، شبکه‌های عصبی¹¹¹، فازی لاجیک¹¹²، آنالیز مولفه‌های مستقل¹¹³، مدل گوسی مختلط¹¹⁴ و روشهای بسیار دیگری اشاره نمود.

همانطور که قبلاً به آن اشاره نمودیم، مجموعه داده‌های آوایی که در اختیار ما قرار دارد، یک مجموعه محدود است. در حالیکه برای بسیاری از روشهای فوق ما نیاز به تعداد داده‌های زیاد برای بدست آوردن یک مدل آماری به ازای هر کلاس هستیم. اما با توجه به اینکه داده‌های متعلق به هر کلاس یا عبارت دیگر تعداد آواها محدود است، مدل آماری بدست آمده با این داده‌های کم نمی‌تواند دقیق باشد. از اینرو ما در اینجا از روش فاصله اقلیدسی برای کلاسه‌بندی استفاده نمودیم. واضح است که در صورت استفاده از طبقه‌بندیهای قدرتمندتر نتایج بهتری حاصل می‌گردد.

همانطور که می‌دانید در یک سیستم تشخیص الگو، یک سری از داده‌ها بعنوان داده‌های آموزشی می‌باشند که برای آموزش سیستم از آنها استفاده می‌گردد. اصولاً هر سیستم تشخیص صوتی نیاز به یادگیری دارد، تا از این طریق بتواند آموزش ببیند. برای آموزش الگوریتم پیشنهادی خود و برای بدست

¹⁰⁸ . Euclidean distance

¹⁰⁹ . Mahalanobis distance

¹¹⁰ . Genetic algorithm

¹¹¹ . Neural Network

¹¹² . fuzzy logic

¹¹³ . Independent Component Analysis (ICA)

¹¹⁴ . Gaussian Mixture Model (GMM)

آوردن داده‌های مرجع از هر کلاس آوایی را بعنوان داده آموزش وارد سیستم می‌کنیم و بردار ویژگی متناظر با آنرا بدست می‌آوریم و آنرا بعنوان داده استاندارد یا مرجع آن کلاس در نظر می‌گیریم.

به این ترتیب با توجه به اینکه هدف ما دسته‌بندی آواهای متفاوت است، در مجموع به ازای هر کلاس یک بردار مرجع بدست می‌آید. حال برای تست سیستم بایستی داده‌های تست وارد سیستم گردند. داده‌های تست همانطور که از اسم آن بر می‌آید برای تست سیستم استفاده می‌گردند و بهتر است، داده‌های تست با داده‌های آموزشی تفاوت داشته باشند. با اینکار دقت تشخیص سیستم بهتر مشخص می‌گردد.

با ورود داده تست که در اینجا یک سیگنال آوا است که متعلق به یکی از ۵ کلاس آوایی مورد نظر است، بردار ویژگی از آن استخراج می‌گردد. حال این بردار بایستی با بردار مرجع مقایسه گردد. در روش فاصله اقلیدسی فاصله برداری بین بردار ویژگی داده ورودی با هر یک از بردارهای مرجع اندازه‌گیری می‌گردد و سپس داده (آوا) به کلاسی تعلق می‌یابد که کمترین فاصله را با بردار مرجع آن کلاس داشته باشد. البته با آزمایشاتی که انجام دادیم به این نتیجه رسیدیم که یک سری از ویژگیهای موجود در بردار ویژگی دارای ارزش بیشتری نسبت به بقیه داده‌ها می‌باشند و با انجام اینکار دقت تشخیص سیستم بالاتر نیز خواهد رفت. در واقع این بردار را که بردار وزندار کردن می‌نامیم با ضرب شدن در بردار ویژگی استخراج شده، ارزش یک سری از ویژگیها را بیشتر می‌کند. ما به کمک این کار حتی توانستیم فضای ویژگی را به ۹ ویژگی کاهش دهیم.

برای آخرین مرحله استخراج ویژگی از دو روش MFCC و PLP استفاده نمودیم که نتایج تشخیص بترتیب در جدولهای ۶-۲ و ۶-۳ آمده است. مجموعه داده ما شامل ۵ کلاس آوایی مختلف است که هر کلاس شامل ۱۲ گوینده مختلف است و هر گوینده هر آوا را چهار بار تکرار می‌کند، پس در مجموع ۴۸ آوا برای هر کلاس داریم و در کل ۲۴۰ نمونه آوایی داریم. ما از این مجموعه تنها ۶ داده را بعنوان داده‌های آموزشی وارد الگوریتم نمودیم و بردارهای مرجع متعلق به هر کلاس را تشکیل دادیم. همانطور که در جدول ۶-۲ نشان داده شده است، در صورت استفاده از الگوریتم پیشنهادی به‌مراه MFCC برای استخراج ویژگی، سه کلاس از پنج کلاس دارای دقت تشخیص ۱۰۰ می‌گردند و کمترین نرخ تشخیص متعلق به کلاس آوایی /AA/ است که حدود ۸۲ درصد است. اما با مقایسه نتایج بدست آمده از ترکیب الگوریتم پیشنهادی با روش MFCC و یا روش PLP، همانطور که در جدول ۶-۳ آورده شده است، روش PLP دارای دقت تشخیص بیشتری نسبت به

روش MFCC است و در اینحالت چهار کلاس آوایی مختلف دارای دقت تشخیص ۱۰۰ درصد و فقط کلاس آوایی /AA/ دارای دقت تشخیص ۸۷,۵ می‌گردد.

همانطور که قبلا اشاره نمودیم، ما با استفاده از بردار وزندار، به یک سری از ویژگیها ارزش بیشتری نسبت به بقیه آنها دادیم. در این بردار در واقع ارزشی که به ویژگیهای ۱ تا ۶ داده می‌شود بیشتر از ارزش ویژگیهای ۷ تا ۱۳ است. با آزمایشاتی که برای مجموعه داده‌های مختلف انجام دادیم به این نتیجه رسیدیم که مقدار ارزش دهی این ضرایب ربطی به مجموعه داده ندارد. برای اطلاعات بیشتر می‌توانید به [۱۲و۱۳] مراجعه نمایید.

در آزمایش بعدی بر تعداد کلاسها افزودیم. در اینحالت تعداد کلاسها برابر با ۸ می‌باشد. همانگونه که در جداول ۴-۶ و ۵-۶ نشان داده شده است، در صورت تلفیق الگوریتم پیشنهادی به همراه روش MFCC دقت تشخیص نهایی سیستم برابر با ۸۸ درصد می‌گردد و در صورتیکه الگوریتم پیشنهادی را با روش PLP تلفیق کنیم دقت تشخیص نهایی سیستم ۸۹,۵ درصد می‌گردد. در قسمت بعد قصد داریم تا الگوریتم پیشنهادی را با روشهای مشابه استخراج ویژگی برای تشخیص آواها مقایسه نماییم.

جدول ۶-۲ نتایج آزمایشات انجام شده با استفاده از الگوریتم پیشنهادی به همراه روش MFCC برای ۵ کلاس مختلف

نتیجه تست		/AO/	/AE/	/AA/	/UH/	/IY/
	/IY/	۰	۰	۰	۰	۴۸
	/UH/	۰	۰	۰	۴۴	۴
	/AA/	۰	۸	۴۰	۰	۰
	/AE/	۰	۴۸	۰	۰	۰
	/AO/	۴۸	۰	۰	۰	۰
	دقت تشخیص	۱۰۰	۱۰۰	۸۲	۹۰	۱۰۰

جدول ۶-۳ نتایج آزمایشات انجام شده با استفاده از الگوریتم پیشنهادی به همراه روش PLP برای ۵ کلاس مختلف

					نتیجه تست
•	•	•	•	۴۸	/IY/
•	•	•	۴۸	•	/UH/
•	۶	۴۲	•	•	/AA/
•	۴۸	•	•	•	/AE/
۴۸	•	•	•	•	/AO/
۱۰۰	۱۰۰	۸۷,۵	۱۰۰	۱۰۰	دقت تشخیص

جدول ۴-۶ نتایج آزمایشات انجام شده با استفاده از الگوریتم پیشنهادی به همراه روش MFCC برای ۸ کلاس مختلف

/ER/	/EH/	/UW/	/AO/	/AE/	/AA/	/UH/	/IY/	
•	•	•	•	•	•	•	۴۸	/IY/
•	۴	•	•	•	•	۴۰	۴	/UH/
•	•	•	•	۸	۴۰	•	•	/AA/
•	۸	•	•	۴۰	•	•	•	/AE/
۴	۴	•	۴۰	•	•	•	•	/AO/
۴	•	۴۴	•	•	•	•	•	/UW/
•	۴۴	•	۴	•	•	•	•	/EH/
۴۰	۴	•	۴	•	•	•	•	/ER/
۸۳,۳	۹۱,۶	۹۱,۶	۸۳,۳	۸۳,۳	۸۳,۳	۸۳,۳	۱۰۰	دقت تشخیص

جدول ۵-۶ نتایج آزمایشات انجام شده با استفاده از الگوریتم پیشنهادی به همراه روش PLP برای ۸ کلاس مختلف

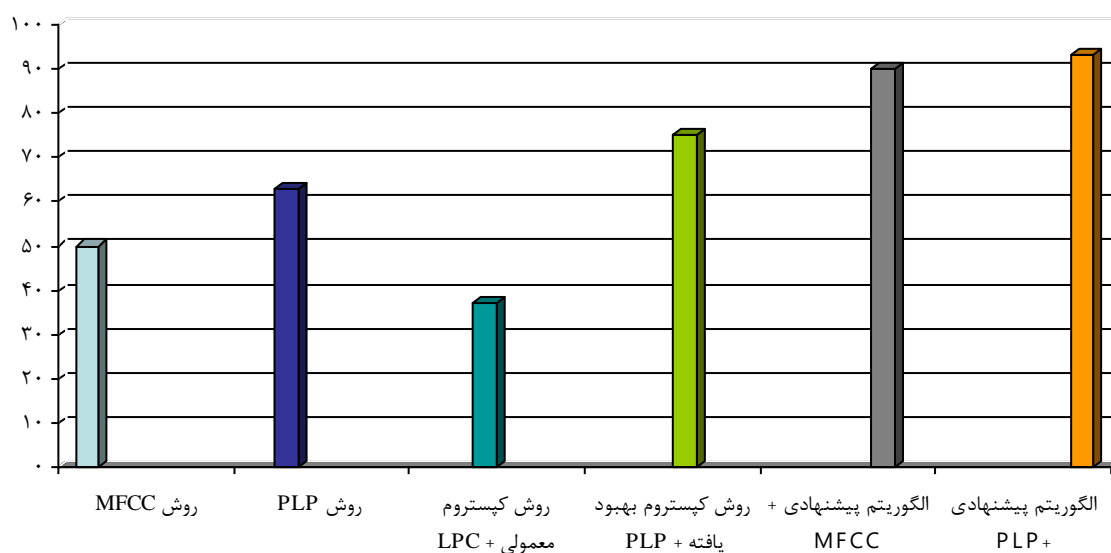
/ER/	/EH/	/UW/	/AO/	/AE/	/AA/	/UH/	/IY/	
۰	۰	۰	۰	۰	۰	۰	۴۸	/IY/
۰	۰	۴	۰	۰	۰	۴۴	۰	/UH/
۰	۰	۰	۰	۶	۴۲	۰	۰	/AA/
۰	۴	۰	۰	۴۴	۰	۰	۰	/AE/
۴	۰	۰	۴۴	۰	۰	۰	۰	/AO/
۴	۰	۴۴	۰	۰	۰	۰	۰	/UW/
۰	۴۴	۴	۰	۰	۰	۰	۰	/EH/
۴۰	۰	۰	۰	۰	۰	۸	۰	/ER/
۸۳,۳	۹۱,۶	۹۱,۶	۹۱,۶	۹۱,۶	۸۷,۵	۸۷,۵	۱۰۰	دقت تشخیص

۴-۶- مقایسه‌ای بین روشهای مختلف استخراج ویژگی

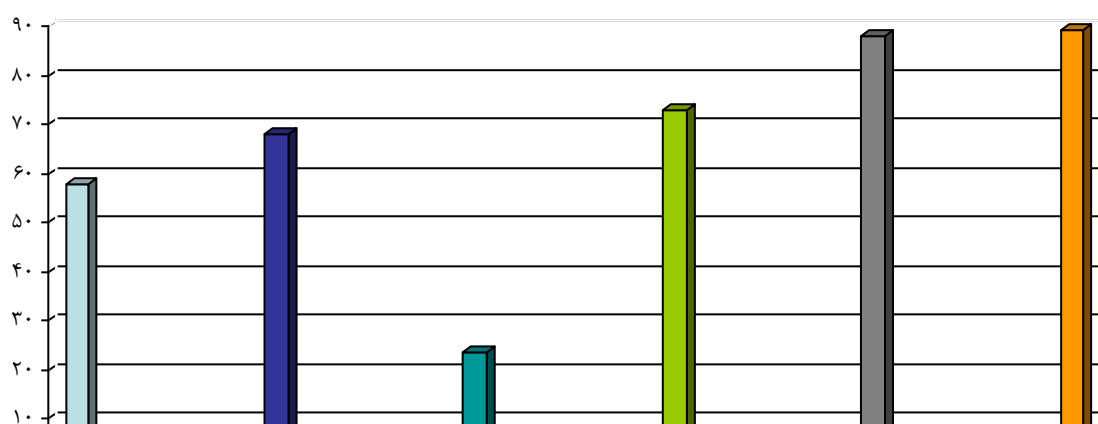
در این بخش جهت نشان دادن کارایی الگوریتم پیشنهادی با روشهای متداول استخراج ویژگی مقایسه‌ای بین این روش با روشهای معروفی مثل MFCC، PLP، کپستروم و PLP و LPC انجام گرفته است. آزمایشات انجام شده برای دو حالت مختلف که یکی با ۵ کلاس و دیگری ۸ کلاس است، صورت گرفته است. در نمودار شکل ۶-۲۹ نتیجه انجام چند آزمایش مختلف برای کلاسه‌بندی ۵ دسته از آواها نشان داده شده است. در تمامی آزمایشات صورت گرفته شرایط یکسانی در نظر گرفته شد. در روش اول ما از روش MFCC به تنهایی جهت تشخیص آواها استفاده نمودیم، در اینحالت دقت تشخیص نهایی سیستم ۵۰ درصد می‌شود. در صورتیکه از روش PLP برای شناسایی آواها استفاده شود، دقت تشخیص ۶۳ درصد می‌شود. اگر از روش کپستروم معمولی به‌همراه کد‌گذاری محاسبات خطی استفاده کنیم، دقت تشخیص سیستم به ازای آواهای مختلف ۳۷,۳ درصد می‌گردد. در آزمایش چهارم ما از روش کپستروم بهبود یافته به همراه PLP که دارای ویژگیهای بهتری نسبت به روش MFCC است، استفاده کردیم. نتایج در اینحالت قابل توجه بود بصورتیکه دقت تشخیص نهایی سیستم با استفاده از این روش در حدود ۷۵ درصد گردید.

همانطور که در شکل ۶-۲۹ نشان داده شده است، در اینحالت الگوریتم پیشنهادی ما در مقایسه با الگوریتمهای دیگری که برای استخراج ویژگی از آواها بکار می‌روند، دارای بیشترین دقت تشخیص است. نکته‌ای که لازم است در اینجا به آن اشاره کنم این است که ما الگوریتم فوق را بر روی مجموعه داده‌های متفاوتی آزمایش نمودیم و در آنها نیز نتایج بدست آمده تطابق زیادی با نتایج ارائه شده زیر داشت.

همچنین با در نظر گرفتن شرایط یکسان برای تمام روشهای استخراج ویژگی، برای ۸ کلاس آوایی مختلف نیز همین آزمایش را انجام دادیم، که نتایج آن در شکل ۶-۳۰ آورده شده است. در اینحالت با وجود اینکه دقت نهایی الگوریتم پیشنهادی در مقایسه با حالت قبل مقداری کاهش دارد، اما باز هم دارای بیشترین دقت تشخیص نسبت به روشهای مشابه آن می باشد.



شکل ۶-۲۹ مقایسه‌ای بین درصد تشخیص نهایی سیستم به ازای روشهای مختلف برای ۵ کلاس آوایی



شکل ۶-۳۰ مقایسه‌ای بین درصد تشخیص نهایی سیستم به ازای روشهای مختلف برای ۸ کلاس آوایی

۶-۴- نتیجه‌گیری و کارهای آینده

در این رساله هدف ما، ارائه یک روش جدید برای استخراج ویژگی به کمک تبدیل میلین یا بعبارت بهتر، نوع محدود شده آن یعنی تبدیل مقیاس می‌باشد. برای انجام اینکار ما از تلفیق چندین روش جدید و ابداعی استفاده نمودیم. در اینحالت ویژگیهای استخراج شده با استفاده از الگوریتم پیشنهادی ما که تلفیقی از روشهای کپستروم بهبود یافته به همراه تبدیل مقیاس و یکی از روشهای استخراج ویژگی MFCC یا PLP بدست می‌آید، با توجه به آزمایشات انجام شده و مقایسه آن با ویژگیهای بدست آمده از بقیه روشهای مرسوم می‌توان گفت بیان کننده بهتری برای دسته بندی سیگنال آوایی می‌باشند. در حالیکه این ویژگیها مستقل از نوع، سن و جنسیت گوینده نیز می‌باشد.

یکی از روشهای جدیدی که در الگوریتم پیشنهادی ما معرفی گردید استفاده از روش کپستروم بهبود یافته بود که مزیت آن نسبت به روشهای مشابه آن که برای استخراج پوش طیفی یک سیگنال بکار می‌رود، ارائه یک پارامتر کنترلی برای افزایش دقت تشخیص نهایی سیستم می‌باشد. همانگونه که در شکل ۶-۳۱ می‌بینید، درصد تشخیص نهایی سیستم به ازای تغییرات پارامتر γ که در بخش ۶-۳ راجع به آن صحبت نمودیم، نشان داده شده است. در اینحالت، بیشترین دقت تشخیص سیستم هنگامی بدست می‌آید که $\gamma = 10$ باشد و به ازای مقادیر بیشتر یا کمتر از آن دقت تشخیص کاهش می‌یابد.

مزایای الگوریتم پیشنهادی نسبت به الگوریتمهای مشابهی که برای تشخیص آواها بکار برده می‌شود، دقت و سرعت بالای تشخیص است، علت این امر هم بخاطر تعداد ویژگیهای استخراج شده از هر آوا است. همین امر باعث می‌گردد تا بتوان از این سیستم در تشخیص آواهای زمان واقعی¹¹⁵ استفاده نمود. از جمله مزایای دیگر این روش دقت ویژگیهای استخراج شده است که بعنوان یکی از مهمترین ویژگیهای این الگوریتم می‌توان از آن یاد کرد. ما برای بررسی این امر با استفاده از الگوریتم ویژگیهای آواها را استخراج نمودیم و سپس با استفاده از یک دسته بندی ساده داده‌ها را کلاسه‌بندی نمودیم. دقت بسیار بالای این الگوریتم در مقایسه با روشهای دیگر با توجه به نوع دسته بندی مورد استفاده حاکی از دقت ویژگیهای استخراج شده است.

اما یکی از مشکلاتی که در رابطه با این الگوریتم با آن مواجه بودیم، ارائه یک روش مناسب برای استخراج پوش طیفی اصوات است. با بررسی و آزمایشات انجام شده و روشهای بسیار زیادی که مورد بررسی قرار گرفت به این نتیجه رسیدیم که رابطه دقیق و مشخصی بین پوش طیفی آواهای یکسان و یا حتی اصوات وجود ندارد و نمی‌توان روش کاملاً قطعی را برای بدست آوردن پوش طیفی یکسان برای همه آواهای یکسان ادا شده توسط افراد مختلف بدست آورد.

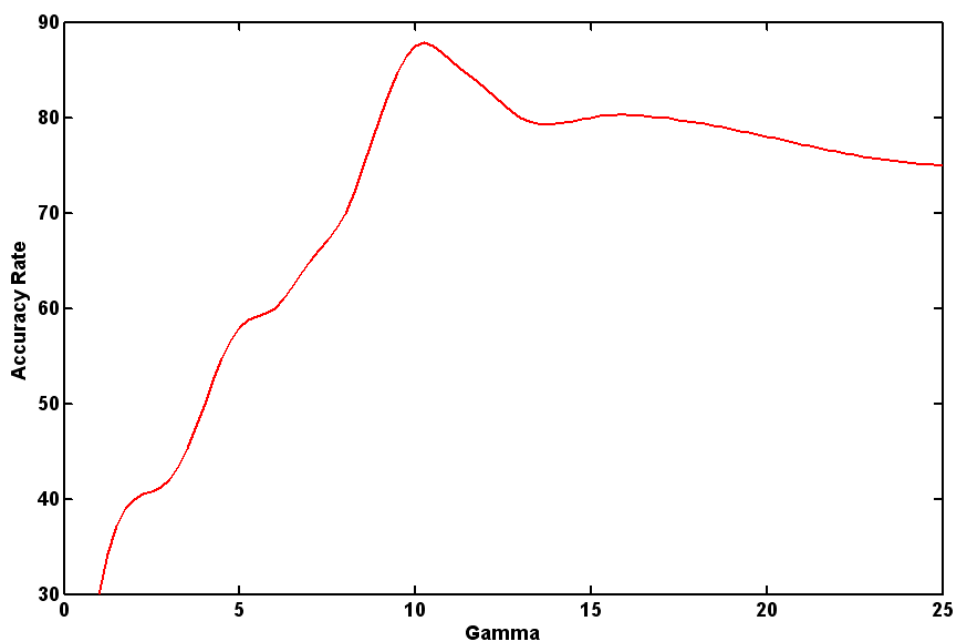
یکی از ویژگیهایی که آواهای مختلف را از یکدیگر متمایز می‌سازد، شکل حفره صوتی است. استفاده از روشهای تقریب پوش طیفی تا حدودی باعث کمرنگ شدن این تغییرات می‌گردد. استفاده از روشهایی که بتواند تغییرات حفره صوتی را در پوش طیفی اعمال کند می‌تواند باعث افزایش کارایی و دقت تشخیص الگوریتم پیشنهادی گردد. روش کپستروم بهبود یافته تا حدی توانست این مشکل را برطرف سازد، اما با توجه به یکسان نبودن پوش طیفی آواهای یکسان سیستم نیازمند آموزش است تا از این طریق بتواند

¹¹⁵ . Real Time

ویژگیهای لازم را از هر کلاس آوایی استخراج نماید. البته نکته‌ای که لازم است در اینجا به آن اشاره کنیم، استفاده از الگوریتمهای قدرتمندتر برای افزایش کارایی سیستم تشخیص آواها است. همانطور که قبلاً اشاره کردید با توجه به محدود بودن مجموعه داده‌هایی که برای آموزش و تست سیستم استفاده گردید، امکان استفاده از الگوریتمهای قدرتمند آماری وجود نداشت، اما با آزمایشات مقدماتی صورت گرفته به این نتیجه رسیدیم که استفاده از الگوریتمهایی مانند HTK می‌تواند دقت تشخیص را بالاتر نیز ببرد.

از جمله اقداماتی که در آینده قصد داریم برای افزایش کارایی دقت تشخیص سیستم استفاده نماییم، بهینه نمودن بردار وزندار مرحله کلاس بندی است و اینکار را می‌توان با استفاده از روشهای معروف بهینه‌سازی، مانند شبکه‌های عصبی، الگوریتم ژنتیک و ... انجام داد.

یکی از مشکلات دیگر، تغییر نرخ دقت تشخیص نهایی سیستم با افزایش تعداد کلاسهای آواهاست. اما در هر حالتی، همواره دقت تشخیص الگوریتم پیشنهادی بیشتر از روشهای مشابه است. در اینحالت با افزایش تعداد کلاسها، دقت تشخیص کاهش می‌یابد که در کمترین حالت برابر با دقت تشخیص نهایی ۸۹٫۵ درصد به ازای ۸ کلاس آوایی مختلف و ۱۲ گوینده مختلف است. علت این امر استفاده از تقریبهای پوش طیفی است که این امر اثر فرکانسهای رزونانس را کاهش می‌دهد. از اینرو بر آن هستیم تا با اعمال تغییرات مناسب در الگوریتم کپستروم بهبود یافته بتوانیم مدل مناسبی برای اصوات مختلف بیان شده توسط افراد مختلف بدست آوریم. در این روش بایستی بتوان تغییرات شکل حفره صوتی را در پوش طیفی حاصل اعمال نمود.



شکل ۶-۳۰ تغییرات دقت تشخیص سیستم به ازای تغییرات پارامتر γ

۶-۵- مراجع

- [1] T. Irino and R. Patterson. "Extracting size and shape information of sound source in an optimal auditory processing model," In CASA workshop, IJCAI-99, August 1999.
- [2] David R. R. Smith and Roy D. Patterson, "The interaction of glottal pulse rate and vocal tract length in judgments of speaker size, sex, and age," J. Acoustic Soc. Am. 118(5), November 2005, Pages 3177-3186.

- [3] D. Arfib, F. Keiler, and U. Zölzer. Source-filter processing. In U. Zölzer, editor "Digital Audio Effects," pages 299-372. John Wiley and sons, Ltd, Chichester Sussex, UK. 2002.
- [4] <http://www.utdallas.edu/~assmann>
- [5] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE transactions on Acoustics, Speech and signal processing, ASSP-28(4): 357-365, 1980.
- [6] M. J. Hunt, "Spectral signal processing for ASR," Automatic Speech Recognition and Understanding Workshop (ASRUW), 1990.
- [7] S. Young and G. Evermann, "Hidden Markov Model toolkit Version 3.1," Published by Cambridge University Engineering Department, December 2001.
- [8] V. Tyagi and C. Wellekens, "On desensitizing the mel-cepstrum to spurious spectral components for robust speech recognition," in Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on, vol. 1, 2005, pp. 529–532.
- [9] S. Sigurdsson, K. B. Petersen and T. L. Schioler, "Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music," In proceedings of the 10th International Conference on Speech and computer (SPECOM 2005), volume 1, pages 191-194, 2005.
- [10] H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis of speech," The Journal of the Acoustical Society of America, 87(4): 1738-1752, 1990.
- [11] J. Psutka, L. Muller, and V. Psutka. "Comparison of MFCC and PLP Parameterizations in the Speaker Independent Continuous Speech Recognition Task," In Proc. Eurospeech, 2001.
- [12] M. Jamaati, H. Marvi, "Vowels Recognition Using Mellin Transform and PLP-based Feature Extraction," Acoustics08-paris 2008, France.
- [13] M. Jamaati, M. Lankarany and H. Marvi, "On the Use of Scale Transform in Modeling of the Spectral Envelope of Vowels," Acoustics08, June 29-July 4, 2008, France.

