# استفادهٔ همزمان از اطلاعات تحریک، پوش طیف و چندی کنندهٔ ماتریسی در بهبود عملکرد سیستمهای تصدیق گوینده

ابوالقاسم صیادیان*

دانشکدهٔ مهندسی برق، دانشگاه صنعتی امیرکبیر

(دریافت مقاله: ۱۳۷۶/۷/۲۶ - دریافت نسخهٔنهایی: ۱۳۷۷/۴/۲۲)

**چکیده** - *روشهای تصدیق هویت افراد از طریق بیان کلمات یا جملات مشخص، دارای کاربرد فراوانی‌اند. وقتی تعدادکلمات یا جملات مورد توافق یا تعداد افراد برای تعیین هویت گفتاری کم باشد، از روشهای بهینه DTW[1] برای این منظور می‌توان استفاده کرد. چنانچه تعداد کلمات یا جملات زیاد باشد و یا اینکه سیستم برای تعداد افراد زیاد طراحی شود، استفاده از روش DTW مقرون به صرفه نیست (به‌علت نیاز به بار محاسباتی و حجم حافظهٔ بالا). در چنین مواقعی از چندی کننده‌های برداری یا ماتریسی می‌توان استفاده کرد. در طی این تحقیق از روش چندی کردن ماتریسی به‌علت کارایی بالاتر استفاده کرده‌ایم. همچنین برای افزایش کارایی سیستم از اطلاعات تحریک (شامل کانتورگین و فرکانس پیچ) نیز به نحو مناسبی استفاده کرده‌ایم. استفاده از اطلاعات تحریک به‌قسمی انجام پذیرفته که سیستم در مقابل تقلید ماهرانه مقاوم[2] است.*

# Using Exciting and Spectral Envelope Information and Matrix Quantization for Improvement of the Speaker Verification Systems

**A. Sayadian**

Department of Electrical Engineering, Amirkabir University of Technology

**ABSTRACT-** *Speaker verification from talking a few words or sentences have many applications Many methods as DTW, HMM, VQ and MQ can be used for speaker verification. We applied MQ for Its precise, reliable and robust performance with compuational simpblicity. We also used pitch frequency and log gain contour for further improvement of the system performance.*

## ۱- مقدمه

تعیین هویت و یا تصدیق هویت خودکار افراد به روشهای مختلف انجام پذیرفته و برای اهداف متفاوت مورد استفاده قرار می‌گیرند. سیستم تشخیص امضای رسمی، سیستم تشخیص چهرهٔ ساکن از روی کارت شناسایی، سیستم تشخیص چهرهٔ دینامیک توسط دوربینهای ویدیویی و سیستم تشخیص یا تصدیق افراد از

* استادیار

طریق گفتار، از جمله روشهای مورد استفاده برای تعیین هویت (یا تصدیق) افراد به‌طریق مختلف‌اند. هر یک از روشهای ذکر شده بر حسب نیاز می‌توانند مورد استفاده قرار گیرند. آنچه در طی این تحقیق مورد توجه است، استفاده از صدای افراد برای تعیین هویت (و یا تصدیق) شخص متقاضی (یا متکلم) است. از جمله کاربردهای این روش، دسترسی به پایگاههای اطلاعاتی از طریق دستورات گفتاری، ورود به شبکه‌های خصوصی از طریق دستورات و مکالمات گفتاری، برداشت و یا جابه‌جایی در حساب بانکی از طریق مکالمهٔ تلفنی و ... است.

سیستمهای تشخیص گوینده از نظر برخورد با گوینده به دو کلاس کلی: الف - سیستمهای تعیین هویت گوینده و ب- سیستمهای تصدیق گوینده، تقسیم می‌شوند. همچنین از نظر پاسخ این سیستمها به جمله یا کلمهٔ مورد استفاده در هنگام تشخیص، به دو کلاس وابسته به متن و یا مستقل از متن درنظر گرفته می‌شوند [۱-۳]. آنچه در این تحقیق مورد توجه است، سیستمهای تصدیق گوینده (امضای گفتاری) وابسته به تعداد کلمات و جملات معین و مشخص است. دراین سیستمها شخص یا اشخاص برای ورود یا استفاده از سیستم باید یک تا چند کلمه (یا جمله) مجاز را بیان کنند. سیستم پس از بررسی صدای متکلم یا متقاضی، او را به‌عنوان شخص مجاز پذیرفته یا رد می‌کند (تصمیم‌گیری باینری). طبیعتاً صدای افراد مجاز قبلاً به نحو مقتضی به سیستم آموزشی داده می‌شود. بنابراین هر سیستم تصدیق گوینده دارای دو فاز عمده است: الف - فاز آموزش و ب - فاز تشخیص و تصدیق. نحوهٔ عمل سیستم در طی این دو فاز به‌طور اجمال به‌قرار زیر است. در فاز آموزش، ابتدا جملات (یا کلمات) مورد نظر توسط گویندگان مجاز در چند پریود زمانی مختلف و با چند حالت بیانی مختلف (و با چند بار تکرار) بیان شده و ضبط می‌شود. آنگاه برای هر گویندهٔ مجاز، یک کتاب کد متمایز و اختصاصی طراحی می‌شود. در فاز تشخیص و تصدیق ، به‌ازای هر فریم (۲۰-۱۰ms) از سیگنال آزمون ورودی (مربوط به فرد ناشناسی یا مدعی)، ویژگیهای لازم استخراج شده و به‌عنوان بردار ورودی سیستم در نظر گرفته می‌شود. نزدیکترین الگوی مرجع از هر کتاب کد به الگوی آزمون ورودی تعیین می‌شود (با تعریف و استفاده از یک تابع فاصلهٔ مناسب). برای هر گوینده مجاز مقدار تابع درستنمایی[۳] متناظر (مانند تابع اعوجاج تجمعی[۴]) محاسبه می‌شود. در پایان فرایند تشخیص (پایان جمله یا کلمه)، تابع درستنمایی با حداقل مقدار اعوجاج تجمعی تعیین می‌شود. آنگاه مقدار متوسط تابع درستنمایی انتخاب شده با یک سطح آستانهٔ معین مقایسه می‌شود. اگر مقدار تابع درستنمایی انتخاب شده از سطح آستانه مورد نظر کمتر باشد، شخص انتخاب شده تصدیق می‌شود. در غیر این صورت فردمتقاضی که جمله (یا کلمه) را بیان کرده است به‌عنوان فرد غیرمجاز در نظر گرفته می‌شود. بنابراین مسایل عمده‌ای که در طراحی سیستم تصدیق گوینده با استفاده از روش چندی کردن ماتریسی مطرح هستند عبارت‌اند از: نوع و تعداد ویژگیها - نوع و نحوهٔ محاسبهٔ تابع فاصله - طول و تعداد فریمها که در هر مرحله یک‌جا کد می‌شود - نحوهٔ محاسبهٔ تابع اعوجاج تجمعی - نحوهٔ انتخاب سطح آستانه برای تصدیق یا رد گوینده - نحوهٔ طراحی الگوهای مرجع برای هر گوینده- انتخاب تعداد الگوهای مرجع برای هر گوینده - انتخاب تعداد کلمات و یا جملات برای آموزش سیستم - انتخاب تعداد تکرار و نحوهٔ بیان و تعداد دفعات زمانی ضبط جملات و... . در طی بخشهای مختلف این نوشتار، ضمن اشاره به روشهای عمومی برای حل مسایل طرح شده، به روش خاص مورد استفاده در طی این تحقیق اشاره خواهیم کرد.

## ۲- انتخاب نوع و تعداد ویژگیها

در اغلب کاربردهای سیگنال گفتار از جمله، بازشناسی گفتار، بازشناسی گوینده، کد کننده‌های پارامتریک و سیستمهای تبدیل متن به گفتار، دو دسته پارامتر شامل پارامترهای مربوط به دامنهٔ پوش طیف و پارامترهای مربوط به سیگنال تحریک استخراج و مورد استفاده قرار می‌گیرند. پارامترهای مهم سیگنال تحریک شامل گین، فرکانس پیچ، نسبت زمان باز و بسته بودن تارهای صوتی برای فریمهای باصداست [۴]. البته به‌خاطر بار محاسباتی و همچنین عدم امکان تخمین دقیق نسبت دورهٔ زمان باز و بسته بودن تارهای صوتی (در طی یک پریود پیچ) در عمل کمتر از آن استفاده می‌شود [۲و۴]. پارامترهای مربوط به پوش طیف به‌دلیل آنکه از یک جنس هستند، نوعاً توسط تعداد محدودی ویژگی از جمله؛ ضرایب کپسترم CC[۵] - ضرایب LSPF[۶] - ضرایب CG[۷] خروجی تحلیل بانک فیلتری[۸] و... بازنمایی شده و مورد استفاده قرار



[ DOR: 20.1001.1.2251600.1378.18.1.3.4 ]

می‌گیرند [۵-۷]. از میان ویژگیهای طیفی، ضرایب کپسترم بیشترین کاربرد را در بازشناسی گفتار و گوینده دارند [۶]. این مطلب به‌خاطر تحلیل‌پذیری رفتار پوش طیف نسبت به ضرایب کپسترم، امکان لیفتر کردن[9] آنها، تفسیر مناسب در تعریف تابع فاصلهٔ اقلیدسی بر روی آنها، بازدهی مناسب در بازشناسی گفتار و گوینده و سهولت محاسباتی در تخمین آنهاست. با توجه به مطالب بیان شده، در طی این تحقیق از ضرایب کپسترمLPC[10] استفاده کرده‌ایم [۶]. البته می‌توان از ضرایب کپسترم FFT[11] نیز استفاده کرد [۸ و ۹]. ولی چون روش تحلیل LPC، اطلاعات پیکهای پوش طیف را بهتر مدل کرده (و بعضاً تقویت نیز می‌کند) و همچنین بار محاسباتی آن نیز کمتر است، و روابط برگشتی ساده بین ضرایب فیلتر LPC و ضرایب کپسترم وجود دارد، نوعاً از ضرایب کپسترم LPC برای اهداف بالا استفاده می‌شود. درجهٔ فیلتر LPC در کاربردهای بازشناسی گفتار برابر ۸ تا ۱۰ و همچنین تعداد ضرایب کپسترم نوعاً ۱/۵ برابر درجه فیلتر LPC در نظر گرفته می‌شود [۱۰]. علت این امر آن است که در بازشناسی گفتار توجه عمده بر آن است که ویژگیهای استخراج شده حتی‌الامکان اطلاعات زبان را در برداشته و نسبت به تغییرات ناشی از گوینده (یا سیستم انتقال) حتی‌المقدور نرمالیزه باشند. ولی در بازشناسی گوینده ما به اطلاعات وابسته به گوینده بیشتر علاقه‌مند هستیم. در نتیجه به‌نظر می‌رسد که با افزایش درجهٔ فیلتر LPC (و در نتیجه افزایش تعداد ضرایب کپسترم) می‌توان اطلاعات بیشتری از گوینده را در ضرایب کپسترم به‌حساب آورد. برای مثال با افزایش مناسب و کنترل شدهٔ درجهٔ فیلتر، صفرهای فیلتر دستگاه صوتی و پهنای باند فرمنتها بهتر مدل می‌شوند. صفرهای فیلتر صوتی و پهنای باند فرمنتها (قطبهای فیلتر صوتی) از جمله اطلاعات خوب و مفید برای تمایز گویندگان هستند. افزایش نامناسب درجهٔ فیلتر LPC موجب می‌شود که پارامترهای منتج، شامل بخشی از اطلاعات هارمونیکهای طیف شوند. این امر برای بازشناسی گفتار و گوینده مناسب نیست. زیرا اطلاعات هارمونیکها جداگانه توسط فرکانس پیچ مورد توجه قرار گرفته است. با شبیه‌سازیهای زیادی که طی این تحقیق انجام پذیرفته، مشخص شده است که اگر درجهٔ فیلتر p کمتر از نصف پریود پیچ انتخاب شود، مشکل ذکر شده به‌وجود نخواهد آمد. از طرفی اگر فرکانس نمونه‌برداری را ۸KHZ در نظر بگیریم، حدود ۴ تا ۵ فورمنت (قطب) و حداکثر ۲ صفر در مشخصهٔ فرکانسی فیلتر دستگاه صوتی خواهیم داشت. اگر شیب سرتاسری طیف را هم توسط حداکثر یک فیلتر درجهٔ ۲ مدل کنیم، به‌نظر می‌رسد که انتخاب درجهٔ فیلتر در گسترهٔ $12 \leq p \leq 16$ محدودهٔ مناسبی برای بازشناسی گوینده خواهد بود. حال فرض می‌کنیم که $p = 16$ باشد، در این صورت پریود پیچ باید بیشتر از ۳۲ نمونه باشد. اگر فرض کنیم کمترین پریود پیچ متکلم (بچه‌ها) برابر ۲/۵ms (یعنی ۴۰۰HZ) باشد، در این صورت حداقل پریود پیچ برابر ۲۰ نمونه خواهد بود. (در فرکانس نمونه‌برداری ۸KHZ) مشاهده می‌شود که اگر پریود پیچ بزرگتر از ۴ms (۳۲ نمونه یا ۲۵۰HZ) باشد، انتخاب درجهٔ فیلتر $p = 16$ مشکلی به‌وجود نخواهد آورد. ولی چنانچه پریود پیچ گوینده کمتر از ۴ms شود (زنها و بچه‌ها)، ضرورت دارد که تمهیدات خاصی اندیشیده شود. در طی این تحقیق وقتی پریود پیچ کمتر از ۴ms شود، ابتدا با استفاده از درونیابی با باند محدودBLI[12]، طول یک بلوک (حدود ۱۰ تا ۱۵ms) حول فریم تحلیل را به نسبت (عدد ۴ms بر طول پریود پیچ فریم فعلی بر حسب ms) افزایش می‌دهیم .آنگاه تحلیل LPC برای استخراج ضرایب فیلتر را بر روی فریم درونیابی شده انجام می‌دهیم. در عمل به‌خاطر امکان ایجاد ناپایداری در تخمین ضرایب فیلتر LPC تحلیل LPC را در دو مرحله انجام می‌دهیم. در مرحلهٔ اول از یک فیلتر درجه دو $p_1 = 2$ استفاده می‌کنیم . بعد از اعمال فیلتر معکوس درجه دو به نمونه‌های سیگنال، آنگاه از یک فیلتر درجه $p_2 = 14$ برای تحلیل سیگنال باقیمانده استفاده می‌کنیم. با استفاده از روابط برگشتی بین ضرایب فیلتر LPC و ضرایب کپسترم [۶]، ۲۴ ضریب کپسترم را به‌دست می‌آوریم. در سیستمهای بازشناسی گفتار پس از تخمین ضرایب کپسترم، آنها را در یک پنجرهٔ مناسب (مانند پنجرهٔ جانگ یامیر [۱۰]) ضرب کرده و سپس مورد استفاده قرار می‌دهند. یعنی اصطلاحاً آنهارا لیفتر می‌کنند. با توجه به اینکه قسمتهای لیفتر شده تا حدودی اطلاعات مربوط به گوینده را در بردارد، بنابراین در سیستمهای تشخیص و تصدیق گوینده، ضرورتی بر اعمال لیفتر نیست.

## ۳- نحوهٔ استفاده از پارامترهای سیگنال تحریک

طول فریم تحلیل برای سیستمهای بازشناسی گفتار و گوینده



[ DOR: 20.1001.1.2251600.1378.18.1.3.4 ]

نوعاً بین ۵ تا ۲۰ms در نظر گرفته می‌شود در طی این تحقیق ما طول فریم تحلیل را برابر ۱۲/۵ms (برابر بالاترین پریود پیچ سیستمهای عملی) در نظر گرفتیم. در سیستمهای بازشناسی گوینده که از روش چندی کردن برداری استفاده می‌کنند، به‌کارگیری سیگنال تحریک (فرکانس پیچ و گین) به نحو مطلوب وجود ندارد . حال فرض می‌کنیم f(n+i) و g (n+i) و X ( n+i)، (i = ۰ تا k)، به‌ترتیب فرکانس پیچ، گین و پارامترهای طیفی k فریم متوالی بعد از فریم nام باشند. در روش چندی‌کردن ماتریسی، برای هر فریم n، پارامترهای k فریم متوالی به‌طور یک‌جا و به‌صورت ماتریسی کد می‌شوند. برای این منظور ابتدا با استفاده از پارامترهای طیفی وتحریک یک بردار تعمیم یافته به‌صورت [f(n+i) و g (n+i) و X (n+i)] = Y (n+i) تشکیل می‌دهیم. یعنی اگر بردار X (n+i) دارای بعد p باشد، بردار Y (n+i) دارای بعد p+۲ خواهد بود.

### ۳-۱- نرمالیزه کردن پارامترهای تحریک

می‌دانیم که گین و فرکانس پیچ یک گوینده دارای تغییرات زیاد و بعضاً شدیدند. بنابراین بین گسترهٔ دینامیکی تغییرات گین و فرکانس پیچ گویندگان مختلف، همپوشانی زیادی وجود دارد. در نتیجه، استفاده از مقدار مطلق گین و فرکانس پیچ نتایج مفیدی برای تشخیص و تصدیق گوینده به‌بار نمی‌آورد. از طرفی اگر از فرکانس پیچ مطلق برای جداسازی گویندگان مختلف استفاده کنیم، امکان فریب سیستم به‌روشهای تقلید ماهرانه وجود دارد. برای رفع مشکلات ذکر شده در هنگام استفاده از پارامترهای تحریک، آنها را به روش مناسبی که بیان می‌شود نرمالیزه می‌کنیم: ۱- ابتدا فرکانس پیچ کلیه فریمهای بی‌صدا[۱۳] را برابر صفر قرار می‌دهیم. ۲- فرکانس پیچ حداکثر در طی بلوک nام (که شامل k فریم متوالی است) را محاسبه می‌کنیم. ۳- مقادیر فرکانس پیچ در طول بلوک nام را بر مقدار حداکثر آن تقسیم می‌کنیم. ۴- مقدار متوسط کانتور فرکانس پیچ نرمالیزه شده را به‌دست می‌آوریم. ۵- مقدار متوسط کانتور فرکانس پیچ نرمالیزه شده را از مقادیر کانتور کم می‌کنیم. ۶- مقادیر کانتور جدید (نرمالیزه شده نسبت به فرکانس پیچ حداکثر و همچنین نرمالیزه شده نسبت به مقدار متوسط کانتور پیچ) به‌علاوه مقادیر حداکثر و متوسط کانتور پیچ به‌عنوان پارامترهای متناظر با فرکانس پیچ مورد استفاده قرار می‌گیرند. ۷- کلیه مراحل ۲ تا ۶ را برای کانتور گین بلوک نیز انجام می‌دهیم. در این حالت به‌جای استفاده از گین مطلق از لگاریتم (در مبنای ۱۰) گین استفاده می‌کنیم. البته مقدار حداکثر گین در بلوک نرمالیزه شده متناظر گینها دخالت داده نمی‌شود (برخلاف کانتور پیچ). بنابراین به‌ازای هر بلوک k فریمی، تعداد k+۲ پارامتر مرتبط با فرکانس پیچ و ۱ + k پارامتر مرتبط با گین (لگاریتم گین)، علاوه بر k.p تعداد پارامتر طیفی خواهیم داشت (p برابر تعداد ضرایب طیفی و k تعداد فریمهای هر بلوک است).

## ۴- تعریف تابع فاصلهٔ مناسب

برای استفاده از روش چندی‌کردن ماتریسی (یا برداری) در تشخیص یا تصدیق گوینده، ضرورت دارد که یک تابع فاصلهٔ مناسب برای مقایسه بلوکهای سیگنال (هم در فاز کد کردن و هم در فاز تولید الگوهای مرجع) تعریف شود. ساده‌ترین و در عین حال عمومیترین تابع فاصله برای تعیین مقدار فاصلهٔ بین دو بردار $X = (x_1, \dots, x_p)$ و $Y = (y_1, \dots, y_p)$ تابع فاصلهٔ اقلیدسی، معادلهٔ (۱) است [۱۱ و ۱۲]

$$d(x,y) = \sum_{i=1}^{p} (x_i - y_i)^2 \qquad (1)$$

در مواقعی که جنس و ارزش پارامترهای $x_i$ ($i = 1$ تا $p$) از یک نوع و مساوی بوده و همچنین گسترهٔ دینامیکی (واریانس آنها) یکسان باشند، استفاده از تابع فاصلهٔ اقلیدسی به نتایج مناسبی منجر می‌شود. متأسفانه در اغلب کاربردها شرایط بالا برقرار نیست. برای مثال، در کاربرد اخیر، p پارامتر اول از نوع کپسترم و با گسترهٔ دینامیکی متفاوت و همچنین فرکانس پیچ و گین از جنس متفاوت و هم دارای گسترهٔ دینامیکی متمایز با همدیگر و همچنین با ضرایب کپسترم‌اند. روش کلاسیک برای یکسانسازی گسترهٔ دینامیکی پارامترها، نرمالیزه کردن آنها نسبت به مقادیر متوسط واریانس است. فرض می‌کنیم: $\sigma_x = [\sigma_{x1}, \dots, \sigma_{xp}]$ و $\mu_x = [\mu_{x1}, \dots, \mu_{xp}]$ مقادیر انحراف معیار و میانگین بردارهای $X = (x_1, \dots, x_p)$ باشند. در این‌صورت بردار نرمالیزه شده به میانگین و واریانس به‌صورت $Y = (y_1, \dots, y_p)$ تعریف می‌شود که $y_i$ ها با استفاده از معادلهٔ زیر به‌دست می‌آیند

$$y_i = \frac{x_i - \mu_{xi}}{\sigma_{xi}} \quad ; \quad i = ۱ \text{ تا } p \qquad (۲)$$

حال فرض می‌کنیم که $X = [x(n),... x(n+k)]$ یک بلوک طیفی k فریمی نرمالیزه شده نسبت به میانگین و واریانس پارامترهای طیفی باشد. همچنین $F(n) = [f(n) , ... f_{ave}(n)]$, $f_{max}(n)$ و $f(n+k)$ ، $G(n) = [g(n) , ... f(n+k)$ بردار نرمالیزه شده کانتور فرکانس پیچ و گین متناظر با بلوک طیفی x باشند. برای اینکه بتوانیم مجموعه پارامترهای فوق‌الذکر را نسبت به یکدیگر ارزش‌گذاری جداگانه‌ای کنیم، تابع فاصلهٔ اقلیدسی وزن داده شده[۱۴] بین دو بلوک متمایز Z و Y را با معادلهٔ زیر تعریف می‌کنیم:

$$d(Y,Z) = \alpha_۱^۲ . d_S(Y,Z) + \alpha_۲^۲ . d_F(Y,Z) + \alpha_۳^۲ . d_G(Y,Z) \qquad (۳)$$

که در آن داریم:

$$d_S(Y,Z) = \frac{۱}{p} \sum_{i=۱}^{k} w_i^۲ . [\sum_{j=۱}^{p} (y_{ij} - z_{ij})^۲] \qquad (۴)$$

$$d_F(Y,Z) = \frac{۱}{k+۲} \sum_{i=۱}^{k+۲} (fy_i - fz_i)^۲ \qquad (۵)$$

$$d_G(Y,Z) = \frac{۱}{k+۱} \sum_{i=۱}^{k+۱} (gy_i - gz_i)^۲ \qquad (۶)$$

$$\alpha_۱ + \alpha_۲ + \alpha_۳ = ۱ \qquad (۷)$$

$$\sum_{i=۱}^{k} w_i = ۱ \qquad (۸)$$

$$w_i = \frac{gzi}{\sum_{j=۱}^{k} gzj} \qquad (۹)$$

در معادله‌های بالا، $d_S(Y,Z)$ فاصلهٔ بین پارامترهای طیفی دو بلوک و $d_F(Y,Z)$ فاصلهٔ بین پارامترهای مربوط به فرکانس پیچ دو بلوک و $d_G(Y,Z)$ فاصلهٔ بین پارامترهای مربوط به گین (لگاریتم گین) دو بلوک Z و Y هستند. در تعیین فاصله بین پارامترهای طیفی دو بلوک از یک بلوک وزنی $w_i$ ($i = ۱$ تا k) استفاده کردیم که متناسب با لگاریتم گین فریمهای داخل بلوک‌اند. این وزنها به‌خاطر آن به پارامترهای طیفی اعمال شده، که از بایاس شدن نتیجهٔ نهایی به‌خاطر وجود نویزهای زمینه جلوگیری شود، چون گین نویزهای زمینه نسبت به قسمتهایی که سیگنال صحبت داریم کمتر است. همچنین با این وزنها به قسمتهای با صدای صحبت (که نوعاً دارای گین بیشتری هستند) نسبت به قسمتهای بی‌صدای صحبت وزن بیشتری داده می‌شود. وزنهای $\alpha_۱$ و $\alpha_۲$ و $\alpha_۳$ با استفاده از روش DA[15] [۱۰، ۱۳ و ۱۴] و به‌طور تجربی تعیین می‌شوند. با استفاده از تحلیل DA و برای کاربردهای این تحقیق (با داده‌های استفاده شده) مقادیر وزنهای بالا به‌شرح $\alpha_۱ = ۰/۷$ و $\alpha_۲ = ۰/۱۸$ و $\alpha_۳ = ۰/۱۲$ تخمین زده شدند (برای کاربردهای مختلف می‌بایستی جداگانه محاسبه شوند).

## ۵- نحوهٔ محاسبهٔ تابع درستنمایی

فرض می‌کنیم $C_i$ کتاب کد ماتریسی متناظر با گوینده مجاز iام باشد. هر کتاب کد $C_i$ ($i=۱$ تا M) شامل $N_i$ الگوی مرجع ماتریسی $C_{ij}$ است. ($j=۱$ تا N) . حال فرض می‌کنیم Y(n) ($n=۱$ تا L) دنبالهٔ بلوکها (یا کلمه) بیان شده از طرف متکلم باشد (یعنی ابتدا تا انتهای جمله شامل L بلوک باشد). در مرحلهٔ تشخیص گوینده، برای هر بلوک ورودی Y(n) ، نزدیکترین الگو از میان هر یک از M کتاب کد را به‌دست می‌آوریم (چون تعداد گویندگان مجاز برابر M فرض شده است پس Mکتاب کد ماتریسی خواهیم داشت). فاصلهٔ بین بلوک ورودی Y(n) و نزدیکترین الگوی متناظر در کتاب کد $C_i$ ($i=۱$ تا M) را $d(Y(n), C_{ij})$ می‌نامیم. تابع اعوجاج تجمعی وزن داده شده را به‌صورت زیر به‌عنوان تابع درستنمایی تعریف کرده و مورد استفاده قرار می‌دهیم

$$D(i) = \sum_{n=۱}^{L} u_n^۲ . d(Y(n), C_{ij}) \qquad (۱۰)$$

که در آن وزنهای $u_n$ به‌صورت زیر از روی گین حداکثر هر بلوک محاسبه می‌شود:

$$u_n = \frac{gmax(n)}{\sum_{j=۱}^{L} gmax(j)} \qquad (۱۱)$$

فرض می‌کنیم در پایان جمله، کمترین مقدار D(i) (i = ۱ تا M) برابر $D_{min}$ باشد. اگر $D_{min}$ از یک سطح آستانهٔ مشخص مانند TD کمتر باشد، شخص متکلم یا مدعی به‌عنوان فرد مجاز در نظر گرفته می‌شود، در غیر این‌صورت به‌عنوان فرد غیرمجاز معرفی خواهد شد. چنانچه هویت شخص گوینده نیز مورد نظر باشد، ایندکس کتاب کد متناظر با $D_{min}$ به‌عنوان شخص متقاضی اعلام خواهد شد. همان طوری‌که ملاحظه می‌شود، مقدار سطح آستانه TD یک پارامتر کلیدی در نتیجهٔ عملکرد سیستم است.

## ۶- ملاحظات نظری و عملی در تعیین سطح آستانهٔ TD

فرض می‌کنیم که سیستم برای M فرد مجاز طراحی شده است. در نتیجه M کتاب کد ماتریسی $C_i$ (i = ۱ تا M) خواهیم داشت. سطح آستانهٔ بهینهٔ TD مقداری است که کمترین خطای ممکن برای کتابهای کد $C_i$ (i = ۱ تا M) را در مرحلهٔ تصدیق گوینده تولید کند. یعنی اگر گوینده مجاز j یک جمله مورد قبول را بیان کرده‌است؛ اولاً سیستم او را تصدیق کند (یعنی مقدار اعوجاج تجمعی متناظر او کمتر از TD شود). ثانیاً ـ ایندکس کتاب کد آشکارسازی شده برای j شود (یعنی فرایند تعیین هویت فردگوینده صحیح انجام پذیرد). ثالثاً اگر شخص گوینده فردی خارج از مجموعهٔ افراد مجاز باشد، مقدار اعوجاج تجمعی متناظر با آن بیشتر از سطح آستانهٔ TD شود. بنابراین عملکرد سیستم به‌شرح زیر را همراه با خطا می‌نامیم: ۱- گوینده از افراد مجاز بوده ولی سطح آستانهٔ تجمعی از TD بیشتر شود. ۲- گوینده از افراد مجاز بوده و سطح آستانه تجمعی نیز از TD کمتر شده است ولی هویت گوینده اشتباه نسبت داده شود. ۳- گوینده از افراد غیرمجاز باشد ولی سطح آستانه تجمعی کمتر از TD شود. ۴- گوینده غیرمجاز با تقلید ماهرانه قادر به تولید سطح آستانه تجمعی کمتر از TD باشد. ملاحظه می‌شود که انتخاب سطح آستانهٔ بهینهٔ TD یعنی انتخاب مقداری است که کمترین خطای ممکن را تحت شرایط ذکر شده ایجاد کند. انتخاب سطح آستانهٔ بهینهٔ TD به‌روش تحلیلی کاری مشکل است. در عمل به‌روش سعی و خطا و با استفاده از پایگاههای داده مناسب مقدار مناسب TD را تعیین می‌کنند. روشی که در طی این تحقیق برای تخمین مقدار اولیه TD انجام پذیرفته، استفاده از روش TCPA-AS[16] است که در طی این تحقیق ارائه و مورد استفاده قرار گرفته است. جزییات روش بالا در قسمت مربوط به طراحی کتابهای کد ماتریسی توضیح داده خواهد شد. فرض می‌کنیم TR مقدار سطح آستانهٔ تعیین شده برای تولید کتابهای کد ماتریسی افراد مجاز (از روی جملات آموزشی) به روش TCPA-AS باشد. در این‌صورت سطح آستانهٔ تجمعی را در محدودهٔ TR ≤ TD ≤ ۳TR در نظر می‌گیریم. آنگاه به روش سعی و خطا، مقداری از TD را تخمین می‌زنیم که سیستم تصدیق گوینده کمترین خطا را تولید کند.

## ۷- روش تولید کتابهای کد ماتریسی افراد مجاز

تولید کتابهای کد ماتریسی یکی از مراحل اساسی فاز آموزش سیستم تصدیق گوینده به روش چندی کردن ماتریسی است. در روشهای کلاسیک برای تولید کتاب کد ماتریسی ، از همان روشهای طراحی کتاب کد برداری [۱۵ و ۱۶] استفاده می‌کنند. مهمترین روشهای طراحی کتاب کد، روش LBG[17] و روش COCA[۱۷][18] هستند. دقت روش COCA بیش از روش LBG بوده و همچنین عملکرد نهایی آن تا حدود زیادی مستقل از انتخاب شرایط اولیه است. در عوض حجم محاسبات آن نسبت به روش LBG بیشتر است.

اگر بخواهیم از روش LBG برای طراحی کتاب کد استفاده کرده و تقریباً همیشه به یک نتیجهٔ یکسان برسیم، لازم است که از روش BS[19] برای انتخاب شرایط اولیه استفاده کنیم. استفاده از روش BS-LBG یا روش COCA هر دو مستلزم اطلاع از $N_i$ (i = ۱ تا M) تعداد الگوهای مرجع کتابهای کد است. علاوه بر آن در روش BS-LBG ضرورت دارد که $N_i$ها به‌صورت توانی از ۲ باشند. استفاده از دو روش بالا برای تولید کتاب کد در سیستمهای بازشناسی گفتار یا کدکننده‌های برداری (یا ماتریسی) امری رایج است. این امر به‌خاطر آن است که عملکرد این سیستمها وابسته به سطح آستانه‌ای مانند TD (که در سیستم تصدیق گوینده به آن نیازمندیم) نیست. بنابراین با انتخاب مناسب N (به‌صورت تجربی) می‌توان کتاب کد مناسبی برای این سیستمها طراحی کرد. اما متأسفانه اعمال روش فوق برای سیستمهای تصدیق گوینده نتایج جالبی به‌بار نمی‌آورد. یا حداقل سیستم طراحی شده چندان مقاوم نبوده و امکان افزایش خطا با تغییر زمان بازشناسی و تغییر شرایط مختلف به‌شدت افزایش می‌یابد [۱۹]. برای حل مشکل مذکور،

روشی بنام APCT[20] در طی این تحقیق ارائه و مورد استفاده قرار دادیم. فرض می‌کنیم $X_1$ , ... , $X_L$ ماتریسهای پارامترهای آموزشی مربوط به یک گویندهٔ مجاز باشند (با استفاده از جملات آموزشی بیان شده توسط گویندهٔ مجاز تولید می‌شوند). روش TCPA برای تولید الگوهای مرجع ماتریسی مربوط به این گوینده به‌شرح زیر است:

۱- اولین الگوی آموزشی را به‌عنوان اولین الگوی مرجع منظور کرده و تعداد الگوهای مرجع را برابر یک قرار می‌دهیم.

۲- فاصله نزدیکترین الگوی مرجع (از میان N الگوی مرجع تولید شده تاکنون) به الگوی آموزشی بعدی را به‌دست می‌آوریم. اگراین فاصله از سطح آستانهٔ TR بیشتر باشد، یک کد جدید تولید می‌شود (یعنی $N \leftarrow N+1$ ، $C(i) \leftarrow X_j$).

۳- اگر فاصلهٔ مذکور از سطح آستانهٔ TR کمتر باشد، کدی تولید نخواهد شد و عملیات مرحلهٔ (۲) برای الگوی آموزشی بعدی ادامه خواهد یافت.

۴- فرایند مرحلهٔ (۲) و (۳) تا اتمام کلیه الگوهای آموزشی ادامه می‌یابد.

همان طوری که ملاحظه می‌شود N یعنی تعداد الگوهای مرجع تولید شده در روش TCPA به سطح آستانه TR و به حجم و تنوع الگوهای آموزشی وابسته است. بنابراین مثل روشهای BS-LBG یا COCA از قبل تعریف شده و معین نیست. برای تعیین تجربی مقدار TR از روش تحلیل و سنتزبرداری به‌شرح زیر استفاده کردیم. ابتدا TR رامساوی یک مقدار حداقل قرار می‌دهیم. با استفاده از سطح آستانهٔ بالا برای کلیه گویندگان مجاز تعدادی کد تولید می‌کنیم. با استفاده از کتابهای کد تولید شده، کلیه جملات آموزشی و آزمون را به‌روش برداری تحلیل و سنتز می‌کنیم. با استفاده از آزمونهای شنیداری کیفیت سیگنال بازسازی شده را هم از نظر تشخیص گوینده و هم از نظر فهم کلام ارزیابی می‌کنیم. برای این منظور از آزمونهای شنیداری نظیر DRT یا MOS یا DAM می‌توان استفاده کرد [۱۸]. با توجه به اینکه تمامی اطلاعات و پارامترهای لازم برای سنتز (یعنی گین، فرکانس پیچ و پارامترهای طیفی شامل ضرایب کپسترم یا ضرایب فیلتر LPC) در اختیار است، به‌راحتی می‌توان فرایند سنتز را انجام داد. فریمهایی که دارای فرکانس پیچ صفرند، به‌عنوان فریمهای بی‌صدا منظور شده و پریود پیچ آن را در هنگام سنتز ۱۰ میلی ثانیه در نظر می‌گیریم. نتایج تجربی حاصل در طی این تحقیق نشان می‌دهد که اگر $0/1 \geq TR \geq 0/01$ انتخاب شود، جملات سنتز شده (با استفاده از کتابهای کد تولید شده) دارای قابلیت فهم و کیفیت تشخیص گویندهٔ خیلی خوب تا خوب‌اند. محدوده مذکور برای TR با فرض این مطلب است که تمامی پارامترها نسبت به مقدار میانگین و واریانس خود مطابق معادلهٔ (۲) نرمالیزه شده‌اند. انتخاب TD کوچکتر موجب می‌شود که سیستم در مقابل تقلید بسیار مقاومتر شده و همچنین خطای سیستم در تعیین هویت کمتر شود. مشکل انتخاب TD خیلی کوچک آن است که وقتی سیستم با تعداد جملات کم آموزش داده شود و یا تنوع حالات بیانی گوینده در هنگام آموزش کم باشد و یا انجام آموزش در پریودهای زمانی مختلف و متنوع و طولانی انجام نپذیرد ممکن است شخص گوینده مجاز به‌عنوان یک شخص غیرمجاز معرفی شود. بنابراین در طی این تحقیق بر مبنای تجربیات شبیه‌سازی و برای ایجاد یک توازن مناسب $TD = 0/05$ و $TR = 0/03$ انتخاب شدند.

## ۸- نتایج شبیه‌سازی

برای ارزیابی کارایی روشهای پیشنهادی، شبیه‌سازیهایی به‌شرح زیر انجام پذیرفته است. سیستم برای حالت وابسته به متن طراحی شده است. متن مورد استفاده شامل ۱۰ جملهٔ مشخص و ۱۰ کلمهٔ صفرتا ۹ بوده است. سیستم برای ۳۰ نفر گوینده طراحی شده است. ۲۰ نفر از گویندگان مرد (بین ۱۶ تا ۶۰ سال) و ۱۰ نفر بقیه زن (بین ۱۵ تا ۵۰ سال) بوده‌اند. هر گوینده هر کلمه یا جمله را با ۵ حالت بیانی تقریباً متمایز و ۴ بار تکرار بیان کرده است. فرایند ضبط جملات و کلمات برای هر گوینده ۵ بار (طی دو ماه) انجام پذیرفته است. از مجموع ۴ بار تکرار کلمات و جملات ضبط شده (طی ۵ دوره زمانی متمایز) دو تکرار برای آموزش سیستم و دو تکرار برای آزمون و ارزیابی مورد استفاده قرار گرفت. ضبط کلمات و جملات تحت شرایط عادی $S/N \geq 24$ dB انجام پذیرفته است. فرکانس نمونه‌برداری ۸ KHZ و نمونه‌های سیگنال به‌صورت ۲ بایتی بوده‌اند. برای ضبط دیجیتالی از ساندبلاستر تجاری استفاده کردیم. در تمامی مراحل ضبط کلمات و جملات، AGC[21] سیستم ساندبلاستر روشن بوده است. طول فریم تحلیل ثابت و برابر ۱۲/۵ ms

جدول ۱- خطای تصدیق گوینده (برحسب درصد) با بیان کلمات صفر تا ۹

| ۱کلمه | ۲کلمه | ۳کلمه | ۴کلمه | ۵کلمه | بیشتر از ۵کلمه |
|---|---|---|---|---|---|
| ۲/۱ | ۱/۳۴ | ۰/۷۵ | ۰/۱۱ | ۰ | ۰ |

جدول ۲- خطای تصدیق گوینده (برحسب درصد) با بیان جملات مجاز

| یک جمله | دو جمله | سه جمله | بیش‌از سه جمله |
|---|---|---|---|
| ۰/۲۲ | ۰ | ۰ | ۰ |

و طول پنجرهٔ تحلیل متغیر و برابر ۳-۲/۵ برابر پریود پیچ بوده است. طول پنجرهٔ تحلیل برای فریمهای بی‌صدا مقداری ثابت و برابر ۲۵ms بوده است. پنجرهٔ مورد استفاده از نوع همینگ و تحلیل LPC از نوع اتوکرلیشن بوده است. اعمال تحلیل LPC با درجهٔ ۱۶ گاه موجب ناپایداری در تخمین ضرایب فیلتر پیشگویی شده‌است. برای رفع این مشکل، تحلیل LPC را در طی دو مرحله انجام دادیم. در مرحله اول از یک فیلتر پیشگویی درجهٔ دو استفاده کردیم. پس از اعمال فیلتر معکوس درجهٔ دو به نمونه‌های پنجره شدهٔ سیگنال، مجدداً از یک فیلتر پیشگویی درجهٔ ۱۴ برای تخمین ضرایب فیلتر استفاده کردیم. با استفاده از روابط برگشتی بین ضرایب فیلتر LPC و ضرایب کپسترم، به‌ترتیب ۳ و ۲۱ ضریب کپسترم را از روی ضرایب فیلتر درجهٔ ۲ و ۱۴ به‌دست آوردیم (در مجموع ۲۴ ضریب کپسترم). برای تخمین گین فریمها از معادلهٔ زیراستفاده کردیم:

$$g = \frac{1}{I}\sum_{i=1}^{I} \left| x(i) \right| \qquad (۱۲)$$

که در آن x(i) یک قطعهٔ پنجره شدهٔ سیگنال و I تعداد نمونه‌های پنجره است. برای تخمین فرکانس پیچ از روش دقیق و مقاوم NCC[۲۲] مندرج در مرجع [۲۰] استفاده کردیم. با توجه به نتایج تحقیقات مندرج در مرجع [۱۹] که اذعان می‌دارد طول مناسب برای استفاده از تغییرات زمانی[۲۳] در سیستمهای چندی کننده ماتریسی ۶۰ ms تا ۱۲۰ است، تعداد فریمهای هر بلوک برای چندی کردن ماتریسی را برابر $k = 8$ در نظر گرفتیم (یعنی طول قطعهٔ سیگنال ۱۰۰ ms است). مقدار سطح آستانه TD برای تصدیق گوینده را برابر ۰/۰۵ و سطح آستانه TR برای تولید کتابهای کد ماتریسی را برای ۰/۰۳ در نظر گرفتیم. جدول (۱) نتایج حاصل برای تصدیق گوینده با فرض اینکه گوینده صرفاً اعداد بین صفر تا نه را بیان کرده است را نشان می‌دهد. همان طوری‌که از جدول (۱) ملاحظه می‌شود، برای اینکه سیستم بدون خطا عمل کند ضرورت دارد که گوینده حدود ۴ تا ۵ کلمه به‌طور متوالی بیان کند (البته مقدار بالا به تعداد گویندگان مجاز سیستم هم بستگی دارد). تمامی خطاهای اتفاق افتاده مربوط به متمایزسازی افراد از یک جنس بوده است (حدود ۷۰٪ مربوط به مردهاو ۳۰٪ مربوط به زنها). در آزمون دوم از جملات برای تصدیق گوینده استفاده کردیم (گوینده یک یا چند جمله از مجموع ۱۰ جمله مجاز را بیان می‌کند). جدول (۲) نتایج حاصل را نشان می‌دهد. با ملاحظهٔ جدول (۲) مشاهده می‌شود که خطای سیستم با دریافت صرفاً یک جمله حدوداً ۰/۲۲ درصد است (بسیار ناچیز) و با افزایش جملات به دو جمله، خطای سیستم صفر شده است. یعنی هر گوینده با بیان دو جمله (از میان جملات مجاز) مورد تصدیق یا رد سیستم قرار می‌گیرد. خطای سیستم در این‌حالت برابر صفر شده است. نتایج حاصل نمایانگر قابلیت روش پیشنهادی برای سیستمهای تصدیق گوینده است. در آزمون سوم، از صدای یک فرد که تلاش کرده است با تقلیدماهرانه صدای یکی از افراد مجاز (مرد) را ایجاد کند استفاده کردیم. برای تمامی کلمات و جملات تقلید شده، سیستم شخص گوینده را به عنوان فردی غیرمجاز اعلام کرده است. البته برای ارزیابی دقیق کارایی این قسمت از سیستم، به داده‌های بیشتری (تقلید ماهرانهٔ صدای افراد) نیازمندیم که در طی این تحقیق در اختیار ما نبوده است.

## ۹- جمعبندی و نتیجه‌گیری

در طی این تحقیق روشهایی برای بهبود عملکرد سیستمهای تصدیق گوینده (امضای گفتاری) ارائه شد. این بهبود با توجه به

عوامل زیر حاصل شده است: الف - تمامی پارامترهای دخیل در تولید سیگنال گفتار برای بازنمایی پارامتر یک مشخصهٔ گوینده به نحو مناسب و شایسته مورد استفاده قرار گرفت. ب - اطلاعات مربوط به گوینده که در پارامترهای استخراج شده نهفته است به نحو مناسبی نگهداری و مورد استفاده قرار دادیم (برخلاف سیستمهای بازشناسی گفتار که با هموارسازی پارامترها و کوچک در نظر گرفتن درجهٔ فیلتر پیشگویی، اطلاعات گوینده را تا حدودی از دست می‌دهند). ج - فضای پارامتر یک مربوط به هر گوینده با دقت[۲۴] زیاد توسط چندی کردن ماتریسی گسسته شده است. این امر توسط ارائه روش مناسبی بنام TCPA برای تولید کتابهای کد ماتریسی حاصل شده است. د - تابع درستنمایی مورد استفاده نسبت به نویزهای زمینه حساسیت کمتری دارد (به‌خاطر استفاده از پارامترهای وزنی متناسب با گین فریمها). هـ- آموزش سیستم با دقت و ملاحظات عملی مناسب انجام پذیرفته است.

با توجه به نتایج حاصل، به‌نظر می‌رسد که استفاده از روشهای پیشنهادی بهبود ارزنده‌ای در استفاده از سیستمهای تصدیق گوینده می‌تواند ایجاد کند.

## واژه نامه

1. dynamic time warping
2. robust
3. likelihood function
4. accumulative distortion function
5. cepstrum coefficient
6. line spectrum pair frequency
7. channel gain
8. filter bank analysis
9. lifter
10. linear predictive coding
11. fast fourier transform
12. band limited interpolation
13. unvoiced
14. weighted
15. discriminative analysis
16. threshold based code production algorithm analysis and synthesis
17. Linde, Bazzo, Gray
18. codebook optimization with Couchy anealing
19. binary splitting
20. threshold based code production algorithm
21. automatic gain control
22. normalized cross correlation
23. temporal variation
24. resolution

## مراجع


1. Atal B.S., "Automatic Recognition of Speakers from Their Voices," *IEEE Proc.* Vol. 64, pp. 460 - 475, 1976.
2. Rosenberg, A.E.,"Automatic Speaker Verification : A Review," *IEEE Proc.*, Vol. 64, pp. 475-487, 1976.
3. Dante, H.M., and Sarma, V.V., "Automatic Speaker Identification for A Large Number of Speaker," *ICASSP*, pp. 295-298, 1978.
4. Alku, P., "An Automatic Method to Estimate the Time Based Parameters of the Glotal Pulseform," *ICASSP*, pp. II 29-32, 1992.
5. Line, W.C., and Pillay, S.K., "Feature Evaluation and Selection for on-line Adaptive Speaker Verification System," *ICASSP*, pp. 734-737, 1976.
6. Furui, S., "Cepstral Analysis, Technique for Automatic Speaker Verification," *IEEE Trans.* ASSP, Vol. 29, No. 2, April 1981.
7. lia, C.S., Lin, M.T., and Wang, W.J., "Study of Line Spectrum Pair Frequencies for Speaker Recognition," *ICASSP*, pp. 277-280, 1990.
8. Rosenberg, A.E., and Soong, F.K., "Evaluation of a Vector Quantization Talker Recognition System in Text Independent and Text Dependent Modes," *Comput. Speech Language* Vol. 22, pp. 143-157, 1987.
9. Soong, F.K., Rosenberg, A.E., Rabiner, L.R., and Juang, B.H., "A Vector Quantization Approach to Speaker Recognition," *AT & T Tech. Journal* Vol. 66, pp. 14-26, 1987.
10. Rabiner, L., and Juang, B.H., *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.
11. Nocerino, N., Soong, F.K., Rabiner, L.R., and Klatt, D.H., "Comparative Study of Several Distortion Measures for Speech Recognition," *ICASSP*, pp. 25-28, 1985.
12. Hua, W.R., Shen, H.L., and Fujisaki, H., "A

Weighted Distance Measure Based on the Fine Structure of Feature Space : Application to Speaker Recognition," *ICASSP*, pp. 273-276, 1990.

13. Gish, H., "Robust Discrimination in Automatic Speaker Identification," *ICASSP*, pp. 289-292, 1990.
14. Johnson, R.A., and Wichern, D.W., *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey, 1988.
15. Gray, R.M., "Vector Quantization," *IEEE ASSP Magazine*, April 1984.
16. Juang, B.H., and Soong, F.K., "Speaker Recognition Based on Source Coding Approoches," *ICASSP*, pp. 613-616, 1990.
17. Wang, Z., and Hanson, J.V., "Codebook Optimization by Cauchy Annealing," Signal Processing Tech. and Applic., *IEEE Tech.* Activities Board, piscataway, New Jersey, 1995.
18. Furui, S., *Advances in Speech Signal Processing*, Marcel Dekker, Inc, Newyork, 1992.
19. Rosenberge, A.E., and Soong, F.K., "Recent Research in Automatic Speaker Recognition," *AT & T Bell lab. Tech. Report*, Murray Hill, New Jersey, 1996.
20. Medan, Y., Yair, E., and Chazan, D., "Super Resolution Pitch Determination of Speech Signals," *IEEE Trans. on ASSP*, Vol. 39, No. 1, Janu. 1991.

[ DOR: 20.1001.1.2251600.1378.18.1.3.4 ]