

## بررسی عملکرد سیستم شنوایی انسان در ماسک اطلاعات نامربوط سیگنال صوتی جهت کاهش بهینه حجم حافظه مورد نیاز جهت فشرده سازی

محمد بهرفتار<sup>۱</sup>، محسن حیاتی<sup>۲</sup>

۱: دانشجوی کارشناسی ارشد الکترونیک دانشگاه رازی کرمانشاه، ۲: عضو هیأت علمی گروه برق دانشگاه رازی کرمانشاه

E-mail: behraftar@mail.znu.ac.ir , hayati@razi.ac.ir

چکیده - درک درستی از مکانیزم سیستم شنوایی انسان الگوی مناسبی جهت طراحی بهینه دستگاههایی که به نحوی در ارتباط با ذخیره و پردازش اطلاعات صوتی هستند، در اختیار طراحان قرار می دهد. می دانیم که گوش انسان در هر دو حوزه زمان و فرکانس دارای محدودیتهای جهت احساس شنوایی می باشد؛ برای مثال بیشتر مواقع متوجه شده ایم که بعضی از اصوات ضعیف در حضور سیگنالهای قوی غیرقابل شنیدن می شوند. در حقیقت هدف ما تشخیص قسمت های نامربوط صوت است که از لحاظ فیزیکی وجود دارند ولیکن گوش انسان قادر به تشخیص آنها نیست و رمزگذاری و ذخیره سازی اطلاعات مربوط به این سیگنالها در حقیقت یک کار بی نتیجه ای است و به جز اتلاف زمان و توان در یک سیستم سود دیگری ندارد. در این مقاله ابتدا توضیح مختصری درخصوص فیزیولوژی گوش ارائه شده و در ادامه گستره فرکانسی و دینامیکی قابل تشخیص توسط سیستم شنوایی، به همراه مفاهیم فرآیند ماسک فرکانسی و ماسک زمانی مورد بررسی قرار می گیرد.

کلید واژه- صوت، آستانه شنوایی مطلق، باند بحرانی، ماسک صوتی، تابع انتشار ماسک.

### ۱- مقدمه

صوت<sup>۱</sup>، در واقع نتیجه ارتعاشات مکانیکی در مولکولهای اجسام می باشد؛ این ارتعاشات بر مولکولهای هوا ضربه وارد کرده و باعث انرژی گرفتن مولکولهای هوا می شوند و بدین ترتیب امواج صوتی را ایجاد می کنند که شدت و دامنه این اصوات به صورت فشار متغیر با زمان و با واحد پاسکال (Pa) بیان می شود. کمترین تغییر فشار هوایی که گوش انسان قادر به تشخیص آن است، برابر ۱۰ میکروپاسکال می باشد که این شدت در فرکانسهایی اندازه گیری شده است که گوش انسان بیشترین حساسیت را نسبت به تغییرات فشار هوا دارد و اغلب به عنوان سطح مرجع، هنگام توصیف شدت صوت به کار می رود و منحنی آستانه شنوایی مطلق<sup>۲</sup> را تشکیل می دهد. در مقابل، ماکزیمم فشار هوایی که گوش انسان قادر به تشخیص آن است برابر ۱۰۰ پاسکال است که

در این تراز از شدت صوت، منحنی آستانه درد<sup>۳</sup> را خواهیم داشت.

### ۲- آستانه شنوایی مطلق:

آستانه شنوایی مطلق یا آستانه سکون<sup>۴</sup>، کمترین تراز شدت صوتی را نشان می دهد که یک صوت در غیاب اصوات دیگر و در محیطی کاملاً ساکت، باید دارای این حداقل شدت صوت باشد تا قابل تشخیص توسط گوش گردد. منحنی آستانه سکون، معمولاً توسط یک سری آزمایشات تجربی و در حضور شنونده های متعدد و متوسط گیری از نتایج بدست آمده، ترسیم می شود؛ شکل (۱) گستره فرکانسی قابل شنود را به صورت تابعی از فرکانس نشان می دهد.

<sup>1</sup> Sound

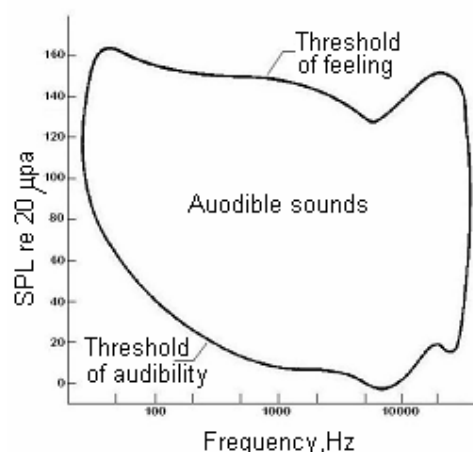
<sup>2</sup> Absolute threshold of hearing

<sup>3</sup> Threshold of pain

<sup>4</sup> Threshold in quiet

متناظر با مقدار پهنای باند بحرانی مربوطه می باشد که این فیلترها با نام فیلترهای صوتی معروف شدند.

اگر بخواهیم گوش داخلی را بصورت مجموعه ای گسسته از فیلترهای صوتی نمایش دهیم که با هم هیچگونه همپوشانی نداشته باشند، در اینصورت جهت نمایش کل گستره فرکانسی قابل شنوایی، به تعداد ۲۵ باند بحرانی کفایت خواهد کرد؛ پهنای هر یک از این باندهای بحرانی با پوشش فرکانس مرکزی ای در گستره فرکانسی قابل شنود، در جدول (۱) لیست شده اند. همچنین فاصله بین باندهای بحرانی با مقیاس critical band rate و با واحد "بارک"<sup>۷</sup> بیان می شود که افزایشی به اندازه یک بارک، معادل با یک واحد افزایش در شماره باند بحرانی خواهد بود. در شکل (۲) رابطه بین مقیاس فرکانسی بارک و هرتز، در یک نمودار ارائه شده است.



شکل ۱: گستره فرکانسی قابل شنود [۷]

## ۲-۱- باند بحرانی<sup>۵</sup>:

مفهوم باند بحرانی جهت تعیین محدوده فرکانس نامی در محدوده تغییرات فرکانسی محرک های صوتی معرفی شد که این محدوده فرکانسی، قوه ادراک انسان را بیشتر تحت تأثیر قرار می دهد؛ همچنین اظهار شد که گوش انسان از انرژی صوت در داخل محدوده فرکانسی باند بحرانی انتگرال می گیرد؛ به عبارتی زمانی که دو صوت در محدوده یک باند بحرانی واقع می شوند، صوتی که سطح انرژی بالاتری دارد، بر قوه ادراک مسلط شده و صوت با سطح انرژی پائین تر ماسک می شود.

فلتچر<sup>۶</sup>، مفهوم باند بحرانی را برای اولین بار در سال ۱۹۴۰ معرفی کرد؛ وی آستانه شنوایی یک نوای خالص سینوسی را در حضور نویزی با پهنای باند باریک و با فرکانس مرکزی ای برابر با فرکانس موج سینوسی اندازه گرفت؛ در شرایطی که چگالی توان نویز ثابت نگه داشته شده بود، وی پهنای باند نویز را بتدریج افزایش می داد؛ در این شرایط مشاهده کرد که آستانه شنوایی همراه با افزایش پهنای باند نویز به تدریج افزایش می یافت تا جاییکه در یک نقطه معین افزایش بیشتر در پهنای باند نویز، تأثیر بسیار کمتری بر روی آستانه شنوایی ایجاد می کرد؛ از اینرو فلتچر پهنای باند آستانه شنوایی گوش برای موج سینوسی مزبور را Critical Bandwidth (پهنای باند بحرانی)، نامید. پیرو این آزمایشات فلتچر اعلام کرد که سیستم شنوایی انسان شبیه به یک مجموعه از فیلترهای میان گذر عمل می کند که پهنای باند هر فیلتر

جدول ۱: لیست باندهای بحرانی که توسط اسکارف بدست آمدند. [۸]

Critical Band Number	Lower Freq. (Hz)	Center Freq. (Hz)	Upper Freq. (Hz)	Bandwidth (Hz)
1	0	50	100	100
2	100	150	200	100
3	200	250	300	100
4	300	350	400	100
5	400	450	510	110
6	510	570	630	120
7	630	700	770	140
8	770	840	920	150
9	920	1000	1080	160
10	1080	1170	1270	190
11	1270	1370	1480	210
12	1480	1600	1720	240
13	1720	1850	2000	280
14	2000	2150	2320	320
15	2320	2500	2700	380
16	2700	2900	3150	450
17	3150	3400	3700	550
18	3700	4000	4400	700
19	4400	4800	5300	900
20	5300	5800	6400	1100
21	6400	7000	7700	1300
22	7700	8500	9500	1800
23	9500	10500	12000	2500
24	12000	13500	15500	3500
25	15500	19500		

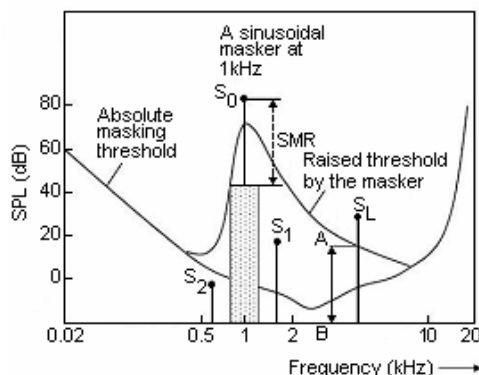
<sup>5</sup> Critical Band

<sup>6</sup> Fletcher

<sup>7</sup> Bark

که انرژی آنها زیر منحنی آستانه‌ماسک قرارداشته باشند، به خاطر حضور سیگنال قوی‌تر  $S_0$ ، ماسک خواهندشد؛ از اینرو تحت این شرایط، سیگنالهای ضعیف‌تر  $S_1$  و  $S_2$  بکلی غیرقابل شنود خواهندبود که این بدلیل پایین بودن تراز شدت صوت مربوط به هریک از این سیگنال‌ها از منحنی آستانه‌شنوایی مربوط به سیگنال  $S_0$ ، می‌باشد. در این شرایط، سیگنال  $S_L$  نیز تا اندازه‌ای ماسک می‌شود ولی بخش‌دریافتی‌سیگنال<sup>۱۲</sup>، بالاتر از منحنی ماسک قراردارد؛ از اینرو می‌توان در زیرباندی که حاوی سیگنال  $S_L$  می‌باشد، میزان نویز حاصل از کوانتیزاسیون را بالاتر فرض کرد که این به معنی نیاز به تعداد بیت‌های کمتر جهت نمایش سیگنال در این زیرباند می‌باشد و بیشتر نویزها بدلیل قرارگرفتن انرژی‌شان در محدوده AB ماسک خواهندشد؛ البته مشخص است که حتی بدون حضور هیچ‌نوع ماسک‌کننده‌ای، سیگنالهایی نظیر  $S_2$  که انرژی آنها زیر آستانه‌شنوایی مطلق می‌باشند، نیز ماسک خواهندشد.

به این ترتیب در ساختار الگوریتم‌های کدگذاری صوت که بر اساس اصول ادراکی سیستم‌شنوایی عمل می‌کنند، مدل آواشنودی، اطلاعاتی درباره نحوه کوانتیزاسیون خطوط فرکانسی سیگنال صوتی را در اختیار بلوک کوانتیزاسیون غیریکنواخت<sup>۱۳</sup> قرار می‌دهد و این بلوک با استفاده از اطلاعات موجود، مؤلفه‌های طیفی صوت را با در نظر گرفتن محدودیت‌های ادراکی گوش، کوانتیزه می‌کند [۲].



شکل ۳: نمایش آستانه ماسک فرکانسی همراه با آستانه ماسک مطلق شنوایی [۸]

<sup>8</sup> Auditory masking

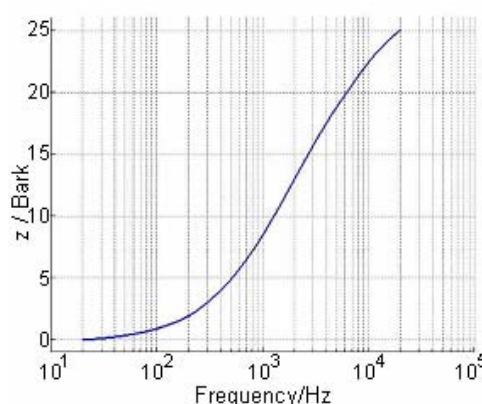
<sup>9</sup> Temporal masking

<sup>10</sup> Simultaneous masking

<sup>11</sup> Narrowband masker

<sup>12</sup> Perceivable portion

<sup>13</sup> Nonuniform quantization block



شکل ۲: رابطه بین مقیاس فرکانسی بارک و هرتز در نمودار لگاریتمی [۷]

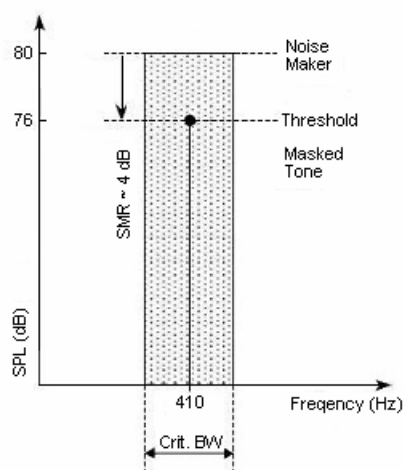
### ۳- ماسک صوتی<sup>۸</sup>:

ماسک صوتی اشاره به فرآیندی دارد که طی آن یک صدا به سبب حضور در مجاورت صداهای دیگر غیرقابل شنود می‌شود. پدیده ماسک در واقع به علت افزایش آستانه‌شنوایی نسبت به آستانه مطلق، به سبب رخداد یک سیگنال با تراز شدت صوت بالاتر در گستره فرکانسی قابل شنود و تحت تأثیر قرارگرفتن سیگنالهای با تراز پایین‌تر نسبت به آستانه‌شنوایی، ایجاد می‌شود که در این فرآیند از سیگنال قویتر با نام ماسک‌کننده و از سیگنال ضعیف‌تر با نام ماسک‌شونده یاد می‌شود. معمولاً پدیده ماسک صوتی در دو طبقه ماسک فرکانسی و ماسک زمانی<sup>۹</sup> تقسیم‌بندی می‌شود.

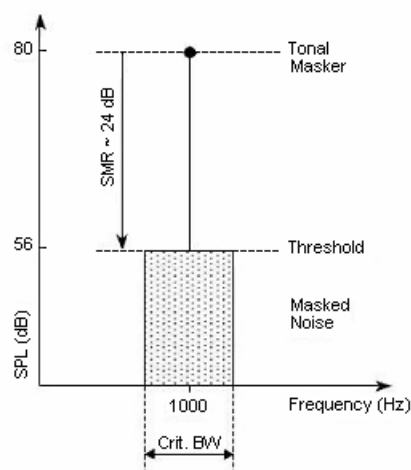
### ۳-۱- ماسک فرکانسی

ماسک فرکانسی که با نام ماسک همزمانی<sup>۱۰</sup> نیز معروف است، یک پدیده در حوزه فرکانس است که طی آن یک سیگنال تراز پایین (ماسک‌شونده)، با وقوع همزمان با یک سیگنال قویتر (ماسک‌کننده)، که طیف فرکانسی آنها نزدیک هم هستند، غیرقابل شنود می‌شود. میزان ماسک توسط سیگنال قویتر بستگی به دو عامل دارد: تراز شدت صوت و فرکانس ماسک‌شونده، که این دو عامل آستانه‌شنوایی سیگنال ماسک‌کننده را تحت تأثیر قرار می‌دهند.

در شکل (۳) مثالی از یک سیگنال ماسک‌کننده دارای پهنای باند باریک<sup>۱۱</sup> که دارای شدت صوتی برابر با  $SPL=80\text{ dB}$  است و با برجسب  $S_0$  مشخص شده، به همراه آستانه ماسک آن نمایش داده شده است؛ در این شرایط تمامی سیگنال‌هایی



شکل ۴: مثالی از عدم تقارن در ماسک همزمانی، ماسک نوا توسط نویز [۴]



شکل ۵: مثالی از عدم تقارن در ماسک همزمانی، ماسک نویز توسط نوا [۴]

در آستانه شنوایی محاسبه شده در محدوده‌ای برابر با باند نویزی که توسط نوای خالص ماسک شده است، مینیمم مقدار SMR، و به عبارتی کمترین اختلاف مابین شدت صوت نوای خالص سینوسی به عنوان ماسک کننده، و شدت صوت سیگنال نویز به عنوان ماسک شونده، زمانی رخ می‌دهد که فرکانس ماسک کننده نزدیک به فرکانس مرکزی سیگنال نویز قرار داشته باشد؛ حداقل مقدار SMR در این حالت در

پس از نقطه نظر حوزه فرکانس، اینکه حضور کدام ناحیه خاص از طیف سیگنال منجر به ماسک انرژی طیف‌های دیگر خواهد شد، توسط میزان دامنه طیف فرکانسی مربوط به سیگنال‌های ماسک کننده و ماسک شونده تعیین می‌شود. از چشم انداز حوزه زمان نیز رابطه فازی مابین محرک‌های صوتی می‌تواند پی‌آمد ماسک صوتی را تحت تأثیر قرار دهد.

از اینرو می‌توان نتیجه گرفت ماسک همزمانی فرآیندی است که طی آن، حضور یک سیگنال نویز قوی و یا یک سیگنال نوای خالص سینوسی به عنوان سیگنال ماسک کننده، موجب ایجاد نیروی برانگیزش کافی در غشاء پایه<sup>۱۴</sup> مربوط به قسمت حلزونی گوش درونی و بدین ترتیب در محل باند بحرانی از این غشاء می‌شود و بدین ترتیب مانع از آشکارسازی مؤثر سیگنال‌های ضعیف‌تر می‌گردد. اگرچه هر یک از خطوط فرکانسی مربوط به سیگنال صوتی، ممکن است شامل سناریوی ماسک همزمانی پیچیده‌ای باشند، ولیکن به منظور شکل‌دهی اعوجاج‌های حاصل از کدگذاری، مناسب است که مابین تنها سه دسته از انواع مختلف ماسک همزمانی تمییز قائل شویم که عبارتند از: ماسک نوا توسط نویز<sup>۱۵</sup> (NMT)، ماسک نویز توسط نوا<sup>۱۶</sup> و ماسک نویز توسط نویز<sup>۱۷</sup> (NMN).

### ۳-۱-۱- ماسک نوا توسط نویز (NMT):

در سناریوی NMT، یک سیگنال نویز پهنای باند باریک<sup>۱۸</sup> و با پهنای باندی برابر با "۱" بارک، نوایی که در باند بحرانی مربوطه قرار دارد را ماسک خواهد کرد، مشروط بر اینکه شدت صوت مربوط به نوای خالص سینوسی به عنوان ماسک شونده، پایین‌تر از آستانه شنوایی محاسبه شده باشد؛ مقدار آستانه در این حالت، مستقیماً به شدت صوت و تا حدی فرکانس مرکزی نویز، بستگی خواهد داشت، شکل (۴).

### ۳-۱-۲- ماسک نویز توسط نوا (TMN):

در این حالت، یک نوای خالص سینوسی که در فرکانس مرکزی یک باند بحرانی واقع می‌شود، تمامی سیگنال‌های نویز موجود در آن باند بحرانی را به همراه اثرات نویز باندهای مجاور ماسک خواهد نمود، مشروط بر اینکه شدت صوت مربوط به طیف نویز، پایین‌تر از آستانه شنوایی محاسبه شده باشد؛ مقدار آستانه شنوایی در این حالت، مستقیماً به شدت و تا حدی فرکانس سیگنال نوای سینوسی بستگی دارد، شکل (۵).

<sup>14</sup> Basilar membrane

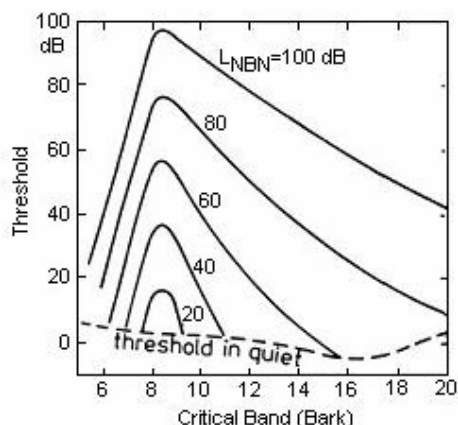
<sup>15</sup> Noise-Making-Tone (NMT)

<sup>16</sup> Tone-Making-Noise (TMN)

<sup>17</sup> Noise-Making-Noise (NMN)

<sup>18</sup> Narrow-band noise signal

گستره ۲۸-۲۱ dB قرار خواهد داشت.



شکل ۶: تغییرات منحنی آستانه ماسک در رابطه با محرک‌های صوتی با دامنه‌های متفاوت [۷]

گستره شنوایی قابل شنود، جهت تشکیل و ساخت منحنی آستانه ماسک جمعی<sup>۲۲</sup> می‌توان از خواص مربوط به NMT، TMN و NMN بهره‌برد. همچنین در کنار روش‌های موجود جهت تخمین آستانه ماسک، شناخت عمیق از خصوصیات عدم تقارن ماسک می‌تواند زمینه را برای دستیابی به مدل‌های ادراکی بهینه‌شده فراهم کند [۳]. در نهایت این موضوع، حاکی از آن است که الگوهای ماسکی که برپایه میزان انرژی سیگنال استوارند، جهت تخمین توان ماسک صوتی در میان کدکننده‌های ادراکی، ممکن است تنها مادامیکه پهنای باند سیگنال ماسک‌کننده، بزرگتر از پهنای باند سیگنال ماسک‌شونده و یا حداکثر مساوی با آن باشد، معتبر باشند؛ در شرایطی هم که پهنای باند ماسک‌شونده از پهنای باند ماسک‌کننده تجاوز می‌کند، در محاسبات مربوط به آستانه ماسک بایستی از معیار پوش<sup>۲۳</sup> استفاده کرد [۳] [۸].

### ۳-۱-۵- تابع انتشار ماسک<sup>۲۴</sup>:

همان‌طوری که پیش‌تر نیز اشاره شد، اثر ماسک هم‌زمانی در حالات مختلف ماسک، تنها محدود به باند بحرانی جاری که حاوی سیگنال ماسک‌کننده است، نمی‌شود بلکه اثر آن در باندهای مجاور نیز نشر می‌یابد. همچنین توضیح دادیم که

### ۳-۱-۳- ماسک نویز توسط سیگنال نویز دیگر (NMN):

توصیف سناریوی ماسک نویز توسط نویز (NMN)، که طی آن یک سیگنال نویز با پهنای باند باریک منجر به ماسک نویزی با پهنای باند محدود دیگر می‌شود، بدلیل روابط فازی پیچیده مابین سیگنال‌های ماسک‌کننده و ماسک‌شونده، کمی سخت‌تر است؛ در واقع روابط فازهای مختلف بین مؤلفه‌های فرکانسی زیاد هریک از این سیگنال‌ها، می‌تواند منجر به آستانه‌های متفاوتی از SMR شود. نتایج مربوط به یکی از تحقیقات انجام‌یافته توسط Hall، پیرامون آستانه‌های بازایی‌شده برای سیگنال‌های نویزی با پهنای باند باریک، که دارای مؤلفه‌های فرکانسی گوناگون با شدت صوت‌های متفاوت بودند، حکایت از مقدار 26dB برای ضریب SMR در این حالت را دارد [۸].

### ۳-۱-۴- عدم تقارن در ماسک:

مثالی از NMT و TMN در اشکال (۴) و (۵)، به وضوح عدم تقارن موجود در توان ماسک اغتشاشات صوتی، توسط سیگنال ماسک‌کننده نویز و سیگنال ماسک‌کننده نوا را در دو حالت متفاوت نشان می‌دهد؛ با وجود آنکه هر دو سیگنال ماسک‌کننده در شدت صوت  $SPL=80\text{ dB}$  ارائه شده‌اند، ولیکن آستانه SMR مربوط به این دو فرآیند، تا ۲۰ dB با هم اختلاف دارند. در واقع وجود این عدم تقارن، انگیزه اصلی از بررسی الگوهای ماسک NMT، TMN و همچنین NMN می‌باشد؛ در این راستا و به منظور ارائه راه کارهای مناسب جهت عدم تشخیص بودن اعوجاج‌های حاصل از کدگذاری صوت توسط سیستم شنوایی انسان، آگاهی از هر سه‌ی این سناریوها می‌تواند بسیار مفید باشد.

برای هریک از فواصل تحلیل زمانی، بلوک مدل ادراکی در ساختار الگوریتم کدگذار صوت، باید مؤلفه‌های طیفی آهنگین<sup>۱۹</sup> و غیرآهنگین<sup>۲۰</sup> مربوط به هر دو سیگنال ماسک‌کننده و ماسک‌شونده و همچنین اعوجاج‌های حاصل از کدگذاری را از میان خطوط طیف فرکانسی مربوطه‌شان استخراج کرده و در مرحله بعد بایستی از معادلات مربوط به ماسک‌صوتی که تابعی از فرکانس می‌باشند، استفاده کند. عطف به روابط حاکم بر چگونگی انتشار منحنی ماسک<sup>۲۱</sup> در

<sup>19</sup> Tone-like components

<sup>20</sup> Noise-like components

<sup>21</sup> Spreading function

<sup>22</sup> Global masking threshold

<sup>23</sup> An envelope-based measure

<sup>24</sup> Spreading function

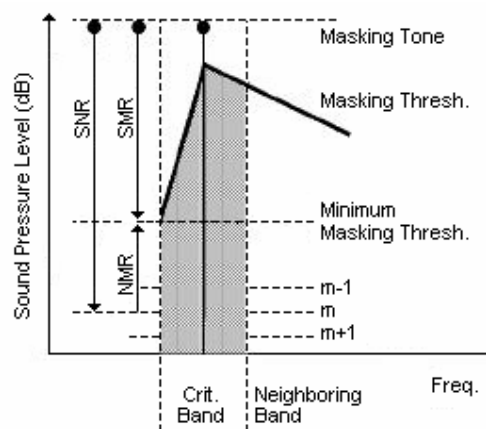
ماسک کننده‌های آهنگین، کمتر از ماسک کننده‌های نویز با پهنای باند محدود می‌باشد بنابراین ماسک کننده‌های نویز با پهنای باند باریک، در مقایسه با ماسک کننده‌های آهنگین ماسک بیشتری را ایجاد خواهند کرد؛ از اینرو مقدار این آفست به شدت وابسته به طبیعت نویز گونه یا آهنگینی سیگنال ماسک کننده خواهد بود؛ این آستانه توسط روابط زیر قابل محاسبه می‌باشد [۵]:

$$\begin{aligned} O_T &= [14.5 + z] \\ O_N &= K \\ TH_N &= E_T - O_T \\ TH_T &= E_N - O_N \end{aligned} \quad (2)$$

که  $O_T$  معرف مقدار آفست مربوط به ماسک کننده آهنگین،  $O_N$  مقدار آفست مربوط به ماسک کننده نویز،  $TH_N$  معرف آستانه ماسک نویز به سبب فرآیند ماسک نویز توسط نوا (TMN)، و  $TH_T$  معرف آستانه ماسک نوا به سبب پدیده ماسک نوا توسط نویز (NMT) می‌باشد. همچنین عبارت  $E_T$  تراز انرژی ماسک کننده نوا در باند بحرانی و  $E_N$  سطح انرژی ماسک کننده نویز در باند بحرانی<sup>۲۶</sup> را نشان می‌دهد و  $K$  در این رابطه عددی بین ۳-۵ dB را داراست؛  $z$  نیز معرف شماره باند بحرانی مورد بررسی است.

قابل ذکر است که آستانه‌های ماسک محاسبه شده از طریق رابطه (۲)، تنها سهم ماسک کننده‌های آهنگین و غیر آهنگین خاص مربوط به طیف فرکانسی سیگنال صوت را در بر می‌گیرند که البته نیازی به توضیح نیست که در سناریوی کدگذاری واقعی، هر فریم به طور نوعی شامل مجموعه‌ای از هر دو نوع ماسک کننده‌ها می‌باشد.

پس از محاسبه آستانه ماسک مربوط به تمامی مؤلفه‌های مربوط به سیگنال نوا و نویز در محدوده هر باند بحرانی، این آستانه‌های ماسک فردی<sup>۲۷</sup> با هم ترکیب می‌شوند تا آستانه ماسک جمعی را بوجود آورند. منحنی آستانه ماسک جمعی بدست آمده، در حقیقت ارزیابی صحیح از تراز را که نویز حاصل از کوانتیزاسیون، در صورت فراتر رفتن سطح انرژی آن از محدوده این تراز، عیناً قابل شنود می‌شود را در اختیار ما قرار می‌دهد؛ در نتیجه منحنی آستانه ماسک جمعی



شکل ۷: نمایش شماتیکی ماسک همزمانی (فرکانسی)، [۹]

برپایه نتایج آواشنودی، اثر ماسک طیفی در فرکانسهای باند میانی، تابعی از فرکانس و تراز سیگنال ماسک کننده است که این موضوع در شکل (۶) به وضوح نشان داده شده است؛ این اثر که با نام "انتشار اثر ماسک"<sup>۲۵</sup> نیز شناخته می‌شود، در کاربردهای کدگذاری صوت و در حوزه فرکانسی بارک، قابل تقریب توسط تابع انتشاری با خروجی مثلثی شکل است که در جهت فرکانسهای پایین دارای شیبی تند و برابر با ۲۵dB/Bark و در سمت فرکانسهای مثبت، شیب آن کم و برابر با ۱۰dB/Bark است، شکل (۷).

عبارت تحلیلی مناسبی که بتواند نمودار مثلثی شکل تابع انتشار را به صورت مستقل از فرکانس و میزان تراز شدت صوت ماسک کننده، با تقریب خوبی مدل سازی کند قابل ارائه در فرم معادله زیر می‌باشد [۴]:

$$SpF_{dB}(z) = 15.81 + 7.5(z + 0.474) - 7.5 \left[ 1 + (z + 0.474)^2 \right]^{0.5} \quad (1)$$

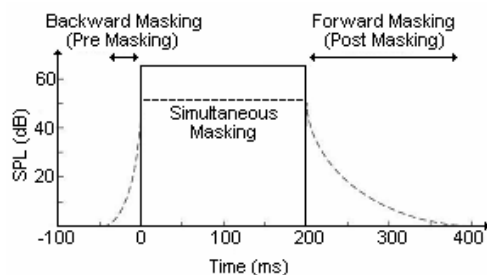
در رابطه مزبور، "z" در واحد فرکانسی بارک و  $SpF(z)$ ، برحسب دسی بل بیان می‌شود. البته در تحلیل دقیق، آستانه ماسک شکلی بسیار غیرخطی دارد که تابعی از فرکانس و دامنه سیگنال ماسک کننده است.

پس از انجام تحلیل‌های مربوط به باندهای بحرانی و محاسبه میزان انتشار ماسک در هریک از این باندها و تعیین تأثیرات آن در باندهای مجاور، آستانه‌های ماسک در کدکننده‌های ادراکی، با کم کردن یک مقدار آفست از الگوی برانگیزش بدست می‌آیند. از آنجاکه که آستانه ماسک معلول

<sup>25</sup> Spread of masking

<sup>26</sup> Critical band noise energy

<sup>27</sup> Individual masking threshold



شکل ۸: ماسک زمانی و ماسک فرکانسی در کنارهم در محور زمان نشان داده شده است. منحنی خط پر سیگنال ماسک کننده و منحنی خط چین آستانه ماسک را نشان می دهد [۶].

لگاریتمی مابین حداقل آستانه ماسک محاسبه شده نسبت به تراز ماسک کننده و تراز سیگنال نویز را نشان می دهند.

### ۳-۲- ماسک زمانی<sup>۳۲</sup>:

ماسک زمانی که با عبارت ماسک غیرهمزمان<sup>۳۳</sup> هم شناخته می شود، نقش مهمی را در ادراک شنوایی انسان در حوزه زمان ایفاء می کند و زمانی رخ می دهد که دو صوت، نسبت به هم با اختلاف زمانی کوتاهی واقع شوند؛ در این صورت سیگنال قوی تر، موجب ماسک سیگنال با سطح انرژی پایین تر خواهد شد حتی اگر سیگنال ماسک شونده قبل از سیگنال ماسک کننده رخ دهد.

اثر ماسک زمانی به دو صورت اتفاق می افتد: در حالتی که یک صدا پس از وقوع یک سیگنال قوی ماسک شود، آن را پس ماسک<sup>۳۴</sup> می نامیم و اگر قبل از آن ماسک شود، آنرا پیش ماسک<sup>۳۵</sup> گوئیم. هر چند عجیب به نظر می رسد ولی پیش ماسک وجود دارد حتی اگر مدت کوتاهی مانند ۲۰ ms قبل از ماسک کننده رخ دهد؛ در مقابل پس ماسک تا ۲۰۰ ms بعد نیز می تواند اثر کند. در شکل (۸) ماسک زمانی به همراه ماسک فرکانسی نشان داده شده که در آن، منحنی خط پر سیگنال ماسک کننده و منحنی خط چین آستانه ماسک را نشان می دهد.

گهگاهی بر میزان تراز "JND"<sup>۳۸</sup> نیز، دلالت دارد.

رویه استاندارد در کدگذاری ادراکی صوت به این صورت است که ابتدا سیگنال های ماسک که شامل هر دو نوع سیگنال نوا و نویز می شود، طبقه بندی شده و سپس آستانه های ماسک مربوط به هر کدام، با استفاده از روابط مناسب بدست می آیند؛ سپس در مرحله بعد، از این اطلاعات جهت شکل دهی طیف نویز در پایین تر از منحنی آستانه ماسک جمعی، یا همان تراز JND، استفاده می شود. از جمله کدکننده های ادراکی بسیار معروف صوت که به این روش عمل می کنند، می توان به مدل های ادراکی استفاده شده در استاندارد صوتی MPEG که در دو گروه psychoacoustic model 1 و psychoacoustic model 2 تقسیم بندی می شود، اشاره کرد. همچنین کدکننده هایی که در ساختار الگوریتم آنها از آنتروپی ادراکی<sup>۳۹</sup> استفاده می شود، نمونه دیگری از این نوع کدکننده ها می باشند.

قابل ذکر است که در حین شکل دهی طیف های نویز، آستانه مطلق شنوایی  $T_q(f)$  نیز مدنظر قرار می گیرد و بنابراین اغلب اوقات، ماکزیمم مقدار از بین  $\max(JND, T_q(f))$ ، به عنوان آستانه اعوجاج مجاز<sup>۴۰</sup> مورد استفاده قرار می گیرد. درک درستی از پهنای باند بحرانی و همچنین فرآیند ماسک همزمانی در زمینه کدگذاری صوت، موجب تسهیل در تفهیم برخی اصطلاحات خاص بیان شده در شکل (۷)، که در آن یک سیگنال نوای سینوسی به عنوان سیگنال ماسک کننده در مرکز باند بحرانی واقع شده است، می شود (تمامی سطوح شدت صوت در شکل مزبور، برحسب واحد دسی بل بیان شده اند). وقوع نوای ماسک فرضی در برخی سطوح ماسک، باعث ایجاد برانگیزشی در طول غشاء پایه می شود که می توان آن را توسط تابع انتشار و منحنی آستانه ماسک مربوط به طیف آن سیگنال، مدل سازی کرد؛ در واقع حداقل آستانه ماسک برای باند مورد بررسی، به معنی حداقل بودن تابع انتشار ماسک در آن باند می باشد.

با فرض اینکه جهت کوانتیزاسیون اطلاعات سیگنال ماسک کننده، از یک کوانتایزر اسکالر یکنواخت<sup>۴۱</sup> که از m-bit جهت کوانتیزاسیون اطلاعات نمونه ها استفاده می کند، استفاده شود، حداکثر مقدار انرژی نویز تولید شده در فرآیند کوانتیزاسیون، برابر با فاصله هر یک از گام های این کوانتایزر خواهد بود. عبارات SMR و NMR نیز به ترتیب فاصله

<sup>۳۸</sup> Just Noticeable Distortion (JND)

<sup>۳۹</sup> Perceptual Entropy (PE)

<sup>۴۰</sup> Permissible distortion threshold

<sup>۴۱</sup> M-bit uniform scalar quantizer

<sup>۴۲</sup> Temporal masking

<sup>۴۳</sup> Non-simultaneous masking

<sup>۴۴</sup> Post-masking or backward masing

<sup>۴۵</sup> Pre-masking or forward masking

پهنای باند مورد نیاز را به  $1/12$  مقدار اولیه تقلیل داد؛ به عبارتی در این شرایط یک آهنگ با استمرار زمانی سه دقیقه، تنها ۲,۵ Mbyte از فضای حافظه را اشغال خواهد کرد.

### سپاسگزاری

در این مرحله جا دارد از خانواده خوبم و دوستان بسیار عزیزم جناب آقایان سعید بیکزاده، خدابخش پیری، مجید جبارزاده و خانم پروانه کاردوست که شرایط مناسبی را جهت نگارش این مقاله فراهم کردند، قدردانی کنم.

### مراجع

[۱] بهرFTAR، محمد "آنالیز، شبیه سازی و بهبود روش های فشرده سازی صوت طبق استاندارد "MPEG-1 Layer III"، پایان نامه کارشناسی ارشد الکترونیک، دانشگاه رازی کرمانشاه، سال ۱۳۸۵.

- [۲] Hung-Chih Lai and Jwu-Shen Hu, "Real-Time Implementation of MPEG-1 Layer3 Audio Decoder on a DSP chip", Department of Electrical and Control Engineering, National ChiaoTung University Hsinchu, Taiwan, R.O.C., June 2001.
- [۳] J. L. Hall, "Asymmetry of masking revisited: Generalization of masker and probe bandwidth," J. Acoust. Soc. Amer., vol. 101, pp. 1023-1033, Feb. 1997.
- [۴] T. Painter, A. Spanias, "Perceptual Coding of Digital Audio", Processing of the IEEE, Vol. 88, No. 4, April 2000.
- [۵] N. Jayant, J. D. Johnstone, and R. Safranek, "Signal compression based on models of human perception," Proc. IEEE, vol. 81, pp. 1385-1422, Oct. 1993.
- [۶] k.Brandenburg, "Mp3 and AAC explained", AES 17<sup>th</sup> International conference on high quality audio Coding, Fraunhofer Institute for Integrated Circuits Fh6-IIS A, Erlangen, Germany.
- [۷] David J. M. Robinson, "Perceptual model for assessment of coded audio," A thesis submitted for the degree of Doctor of philosophy, Department of Electronic Systems Engineering, University of Essex, March 2002.
- [۸] Christopher R.Cave, "Perceptual Modelling for Low-Rate Audio Coding", Department of Electrical and Computer Engineering, McGill University, Montreal, Canada, June 2002
- [۹] P. Noll, "Wideband speech and audio coding," IEEE Commun. Mag., pp. 34-44, Nov. 1993.

برخی مواقع افزایش ناگهانی انرژی سیگنال در بازه کوتاهی از زمان، منجر به ایجاد خطاهای کوانتیزاسیون لحظه ای بزرگی شده و باعث انتشار این خطاها در طول همان بلوک از نمونه های صوتی می گردد که این پیش انعکاس ها در صوت<sup>۳۶</sup>، می تواند بویژه در نرخ بیت های پایین به وضوح قابل شنیده شدن باشد. حال چنانچه این پیش انعکاس ها دارای استمرار زمانی کوتاهی باشند، قابل حذف توسط اثر حوزه زمان پدیده پیش ماسک خواهند بود. از طرفی با توجه به اینکه اثر پس ماسک برای مدتی بیش از ۱۶۰ msec استمرار می یابد، خطاهای کوانتیزاسیون مربوط به افزایش ناگهانی در انرژی سیگنال معمولاً غیر قابل شنود خواهند بود.

معمولاً در الگوریتم های کدگذاری صوتی که بر اساس اصول ادراکی گوش عمل می کنند، از هر دو اثر پیش ماسک و پس ماسک استفاده می شود [۶] [۲].

### ۴- نتیجه گیری

در مجموع بکارگیری اثر ماسک فرکانسی و ماسک زمانی، هر دو امکان می دهد تا داده های اصلی را بطور قابل توجهی و بدون هیچگونه تغییر محسوس در کیفیت ادراکی سیگنال صوتی، کاهش دهیم و بدین ترتیب از حجم مورد نیاز جهت ذخیره اطلاعات صوتی به طور فزاینده ای کاسته می شود. برای مثال جهت ذخیره سازی صدایی با کیفیت صدای ذخیره شده روی یک دیسک فشرده نوری، که در آن جهت ذخیره سازی صوت در حالت استریو دو کاناله، از فرکانس نمونه برداری ۴۴,۱ kHz استفاده شده است، برای پخش هر ثانیه از صوت به فضای حافظه ای در حدود ۱,۴۱۱,۲۰۰ (۲\*۱۶\*۴۴۱۰۰) بیت نیاز خواهد بود؛ به عبارت دیگر جهت اجرای یک فایل صوتی در فرمت PCM از طریق شبکه اینترنت و با کیفیتی معادل صدای ذخیره شده روی CD، به پهنای باندی فراتر از ۱,۴ Mbit/s نیاز خواهیم داشت؛ در این حالت جهت ذخیره یک آهنگ سه دقیقه ای، به فضای حافظه ای برابر با ۲\*۱۶\*۴۴۱۰۰\*۱۸۰ بیت و به عبارتی بالغ بر ۳۰ Mbyte نیاز خواهد بود. حال در الگوهای فشرده سازی صوت (نظیر الگوریتم MPEG)، که از شبیه سازی ویژگیهای ادراکی گوش بهره می برند، می توان

<sup>36</sup> Pre-echo effects

<sup>37</sup> Short MDCT block transform

<sup>38</sup> Long window