بازشناسی گفتار پیوسته فارسی با استفاده از الگوریتم ویتربی و خصوصیات پروزودیک کلمات

فرشاد الماس گنج
دانشجوی دکتری
دانشکده فنی و مهندسی دانشگاه تربیت مدرس

سید محمّدرضا هاشمی گلپایگانی
استاد
دانشکده مهندسی برق، دانشگاه صنعتی امیرکبیر

کارولوکس
استاد
دانشکده مهندسی برق، دانشگاه تهران

محمود بیجن خان
استادیار
دانشکده علوم انسانی، دانشگاه تهران

محمد حسن قاسمیان
استادیار
دانشکده فنی و مهندسی، دانشگاه تربیت مدرس

چکیده

موضوعی که در اینجا به آن می‌پردازیم، ارائه مدلی برای بازشناخت دنباله کلمات گفتار پیوسته فارسی است. این کار در چند مرحله صورت می‌گیرد. در اولین مرحله سیگنال گفتار به پنجره هایی از سیگنال شکسته می‌شود و به هر کدام از پنجره ها، یکی از واجهای زبان فارسی نسبت داده می‌شود. برای این کار از روش های مؤثری که قبلاً گزارش شده اند استفاده می‌شود. در مرحله بعدی عمل برش هجایی و تعیین واجهایی که در هر هجا قرار گرفته اند صورت می‌گیرد. واژگان مورد استفاده از ۱۲۰۰ کلمه تشکیل شده است که همگی از دادگان «فارس دات» انتخاب شده اند. در آخرین مرحله به کمک الگوریتم ویتربی و خصوصیات پروزودیک کلمات، بهترین دنباله کلمات موجود در گفتار تعیین می‌شود. نکته جدیدی که در این مدل گنجانیده شده است، استفاده از پروزودی برای افزایش درصد صحت بازشناخت کلمات و رفع ابهام بین کلماتی است که از لحاظ واج آرایی مشابه یکدیگرند، ولی با یکدیگر تفاوت دارند. نتایج آزمایش ها کارآئی مدل را نشان می‌دهند.

Using viterbi algorithm and prosodic features of words for speech recognition

F. Almasganj
Tarbiat Modarress University

S. M. Hashemi Golpayegani
Amirkabir University

K. Lucas
Tehran University

m. Bijankhan
Tehran University

M.S. Gasemian
Tarbiat Modaress University

Abstract

The purpose of our study is to design a wordspotting model for continuous speech recognition using both the sequence of the phonemes and prosodic features. The main problem is to alter the mistakes which usually occur in the recognition of the phoneme candidates. Each word is coded as a sequence of syllables in the vocabulary, each syllable has a maximum length of four phonemes. The vocabulary is consisted of 1200 words from the FARSDAT database. After extracting phonemes, the syllabification of the utterance is performed according to the syllable structure of Farsi. The sequence of the recognized syllables are fed into a module that extracts a lattice of the best possible words using a search algorithm. This algorithm is founded on two computations: 1) the phonemic similarity degree of the recognized sequence of the syllables with the segmental structure of the words in the vocabulary using phoneme confusion matrix. 2) The prosodic similarity of the recognized sequence of the syllables with the words in the vocabulary. Finally, the optimum path of words is derived from the lattice, using a viterbi algorithm.

واژه‌های کلیدی

بازشناسی گفتار پیوسته ـ بازشناسی کلمات ـ پروزودی

مقدمه

روش‌های زیادی برای بازشناسی کلمات گفتار پیوسته ارائه شده‌اند که هر کدام از مدل خاصی استفاده کرده‌اند. برخی روش‌های ریاضی، الگوریتم‌های جستجوی درختی با الگوریتم‌های جستجوی سریع در واژگان را انتخاب کرده [۹] [۱۰] [۱۱] و برخی دیگر از مدل‌های آماری مانند HMM در ساخت مدل کلمات و بازشناسی گفتار استفاده کرده‌اند [۱۲] [۱۳].

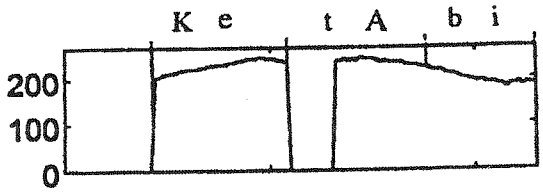
ما در مدل خود بعد از بازشناسی واج‌های گفتار، به جستجو در واژگان می‌پردازیم و پس از ساخت آرایه‌ای از بهترین کلمات، از الگوریتم ویتربی (Viterbi) برای انتخاب بهترین دنباله کلمات در گفتار پیوسته استفاده می‌کنیم.

در بین روش‌های بازشناسی کلمات، آنچه که معمولاً به آن توجه نشده است، استفاده از خصوصیات پروزودیک کلمات در حین بازشناسی آنها است. از ویژگی‌های اساسی در روانشناسی گفتار، زیر و بمی (pitch) و بلندی (loudness) است. اثرات این دو عامل در کنار اثرات ناشی از دو عامل دیگر سرعت و وزن (rhythm)، در مجموع خواصی هستند که به نام خصوصیات پروزودیک زبان شناخته می‌شوند [۶].

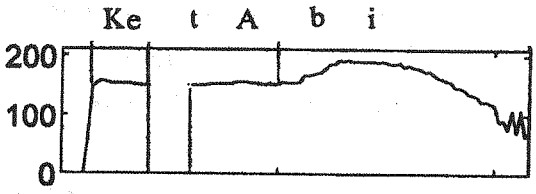
از سوی دیگر بسیاری از زبانشناسان، زبان فارسی را از نوع زبان‌های آهنگ آکسانی (pitch-accented) می‌دانند [۶] [۷]. در چنین زبانهایی الگوی زیر و بمی (pitch) روی واحدهای زبانی قابل پیش‌بینی است و تکیه (stress) از نوع زیر و بم ـ محور (pitch- oriented) است. بنابر این می‌توان تکیه هجاهای کلمات را از قبل پیش‌بینی نمود و نوع تکیه را باتوجه به منحنی مرز و زیر و بمی هجاها تعریف نمود.

باتوجه به مطالب فوق، ما نیز از منحنی‌های نشانگر مرز زیر و بمی هجاهای کلمات برای افزایش درصد صحت بازشناسی کلمات گفتار پیوسته کمک می‌گیریم.

مسئله دیگری که اهمیت پروزودی کلمات را در حین بازشناسی آنها دو چندان می‌کند، وجود کلماتی است که از نظر واج آرایی کاملاً یکسان هستند و تنها با تغییر محل هجای تکیه بر [۱]، یعنی هجایی از کلمه که بالاترین تکیه در آن واقع شده است، معنی آنها تغییر می‌کند. شکل ۱ منحنی مرز زیر و بمی کلمه «کتابی» را با تکیه روی هجای «تا» به مفهوم «یک کتاب» نشان می‌دهد:



شکل (۱) منحنی مرز زیر و بمی کلمه «کتابی» با تکیه روی هجای «تا»

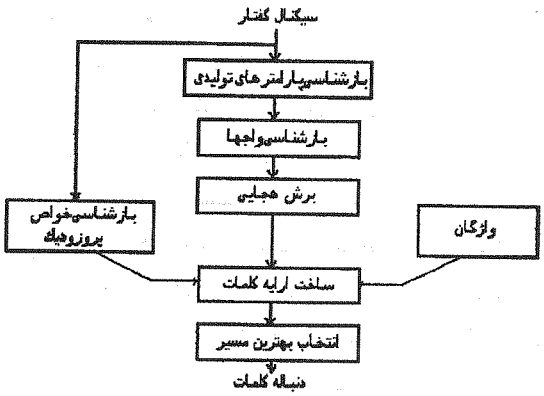


شکل (۲) منحنی مرز زیر و بمی کلمه «کتابی» با تکیه روی هجای «بی»

همین کلمه اگر با تکیه روی هجای «بی» تلفظ شود، به مفهوم «شبیه کتاب» (صفت) خواهد بود (شکل ۲). موارد شبیه به این مثال در گفتار پیوسته زبان فارسی فراوان هستند [۱]. در چنین مواردی، بازشناسی کلمه صحیح منوط به بازشناسی محل هجای تکیه بر (از خصوصیات پروزودیک کلمات) است. این موضوع در سیستمهای درک گفتار که هدف نهایی آنها دریافت مفهوم جملات است، اهمیت بیشتری می‌یابد. در عین حال وارد کردن خواص پروزودیک کلمات در روش بازشناسی آنها می‌تواند درصد صحت بازشناسی را افزایش دهد.

۲ـ روش

شکل ۳ تصویری از مدل بازشناسی گفتار را نشان می‌دهد.



شکل (۳) تصویری از مدل بازشناسی بهترین دنباله کلمات

شرح عملکرد این مدل به صورت زیر است:

الف: سیگنال گفتار به پنجره‌های ۲۳ میلی ثانیه‌ای تقطیع شده و یک شبکه عصبی MLP چهار لایه، به هر پنجره از سیگنال تعدادی مشخصه تولیدی (Articulato-ry features) نسبت می‌دهد [۲]. سپس شبکه عصبی MLP دیگری، با ورودی مشخصات تولیدی نسبت داده شده به هر پنجره سیگنال گفتار، شبیه‌ترین واج زبان فارسی [۲] را برای آن تعیین می‌کند.

دنباله پنجره‌ها به همراه برچسب‌های واجی وارد بلوک تفکیک هجایی می‌شوند. تفکیک هجاها از یکدیگر براساس اطلاعات و قواعد زبانشناسی زبان فارسی صورت می‌گیرد [۸]. الگوهای واجی هجاهای زبان فارسی CVC,CV و CVCC(V معرف «واکه» و C معرف «همخوان») هستند. قبل از اعمال این قواعد به دنباله پنجره‌ها، فیلتر خاصی برای حذف برخی اشتباهات به این دنباله اعمال می‌گردد. برای مثال طول یک واکه نمی‌تواند از حدی کمتر باشد و اگر واکه‌ای تنها در یک پنجره ظاهر شده باشد و اطراف آن را همخوان‌ها گرفته باشند، به عنوان خطای بازشناسی تلقی شده و حذف می‌گردد.

مواردی اینچنین توسط این فیلتر اصلاح می‌گردند. سپس قواعد واجی هجاهای زبان فارسی به آنها اعمال می‌گردند. خروجی این بلوک دنباله هجاهای گفتار است.

در حین فرآیند اولیه ممکن است برخی از واجها حذف شده و بازشناسی نشوند. بنابراین لازم است الگوهای VC,V و VCC نیز برای هجاها مجاز شمرده شوند تا در الگوریتم تفکیک هجایی خللی وارد نشود.

از علل اصلی تقطیع هجایی گفتار در مدل ارائه شده، استفاده از نوع تکیه هجاهای کلمات در مراحل نهایی بازشناخت کلمات است. در این صورت انتخاب واحد هجا در مرحله بازشناخت کلمات، ترکیب عوامل شباهت واجی کلمات و شباهت‌های پروزودیک آنها با دنباله هجایی بازشناسی شده را بسیار ساده‌تر می‌سازد.

ب) واژگان مورد استفاده از ۱۲۰۰ کلمه تشکیل شده است که همگی در دادگان فارس دات [۱۴] موجود هستند. کلمات به کار رفته در این دادگان از بین پربسامدترین کلمات زبان فارسی (حداکثر چهار هجایی) انتخاب شده‌اند. با این انتخاب نه تنها از کلمات پر کاربرد زبان فارسی استفاده کرده‌ایم، بلکه توانسته‌ایم جملات این دادگان را برای آموزش و توسعه مدل خود به کار بریم. اطلاعات مربوط به هر کلمه در واژگان عبارتند از: تعداد هجاها، واجهای هر هجا و نوع تکیه هجاهای هر کلمه.

ج) منحنی نشانگر مرز زیر و بمی هجاهای سیگنال گفتار (pitch contour) با بکارگیری پارامترهای کپستروم تعیین می‌شود [۳]. سپس یک شبکه عصبی از نوع "Recurrent MLP" با توجه به این منحنی، تکیه هجاها را تعیین کرده و محل هجای تکیه بر را

مشخص می‌کند. در واژگان نیز مشابه این برچسب‌های پروزودیک برای هجاهای کلمات وجود دارد و مشخص شده است که هجای تکیه بر هر کلمه کدامست. بازشناسی کلماتی که تفاوت آنها با یکدیگر به خاطر محل هجای تکیه بر است، فقط با استفاده از این برچسب‌های پروزودیک امکان دارد.

د) دنباله هجاها و برچسب‌های پروزودیک وارد بخش دیگری می‌شوند که وظیفه آن پیدا کردن بهترین کلماتی است که می‌توانند در هر محل از دنباله هجایی بازشناسی شده، موجود باشند. برای مثال، جدول ۱ دنباله هجایی جمله «این کتاب را حتماً بخوان» و بهترین کلمات قابل درج در دنباله هجایی آن را نشان می‌دهد:

جدول (۱) دنباله هجائی جمله «این کتاب را حتماً بخوان» و کلمات ممکن در آن:

in/ke/t@b/r@/hat/man/be/x@n

-in-ket@b - - man-be-

-ink- -r@-hat man

-ke-t@b - r@-hat- -bex@n

پیچیدگی مسئله زمانی بیشتر می‌شود که احتمال وقوع خطا در بازشناسی واجها را هم وارد الگوریتم بازشناسی کلمات کنیم. کلماتی هستند که با برخی تغییرات قابل بازسازی از دنباله هجاها هستند. بنابراین لازم است بهترین کلمات موجود در دنباله هجایی انتخاب شده و احتمال وجود آنها در این دنباله محاسبه گردد. در این مدل، آرایه‌ای از کلمات ساخته می‌شود که چهار ستون (برابر حداکثر تعداد هجای کلمات واژگان) دارد و تعداد سطرهای آن برابر تعداد هجاهای دنباله هجایی است.

a_{11}	a_{12}	a_{13}	a_{14}
a	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
$a_{(n-3)}1$	$a_{(n-3)}2$	$a_{(n-3)}3$	$a_{(n-3)}4$
$a_{(n-2)}1$	$a_{(n-2)}2$	$a_{(n-2)}3$	ϕ
$a_{(n-1)}1$	$a_{(n-1)}2$	ϕ	ϕ
a_{n1}	ϕ	ϕ	ϕ

هر عضو این آرایه شماره یک کلمه از واژگان است. ستون اول حاوی شماره کلمات تک هجایی، ستون دوم حاوی شماره کلمات دو هجایی و ... می‌باشد. برای مثال a11 شماره بهترین کلمه تک هجایی است که قابل درج در اولین هجای دنباله است و یا a14 شماره بهترین کلمه چهار هجایی است که قابل درج در هجای اول تا چهارم دنباله هجایی است. اگر در هر محل از دنباله هجایی، کلمه‌ای n هجایی قابل درج نباشد، در معادل آن محل در آرایه کلمات عدد صفر قرار داده می‌شود.

به طور قرینه با این آرایه، آرایه دیگری با همین ابعاد ساخته می‌شود. اعضای آرایه دوم، اعدادی نسبی بین صفر و یک هستند که به طور قرینه با هر عضو از آرایه اول، امتیازی را برای وجود یک کلمه از آرایه اول در دنباله هجایی بازشناسی شده نشان می‌دهند. هر چه این عدد به «یک» نزدیکتر باشد، امتیاز آن عدد بالاتر است و امکان انتخاب آن به عنوان یکی از کلمات بازشناسی شده، افزایش می‌یابد. برای محاسبه این عدد برای هر کلمه، دو معیار اصلی مورد نظر قرار می‌گیرد:

د-۱) اولین معیار شباهت واجی کلمات با واجهای دنباله هجایی بازشناسی شده است. میزان این شباهت به کمک آرایه ابهام (confusion matrix) روش بازشناسی واجها [۲] محاسبه می‌گردد.

با آزمایش‌هایی که بر روی روش بازشناسی پنجره‌های سیگنال گفتار و نسبت دادن آنها به هر واج زبان فارسی انجام شده است، احتمال اشتباه شدن هر واج با واج دیگر محاسبه شده و نتایج در آرایه ابهام این روش درج شده‌اند. باتوجه به اعداد این آرایه، آرایه دیگری به نام آرایه عدم شباهت واجها می‌سازیم. اعضای این آرایه جدید اعدادی هستند که به طور نسبی فاصله‌هایی را برای واجها ارائه می‌کنند.

بدین ترتیب هر چه احتمال اشتباه شدن دو واج بیشتر باشد، عدد معرف فاصله آنها نیز بیشتر خواهد بود. جدول ۲، آرایه عدم شباهت واکه‌ها را برای روش بازشناسی واجها که در این مدل مورد استفاده قرار گرفته است، نشان می‌دهد.

جدول (۲) عدم شباهت واکه ها براساس ماتریس ابهام روش بازشناسی آنها

	آ	أ	ا	اُ	او	ای
آ	۰	۰/۳	۰/۶	۰/۰۷۵	۰/۲۴	۱/۵
أ	۰/۱۵	۰	۰/۱۵	۰/۳۳	۱/۵	۱/۵
ا	۱/۵	۰/۶	۰	۰/۳	۰/۳	۰/۱۵
اُ	۰/۶	۱/۵	۰/۳۳	۰	۰/۰۷۵	۱/۵
او	۱/۵	۱/۵	۰/۳۳	۰/۲۴	۰	۰/۶
ای	۱/۵	۱/۵	۰/۲۴	۱/۵	۰/۶	۰

برای مثال، مطابق این جدول فاصله نسبی واکه «آ» از واکه «آ» برابر ۰/۱۵ و فاصله نسبی واکه «آ» از «أ» برابر ۰/۳ می باشد. علت قرینه نبودن این آرایه این است که معمولاً احتمال اینکه واج «A» اشتباهاً واج «B» بازشناسی شود، با احتمال اینکه واج «B» اشتباهاً واج «A» بازشناسی شود، برابر نیست و در نتیجه فاصله هر کدام از دیگری برابر نخواهد بود.

مشابه آرایه نشان داده شده در جدول ۲ برای همخوان ها نیز تدوین شده و در محاسبات مورد استفاده قرار می گیرد.

همچنین برای حالاتی که یک واج در دنباله واجی بازشناسی شده، حذف (deletion) و یا اضافه (insertion) شده باشد، امتیازهایی به عنوان عدم شباهت رشته واجی بازشناسی شده و کلمات واژگان در نظر گرفته می شوند.

مجموع امتیازهای عدم شباهت واجهای دنباله هجایی و واجهای هر کلمه از واژگان، بعد از تقسیم بر تعداد هجاهای آن کلمه، به عنوان میزان عدم شباهت آن کلمه ذخیره می گردد. کلماتی از واژگان که این عدد برای آنها حداقل باشد، تا این مرحله امتیاز بیشتری را برای قرار گرفتن در دنباله کلمات دارا می باشند.

د ـ ۲) علاوه بر اولین معیار که ملاحظه شباهت واجی است، شباهت پروزودیک کلمات واژگان با دنباله هجایی بازشناسی شده مورد بررسی قرار می گیرد.

در گزارش قبلی [۳] روشی برای بازشناسی نوع تکیه هجاها عنوان شده است. با استفاده از این روش می توان محل هجای تکیه بر را در دنباله هجایی بازشناسی شده، تعیین کرد و با محل هجای تکیه بر کلمات واژگان (نوع تکیه هجاهای هر کلمه در واژگان موجود است) مقایسه نمود.

کلماتی از واژگان که این الگوی پروزودیک آنها با دنباله هجایی بازشناسی شده تفاوت دارد، از فضای جستجو حذف شده و باعث کوچکتر و دقیق تر شدن فضای جستجو می شوند. در عین حال کلماتی که از لحاظ واجی مشابه یکدیگرند، ولی از لحاظ محل هجای تکیه بر متفاوت هستند نیز تفکیک می شوند.

عامل پروزودیک دوم که مورد استفاده قرار گرفته است، بررسی امکان صفر شدن فرکانس پایه گفتار F0 در طول یک جمله است. با مطالعه و آزمایش واجهای زبان فارسی می بینیم که در حین ادای واجهای x, s, sh, ch, p, f و k (ف، پ، ت، چ، ش، س، خ، ک، ء) امکان صفر شدن F0 وجود دارد. از سوی دیگر در طول واج های p, sh, s, t و f (ت، س، ش، پ، ف) حتماً صفر شدنF0 رخ می دهد.

از این عامل برای کوچکتر کردن فضای جستجو استفاده می کنیم. برای مثال اگر در بین هجاهای یک کلمه دو هجایی، صفر شدن F0 را ملاحظه کنیم، در فضای جستجوی کلمات واژگان از کلماتی مانند «اینجا»، «بیرون» و ... صرفنظر نموده و به جستجو در بین کلماتی از قبیل کتاب، انسان و ... می پردازیم.

بازشناسی اینکه F0 در کدام نواحی صفر شده است، بسیار ساده تر از بازشناسی محل هجای تکیه بر (عامل پروزودیک اول) است و همین امر کاربرد عامل دوم را ساده می کند.

با ترکیب دو معیار شباهت واجی و شباهت پروزودیک کلمات واژگان با بخش های مختلف دنباله هجایی بازشناسی شده، بهترین کلمات موجود در این دنباله تعیین گردیده و یک آرایه چهارستونی از این کلمات به همراه آرایه دیگری که حاوی امتیازهای کلمات (عددهایی بین صفر تا یک) است، ساخته می شوند.

ه) بکارگیری الگوریتم ویتربی برای تعیین بهترین دنباله کلمات:

آرایه کلماتی که در قسمت قبلی به دست می آید، شکل مناسبی برای اجرای آلگوریتم ویتربی [۴][۵] ندارد، زیرا این آلگوریتم قابل اعمال به یک فرآیند کاملاً مرحله ای است، در حالی که در این آرایه کلمات اگر دریک سطر آرایه بخواهیم مسیری را دنبال کنیم که از یک کلمه ۳ هجایی عبور می کند، باید دو سطر آرایه را جا گذاشته و به سطر سوم بعد از سطر فعلی مراجعه نماییم.

پس نمی توان الگوریتم ویتربی را به طور مستقیم به این آرایه کلمات اعمال نمود. بنابراین آرایه ۴ ستونی

کلمات را تبدیل به آرایه دیگری می‌کنیم که ۱۱ ستون دارد و شرایط اعمال الگوریتم ویتربی به آن فراهم است.

روش تولید آرایه جدید (از روی آرایه قبلی کلمات) این گونه است که اگر مسیری از دنباله کلمات بخواهد از یک کلمه n هجایی عبور کند، غیر از امتیاز قرار داده شده برای هجای اول آن (که همان امتیاز قرار داده شده برای آن کلمه n هجایی است)، امتیاز (n-1) هجای باقیمانده برابر ۱ قرار داده می‌شود. در نتیجه ستونهایی با عناوین «هجای دوم کلمات دو هجایی»، «هجای دوم کلمات سه هجایی» و ... به ستونهای آرایه اول اضافه می‌شوند و در مجموعه یک آرایه ۱۱ ستونی ساخته می‌شود.

سپس با استفاده از الگوریتم ویتربی بهترین مسیر (با بیشترین امتیاز) را برای این آرایه جدید به دست می‌آوریم. متناظر با این مسیر، بهترین دنباله کلمات تعیین می‌گردد.

۳ـ نتایج عملی

با استفاده از بیش از ۱۰۰ جمله آزمایشی مشاهده شد که با ۷۶٪ صحت بازشناسی واجها، درصد صحت بالای ۸۱٪ برای بازشناسی دنباله کلمات به دست می‌آید. استفاده از پروزودی، به طور متوسط باعث ۵٪ رشد در درصد صحت فرآیند بازشناسی کلمات شده است که قابل توجه می باشد. در برخی موارد دیده می‌شود که دنباله صحیح تر کلمات در مسیری از آرایه کلمات است که امتیاز آن کمی کوچکتر از مسیر اول است.

بنابراین اگر مسیرهایی را که امکان آنها نزدیک به مسیر اول است محاسبه نموده و مورد توجه قرار دهیم، امکان وجود پاسخ بهتر در این مجموعه بیشتر است. با استفاده از اطلاعات نحوی و معنایی و الگوریتم‌های مناسب دیگر می‌توان از بین دنباله های برگزیده کلمات، بهترین دنباله را انتخاب نمود. به این ترتیب درصد صحت بازشناسی کلمات به مقدار فاحشی افزایش می‌یابد.

اضافه کردن بخش‌های تحلیل نحوی و معنایی به این مدل را می‌توان به عنوان روشی کارساز در افزایش توان بازشناسی آن پیگیری نمود.

٤ـ نتیجه گیری

در این گزارش مدلی را برای بازشناسی دنباله کلمات گفتار پیوسته ارائه نمودیم. این مدل عمل بازشناسی کلمات را به روشی ریاضی که جزئیات آن ارائه شد، انجام می‌دهد. در عین حال این مدل از برخی از خواص پروزودیک سیگنال گفتار که معمولاً در چنین فرآیندهایی کمتر به آنها توجه می شود، برای بهبود بازشناسی و همچنین رفع ابهام در بازشناسی کلماتی که دنباله واجی مشابه دارند استفاده می‌کند. آزمایش های انجام شده کارآیی مدل را نشان می‌دهند. موفقیت این مدل در استفاده مؤثر از خصوصیات پروزودیک سیگنال گفتار پیوسته از جمله نکات مهمی است که در این گزارش آمده است.

مراجع

[۱] وحیدیان کامیار، ۱۳۵۷، «نوای گفتار در زبان فارسی»، اهواز، انتشارات دانشگاه جندی شاپور، آذر ماه ۱۳۵۷.

[۲] سید علی سید صالحی، ۱۳۷۴، بازشناخت گفتار پیوسته فارسی با استفاده از مدل عملکردی مغز انسان در درک گفتار، پایان‌نامه دکترای مهندسی برق، دانشگاه تربیت مدرس، بهمن ۱۳۷۴.

[۳] فرشاد الماس گنج، سید محمدرضا هاشمی گلپایگانی، حسن زند، «تشخیص نوع مرز زیر و بمی هجاهای گفتار پیوسته فارسی با استفاده از شبکه عصبی»، کنفرانس بین المللی سیستمهای هوشمند و شناختی ـ سمپوزیوم علوم شناختی، مرکز تحقیقات فیزیک نظری و ریاضیات، ۱۳۵۷.

[4] H. F. Silverman, "The Application of dynamic programming to connected speech Recognition- IEEE ASSP Magazine - July 1990.

[5] D.P. Morgan, "Neural network and speech processing", Kluwer Academic Publishers, (1991).

[6] D. Crystal, "Encyclopedia of language and linguistic", PP. 169-173, (1992).

[7] H. Zand, "Stress in Persian", Manuscript, Linguistics Department, the University of Kansas, Lawrence, USA, (1989).

[8] Y. Samareh, "The arrangement of segmental phonemes in Farsi", University of Tehran, Faculty of letters, Tehran, (1977).

[9] P. Laface, C. Vair, L. Fissore, "A fast segmental viterbi algorithm for large vocabulary recognition", ICASSP95, PP. 560-563, (1995).

[10] H. Ney, X. Aubert, "A word graph algorithm for large vocabulary continuous speech recognition", ICSLP 94, PP. 1355 - 1358, (1994).

[11] P. S. Gopalakrishnan, L. R. Bahl, R.L. Mercer, "A treee search strategy for large-vocabulary continuous speech recognition", ICASSP 95, PP. 572-575, (1995).

[12] J. I. Murakami, "New wordspotting algorithm based on forward decoding", Eurospeech 95, PP. 2153-2154, (1995).

[13] H. Kanagawa, "A hybrid wordspotting method for spontaneous speech understanding using word based pattern matching and phoneme-based HMM", ICASSP 95, pp. 289-292.

[14] M. Bijankhan, J. Sheikhzadegan, "FARSDAT - The Speech Data Base of Farsi Spoken Language", proceedings of the fifth international Australian conference on speech science and technology (SST-94), vol2, PP. 826-831, (1994).