



An Introduction to the Process of the Design and Production of a Standard Persian Emotional Speech Database

Mahsa Ravanbakhsh¹ 

Postdoctoral Researcher, Institute of Cognitive Studies (ICSS), Tehran, Iran

Saeed Setayeshi² 

Full Professor, Department of Energy Engineering and Physics, Amirkabir University of Technology, Tehran, Iran

Abstract

Different environmental stimuli are able to influence and change human emotional states. Speech expresses change in emotional state in two **ways**: verbally through vocabulary and syntax and also non-verbally through tone and intonation. Tone and intonation have a paralinguistic role and can alter the meaning of speech. The concept of affective computing in computer science first introduced the quantitative processing and study of emotions. The main idea was that the machine could recognize and interpret human emotional states and respond or behave appropriately according to them. The quantitative study of emotional speech is known as speech emotion recognition. Recognition or classification of emotional speech means being able to recognize the speaker's emotional state using analysis performed on the speech signal. The first step for conducting this type of study is to have a rich, standard, high-quality, and appropriately sized dataset for evaluating speech emotion recognition algorithms. There are a wide variety of standard speech emotion datasets in popular languages. The absence of such a database in the Persian language has led researchers in the field of speech emotion recognition to have a low-quality understanding of emotional

1. mahsa.ravanbakhsh@gmail.com (Corresponding Author)

2. setayesh@aut.ac.ir

How to cite: Ravanbakhsh, M., & Setayeshi, S. (2024). An Introduction to the Process of the Design and Production of a Standard Persian Emotional Speech Database. *Language and Linguistics*, 20(40), 225- 248. doi: [10.30465/LSI.2025.50640.1787](https://doi.org/10.30465/LSI.2025.50640.1787)

patterns and their impact. This point underscores the necessity of creating and producing a speech emotion dataset in the Persian language. In this paper, we describe the designing, preparing, and producing of a speech emotion dataset in standard Persian, similar to the approach of the Berlin Emotional Speech Database, which can be used in studies of speech emotion recognition.

Keywords: emotional speech, emotional speech database, emotional state, paralinguistic, speech emotion recognition.

1. Introduction

Various environmental stimuli can influence and change human emotional states. These changes in emotional states appear in the human body as changes to bodily postures, facial expressions, heart rate, sweating, breathing, and speech. Speech conveys emotional state changes in two ways: verbally through words and syntax, and non-verbally through tone and intonation. Tone and intonation of speech have a paralinguistic function and can alter the meaning of speech (Alinezhad and Hoseini-Balam, 2013 (1392 SH); Alinezhad, 2010).

The quantitative study of emotional speech is known as Speech Emotion Recognition (SER). SER, or classification, refers to the ability to identify the speaker's emotional state by analyzing the speech signal. In these studies, the focus is on processing, analyzing, and classifying the audio signals of emotional speech by artificial intelligence and machine learning tools. The first step in conducting such studies is to access a standard and high-quality emotional speech dataset, which is the primary and crucial step in speech emotion studies and emotional speech recognition.

In the present paper, we intend to present the process of designing and producing an emotional speech dataset in standard Persian. In this research, we plan to explain how we created an emotional speech dataset in standard Persian that matches the quality and features of the Berlin Emotional Speech Database (Burkhardt et al., 2005), so it can be used in studies on emotional speech recognition and help researchers accurately measure what they want to evaluate.

2. Literature Review

Cognitive science, as an interdisciplinary field, aims to scientifically study the mind. The mind in cognitive science is considered as a set of functions, operations, and brain activities in such a way that it can be described,

explained, formulated, and modeled scientifically (Friedenberg & Silverman, 2006; Ghasemzadeh, 2006 (1385 SH)). Emotion is one of the fundamental topics in this field, whose representation and computation in the mind and brain **are** of particular importance.

An acoustic signal transmits speech, the most accurate and effective form of communication (Deller et al., 1999). Speech, beyond the words and syntax that form it, encompasses a vast amount of information, such as the speaker's age, gender, location, and emotional state. Two main computational components that form the SER studies are acoustic features that are representing the emotional states in speech and the classification methods for those features (Hasrul et al., 2011). However, the significant step for conducting such studies is having access to standardized and appropriate data for examination and processing.

Various institutions and laboratories around the world have developed and produced different types of emotional speech datasets based on their research criteria. Some of these emotional speech datasets include the Berlin Emotional Speech Database (Burkhardt et al., 2005), the Danish Emotional Speech Dataset (Engberg and Hansen, 1996), and the Emotional Speech of Mandarin and Burmese Speakers (Nwe et al., 2003). Recently, several emotional speech datasets have been created in Persian, including the Persian Drama Radio Emotional Corpus (PDREC), the Persian Emotional Speech Database, the Sahand Emotional Speech (SES) Database, the Persian Emotional Speech Database (Persian ESD), and the Sharif Emotional Speech Database (ShEMO).

Unfortunately, Persian language datasets have been rarely used due to their lack of quality and standards. However, researchers frequently use the Berlin Emotional Speech Database, one of the best types of emotional speech datasets. This database consists of a total of 800 emotional speech samples. The target emotions in the emotional speech database include anger, fear, happiness, sadness, boredom, disgust, and a neutral version.

3. Method

The stages of designing, producing, and collecting the Persian emotional speech dataset are similar to those used in the production of the Berlin emotional speech database. The language of this emotional speech dataset is standard Persian. From a technical standpoint, the process of producing the emotional speech dataset includes the following stages:

1. Designing and preparing utterances without emotional content.

2. Selecting a recording laboratory or studio similar to the standards of the Technical University of Berlin.
3. Calling for participation in the project.
4. Selecting participants.
5. Preparing participants.
6. Recording acoustic samples.
7. Labeling the produced acoustic samples.
8. Evaluating the produced samples.

4. Results

During the process of producing samples of emotional speech in standard Persian, we had a total of 22 participants. These participants were recorded in different emotional states of anger, fear, happiness, sadness, boredom, disgust, and a neutral version. We recorded 17 emotionally neutral utterances from each participant, with an average of three recordings. The total recordings provided 7854 initial raw samples.

5. Conclusion

Speech signal carries the message with three main components, i.e., syntax, content, and the coded emotion in the form of intonations. The speech signal intertwines these three components, which influence each other to form the emotional and contextual content of the speech.

To study the emotional content in the Persian language, we have created the procedures to produce a dataset that can be used in future research studies. We then used these procedures to collect samples from 22 participants to form the Persian speech emotional dataset. This paper presents a report on the various stages of designing, preparing, and producing an emotional speech database in standard Persian.

Acknowledgments

We gratefully acknowledge Dr. Shahin Nematzadeh for her invaluable guidance and support. We also thank the utterance evaluators, the participants who provided emotional speech samples, and the Institute of Cognitive Science Studies and Cognitive Sciences and Technologies Council for their support.



مقدمه‌ای بر فرآیند طراحی و تولید پایگاه داده گفتار هیجانی فارسی معیار

مهسا روانبخش  پژوهشگر پسادکتری، مؤسسه آموزش عالی علوم شناختی، تهران، ایران

سعید ستایشی  استاد تمام، دانشکده مهندسی انرژی و فیزیک، دانشگاه امیرکبیر، تهران، ایران

چکیده

محرک‌های محیطی مختلف قادرند تا حالت هیجانی انسان را تحت تأثیر قرار داده و آنرا تغییر دهند. گفتار به دو طریق کلامی بواسطه واژگان و نحو و همچنین غیرکلامی بواسطه لحن و آهنگ گفتار، تغییرات حالت هیجانی را بیان می‌کند. لحن و آهنگ گفتار دارای نقش پیرازبانی هستند و معنای گفتار را دستخوش تغییر می‌کنند. پردازش و مطالعه کمی هیجان نخستین بار با مفهوم محاسبات عاطفی در علوم کامپیوتر مطرح شد. ایده اصلی این بود که ماشین بتواند حالت هیجانی انسان را بازشناسی و تفسیر کند و مطابق با آن پاسخ یا رفتار مناسب ارائه کند. مطالعه کمی گفتار هیجانی با عنوان بازشناسی گفتار هیجانی شناخته می‌شود. بازشناسی یا دسته‌بندی گفتار هیجانی بدین معناست که بتوان حالت هیجانی گوینده را با استفاده از تحلیل انجام شده بر روی سیگنال گفتار مشخص کرد. نخستین گام برای انجام این گونه از مطالعات داشتن مجموعه دادگان غنی، استاندارد، با کیفیت و البته به تعداد مناسب برای ارزیابی الگوریتم‌های بازشناسی هیجان گفتار است. انواع گسترده و استاندارد مجموعه دادگان گفتار هیجانی به زبان‌های مختلف وجود دارند. عدم دسترسی پژوهشگران این حوزه به چنین پایگاه داده‌ای سبب شده است تا دانش ما پیرامون الگوهای هیجانی و تأثیر آن در زبان فارسی به وضوح آشکار نباشد. این نکته ضرورت وجود تهیه و تولید مجموعه دادگان گفتار هیجانی به زبان فارسی را نشان می‌دهد. در این مقاله قصد داریم فرآیند طراحی، تهیه و تولید مجموعه دادگان گفتار هیجانی فارسی معیار مشابه با مجموعه دادگان گفتار هیجانی برلین که قابلیت استفاده در مطالعات بازشناسی گفتار هیجانی را داراست تشریح کنیم.

کلیدواژه: بازشناسی گفتار هیجانی، پایگاه داده گفتار هیجانی، حالت هیجانی، عوامل پیرازبانی، گفتار هیجانی.

مقدمه

محرك‌های محیطی مختلف قادرند تا حالت هیجانی انسان را تحت تأثیر قرار دهند و آنرا تغییر دهند. تغییرات حالت هیجانی در رُست‌های بدنی، میمیک صورت، ضربان قلب، میزان تعرق، تنفس و گفتار نمود پیدا می‌کنند. گفتار به دو طریق کلامی بواسطه واژگان و نحو و همچنین غیرکلامی بواسطه لحن و آهنگ گفتار، تغییرات حالت هیجانی را بیان می‌کند. لحن و آهنگ گفتار دارای نقش پیرایه‌بانی^۱ هستند و قادرند تا معنای گفتار را دستخوش تغییر کند (علی‌نژاد و حسینی‌بالام، ۱۳۹۲؛ علی‌نژاد، ۲۰۱۰).

مطالعه کمی گفتار هیجانی با عنوان بازشناسی گفتار هیجانی^۲ شناخته می‌شود. بازشناسی یا دسته‌بندی گفتار هیجانی بدین معناست که بتوان حالت هیجانی گوینده را با استفاده از تحلیل انجام شده بر روی سیگنال گفتار مشخص کرد. مطالعات فراوانی پیرامون بازشناسی گفتار هیجانی انجام شده است که نشان می‌دهند بیان حالات هیجانی از ماهیت پیچیده‌ای برخوردارند که مطالعه کمی هیجان را دشوار می‌سازد. پژوهش‌های انجام شده پیرامون بازشناسی جنسیت گوینده از دو جنبه قابل بررسی هستند، نخست ویژگی‌های مورد استفاده برای بازشناسی و دوم روش‌ها و مدل‌های مورد استفاده برای بازشناسی.

هیجان یکی از مؤلفه‌های اساسی در علوم شناختی است که بسیار اندک بدان پرداخته شده است و پژوهش‌های اندکی پیرامون آن در علوم شناختی زبان انجام شده است. اما نتایج بدست آمده از این پژوهش‌ها نشان‌دهنده تأثیر بسزای بیان هیجان در معناشناسی و تغییراتی است که در معنی واژگان و پاره‌گفتارها ایجاد می‌کند. پژوهش‌های انجام شده پیرامون محتوای هیجانی گفتار با دو رویکرد کلی کیفی و کمی قابل تفکیک هستند. در پژوهش‌هایی مانند علی‌نژاد (۱۳۸۴؛ ۲۰۱۰) و علی‌نژاد و ویسی (۱۳۸۶) که دارای رویکردی کیفی هستند، نشان می‌دهند که هیجان با تغییر آهنگ و لحن گفتار بر درک معنا، یادسپاری، اندیشه، رفتار و پیام منتقل شده اثر می‌گذارد. اما در پژوهش‌های انجام‌شده با رویکرد کمی پیرامون محتوای هیجانی گفتار، تمرکز بر پردازش، تجزیه و تحلیل سیگنال صوتی گفتار هیجانی و تحلیل و دسته‌بندی آنها با کمک ابزارهای هوش مصنوعی و یادگیری ماشین است (هاسرل^۳ و همکاران، ۲۰۱۱). هدف این دسته از پژوهش‌ها نشان دادن ویژگی‌هایی در سیگنال صوتی است که بارزتر از سایر مؤلفه‌های تشکیل‌دهنده سیگنال گفتار، محتوای هیجانی را بازنمایی می‌کنند و سپس به کمک آنها حالات هیجانی را بازشناسی کنند.

1. paralinguistic

2. Speech Emotion Recognition (SER)

3. M. Hasrul

پژوهش‌های انجام شده در حوزه روان‌شناسی نشان می‌دهند که گفتار بهتر از سایر روش‌های ابراز و بیان هیجان محتوای هیجانی را منتقل می‌کند (اکمن^۱، ۲۰۰۷).

هیجان دارای تأثیر بسزائی بر قابلیت‌های شناختی انسان است. علوم شناختی زبان که به طور خاص به مطالعه زبان و درک چیرستی آن می‌پردازد نیز متأثر از هیجان است. هیجان غیرکلامی قادر است تا محتوای پیام یک پاره‌گفتار بیان شده تحت یک قالب نحوی و واژگانی یکسان را تغییر دهد. این موضوع نشان‌دهنده جایگاه ویژه مطالعات کمی و پارامتریک محتوای هیجان گفتار در علوم شناختی زبان است، به خصوص گونه‌ای از مطالعات که برآمده از ویژگی‌های زیستی درک گفتار و پردازش هیجان در سیستم عصبی و شناختی باشد بسیار حائز اهمیت است.

گفتار بهترین و مؤثرترین روش برای برقراری ارتباط و انتقال اطلاعات است. هر گفتاری که بیان می‌شود علاوه بر نحو و واژگان شامل اطلاعات بسیاری مانند سن، جنسیت و حالت هیجانی گوینده گفتار نیز است. این اطلاعات در قالب ویژگی‌هایی در سیگنال صوتی گفتار حمل می‌شوند و می‌توانند دارای ماهیت ایستا و یا پویا باشند. پژوهش‌های انجام شده با رویکردهای مختلف پیرامون بازشناسی هیجان گفتار اشاره به چند وجهی بودن^۲ و پیچیدگی^۳ در این مسئله دارند (هاسرل و همکاران، ۲۰۱۱). این نکته سبب شده تا مطالعه پیرامون هیجان و بازشناسی هیجان گفتار با دشواری‌های فراوانی همراه شود.

با توجه به آنکه مطالعه و پژوهش پیرامون گفتار هیجانی یکی از موضوعات بسیار کاربردی و حائز اهمیت در علوم شناختی زبان، زبان‌شناسی رایانشی و هوش مصنوعی است و دارای کاربردهای فراوانی در حوزه تعامل انسان و ماشین است، نخستین گام برای انجام این گونه از مطالعات دسترسی به مجموعه‌داده‌گان گفتار هیجانی استاندارد و با کیفیت نخستین و مهم‌ترین گام در مطالعات هیجان گفتار و بازشناسی گفتار هیجانی محسوب می‌شود. در مؤسسات و آزمایشگاه‌های گوناگون در تمامی جهان بر حسب اهداف و کاربردهای مد نظرشان انواع گوناگون مجموعه‌داده‌گان گفتار هیجانی را طراحی، تهیه و تولید می‌کنند. چرا که بازشناسی هیجان گفتار و بررسی ویژگی‌های پارامتریک آن در عرصه‌های گوناگون همچون مطالعات علوم شناختی- زبان، معناشناسی شناختی، پژوهش‌های پیکره‌ای زبان، امور درمانی، برنامه‌های آموزشی، رفتارشناسی و ارتباط مؤثر انسان با ماشین از جایگاه ویژه‌ای برخوردار است. اکثر مجموعه‌داده‌گان تولید شده به زبان‌های انگلیسی، آلمانی و چینی هستند. این مجموعه‌داده‌گان

1. P. Ekman
2. multimodal
3. complexity

مکرراً توسط پژوهشگران مورد استفاده قرار می‌گیرند و نتایج بدست آمده از آنها سبب افزایش دانش پیرامون هیجان و الگوهای هیجانی در این زبان‌ها شده است. همچنین این دانش به آنها کمک کرده که در انواع کاربردهای آموزشی، درمانی و هوش مصنوعی پیشرفت چشمگیری داشته باشند. شایان ذکر است که چندین نمونه مجموعه‌دادگان گفتار هیجانی به زبان فارسی پیش‌تر تهیه و ارائه شده‌اند که از آن جمله می‌توان به پایگاه‌داده گفتار احساسی زبان فارسی (کشتیاری و همکاران، ۲۰۱۵) و دادگان گفتار هیجانی شریف (محمدنظامی و همکاران، ۲۰۱۹) اشاره کرد.

متأسفانه مجموعه‌دادگان موجود به زبان فارسی از استانداردها، کیفیت و ویژگی‌های در نظر گرفته شده در انواع غیر فارسی زبان برخوردار نیستند. عدم دسترسی پژوهشگران حوزه بازشناسی گفتار هیجانی به چنین پایگاه‌داده‌ای سبب شده است تا دانش ما پیرامون الگوهای هیجانی و تأثیر آن در زبان فارسی به عنوان یکی از غنی‌ترین زبان‌های جهان به وضوح روشن و شفاف نباشد. این نکته ضرورت وجود تهیه و تولید مجموعه‌دادگان گفتار هیجانی به زبان فارسی را دو چندان می‌کند. چنین مجموعه‌دادگانی امکان مطالعه علمی زبان فارسی را فراهم می‌آورد و راهنمایی برای استفاده از آن در انواع کاربردهای آموزشی، درمانی و هوش مصنوعی فراهم می‌کند.

در این گزارش قصد داریم فرآیند طراحی و تولید مجموعه‌دادگان گفتار هیجانی به زبان فارسی معیار ارائه می‌کنیم. در این طرح قصد داریم مراحل و چگونگی فرآیند تولید مجموعه‌دادگان گفتار هیجانی مشابه با استانداردها، ویژگی‌ها و کیفیت پایگاه‌داده گفتار هیجانی برلین^۱ (بورکهارت^۲ و همکاران، ۲۰۰۵) به زبان فارسی معیار گام به گام شرح دهیم؛ به گونه‌ای که قابلیت استفاده در مطالعات بازشناسی گفتار هیجانی را داشته باشد و قادر باشد ارزیابی درستی از آنچه مورد نظر پژوهشگران است را فراهم آورد.

۲. پیشینه پژوهش

علم‌شناخت یا شناخت‌پژوهی، علمی میان رشته‌ای است که هدف آن بررسی علمی ذهن است. ذهن در علم‌شناخت همچون مجموعه‌ای از کارکردها، عملکردها و فعالیت‌های مغزی در نظر گرفته می‌شود به گونه‌ای که بتوان ذهن را به صورت علمی توصیف، تبیین، فرمول‌بندی و مدل‌سازی کرد (فردنبرگ^۳ و سیلورمن^۴، ۲۰۰۶؛ قاسم‌زاده، ۱۳۸۵). پژوهشگران این عرصه

1. Berlin Database of Emotional Speech
2. F. Burkhardt
3. J. Friedenberg
4. G. Silverman

پیوسته در تلاش هستند تا فرآیندها و بازنمایی زیرساخت‌های تمامی عملکردهای هوشمندانه را درک کنند. هیجان یکی از موضوعات بنیادین پیرامون ذهن است که شناخت پژوهان مطالعات و پژوهش‌های فراوان پیرامون آن انجام داده‌اند.

بررسی و مطالعه هیجان به طور سنتی در علوم شناختی بسیار اندک مورد توجه قرار گرفته است، علی‌رغم آنکه از زمان‌های گذشته فلاسفه به اثر هیجان بر تفکر اشاره داشته‌اند (تاگارد^۱، ۲۰۰۵). اما در دهه‌های اخیر شاهد افزایش روز افزون توجه، مطالعه و بررسی هیجان و به طور ویژه تأثیر آن بر شناخت و تصمیم‌گیری هستیم. در علوم شناختی چگونگی بازنمایی^۲ هیجان و رایانش^۳ هیجان در ذهن و مغز از اهمیت ویژه‌ای برخوردار است.

ارائه یک تعریف جامع برای هیجان بسیار دشوار است. در یک تعریف ارائه شده، هیجان حاصل تعامل میان عوامل ذهنی، محیطی و فرآیندهای عصبی و هورمونی است (کلینجینا^۴ و کلینجینا^۵، ۱۹۸۱). در این تعریف هیجان از تجارب هیجانی ناشی می‌شود و تغییرات فیزیولوژیک را به دنبال دارد. همچنین هیجان‌ها سبب می‌گردند تا انسان تبیین‌های شناختی خلق کند و همین‌طور باعث فراخوانی رفتارهایی می‌شوند که در اکثر موارد بیانی (مثلاً خندیدن یا گریستن)، هدفمند و انطباقی هستند. بر اساس این تعریف، هیجان‌ها از فرآیندهای زیست‌شناختی، شناختی و یادگرفته شده ناشی می‌شوند. به عبارت دیگر، هیجان را می‌توان به عنوان مجموعه‌ای از احساسات، شناخت و پاسخ‌های جسمانی که بازتابی از کنش نظام‌های ارزشی در درون مغز است، تعریف کرد (ادلن^۶، ۲۰۰۴).

گفتار به عنوان دقیق‌ترین و مؤثرترین روش ارتباطی محسوب می‌شود که در قالب یک سیگنال صوتی^۷ منتقل می‌شود (دلر^۸ و همکارن، ۱۹۹۹). از طریق گفتار انسان‌ها قادرند تا ایده‌ها، آرا، افکار، احساسات، هیجانات و عواطف خود را به واسطه کلمه‌ها، صداها و ترکیب آنها بیان کرده و با دیگران به اشتراک بگذارند. هر گفتاری که بیان می‌شود هر اندازه هم که ساده باشد، علاوه بر معنای ظاهری و آشکار از مضمونی درونی نیز برخوردار است (لوریا^۹، ۱۹۸۲). گفتار را می‌توان یک بسته در نظر گرفت که حاوی حجم عظیمی از اطلاعات که فراتر از واژگان

-
1. P. Thagard
 2. representation
 3. computation
 4. P. R. Kleinginna
 5. A. M. Kleinginna
 6. G. M. Edelman
 7. acoustic
 8. J. R. Deller
 9. A. R. Luria

و نحو گفتار است، از جمله این اطلاعات می‌توان به سن، جنسیت، مکان و حالت هیجانی گوینده گفتار اشاره کرد.

محرك‌های حسی مختلف حالت ذهنی انسان را تحت تأثیر قرار می‌دهند و آنرا دستخوش تغییر قرار می‌دهند. این تغییرات گفتار به صورت کلامی از طریق واژگان و نحو و همچنین به صورت غیرکلامی از طریق آهنگ و لحن گفتار نمود پیدا می‌کنند و معنای گفتار را تحت تأثیر قرار می‌دهند. آهنگ گفتار دارای نقش پیرایه‌بانی^۱ است و حالات هیجانی گوینده گفتار مانند غم، خشم و شادی را نمایش می‌دهد (علی‌نژاد و حسینی‌بالام، ۱۳۹۲). علی‌نژاد (۲۰۱۰) در یک مطالعه آماری نشان داد بسته به نوع آهنگ بیان پاره‌گفتار «بله»، می‌توان معنای آنرا در یک طیف پیوسته از موافقت کامل تا مخالفت کامل قرار دارد.

بازشناسی هیجان در گفتار با عنوان بازشناسی هیجان گفتاری^۲ خوانده می‌شود. مطالعه کمی هیجان نخستین بار در علوم کامپیوتر با مفهوم محاسبات عاطفی^۳ مطرح شد (پیکارد^۴، ۱۹۹۵). ایده اصلی در این مفهوم این است که ماشین بتواند حالات هیجانی انسان را بازشناسی و تفسیر کند و در نهایت، مطابق با آن حالت هیجانی پاسخ یا رفتار مناسب ارائه کند (پیکارد، ۱۹۹۵؛ فرانتی^۵ و همکاران، ۲۰۱۷). مطالعات فراوانی پیرامون بازشناسی حالات هیجانی در گفتار هیجانی انجام شده است. نتایج نشان می‌دهد بیان حالات هیجانی از ماهیتی پیچیده‌ای برخوردار است که بازشناسی حالت هیجانی و مطالعه کمی هیجان را با دشواری همراه می‌کند. بیشتر پژوهش‌های انجام شده پیرامون بیان حالت هیجانی در گفتار هیجانی به بازشناسی هیجان گفتار ختم می‌شود. این پژوهش‌ها با دیدگاه محاسباتی با دو رویکرد اصلی انجام می‌شوند. در رویکرد نخست تمرکز بر ویژگی‌های صوتی بازنمایی‌کننده حالات هیجانی در گفتار است و دومین رویکرد متمرکز بر ارائه طبقه‌بندی‌کننده‌ها^۶ برای بازشناسی انواع حالات هیجان گفتار می‌شود (هاسرل و همکاران، ۲۰۱۱). اما مهم‌ترین گام برای انجام این قبیل از مطالعات در اختیار داشتن داده‌های استاندارد و مناسب برای بررسی و پردازش است.

مؤسسات و آزمایشگاه‌های گوناگونی در دنیا بر حسب اهداف پژوهشی و کاربردی خود انواع مختلفی از مجموعه‌دادگان گفتار هیجانی را تهیه و تولید کرده‌اند. در جدول ۱ برخی از این مجموعه‌دادگان گفتار هیجانی همچون پایگاه داده گفتار هیجانی برلین (بورکهارت و همکاران،

1. paralinguistic
2. Speech Emotion Recognition (SER)
3. affective computing
4. R. W. Picard
5. E. Franti
6. classifier

۲۰۰۵)، مجموعه‌دادگان گفتار هیجانی دانمارکی^۱ (انبرگ^۲ و هسن^۳، ۱۹۹۶)، مجموعه‌دادگان گفتار هیجانی سخنوران چینی و برمه‌ای^۴ (نیو^۵ و همکاران، ۲۰۰۳)، مجموعه‌دادگان گوش کودک^۶ (سلنی^۷ و مک رابرتز^۸، ۱۹۹۸)، و مجموعه‌دادگان طبیعی مرکز تماس (موریسون^۹ و همکاران، ۲۰۰۷) معرفی شده‌اند.

جدول ۱. مشخصات برخی از مجموعه‌دادگان گفتار هیجانی

نام	زبان	تعداد نمونه‌ها	گویندگان گفتار هیجانی	حالات هیجانی بیان شده
پایگاه‌داده گفتار هیجانی برلین (بورکهارت و همکاران، ۲۰۰۵)	آلمانی	۵۳۵	هنرپیشگان غیرحرفه‌ای	خشم، خنثی، شادی، غم، کسلی، انزجار و ترس
مجموعه‌دادگان گفتار هیجانی دانمارکی (انبرگ و هسن، ۱۹۹۶)	دانمارکی	۲۶۰	هنرپیشگان غیرحرفه‌ای	خشم، خنثی، شادی، غم و تعجب
مجموعه‌دادگان گفتار هیجانی سخنوران چینی و برمه‌ای (نیو و همکاران، ۲۰۰۳)	چینی و برمه‌ای (دو زبانه)	۷۲۰	هنرپیشگان غیرحرفه‌ای	خشم، انزجار، ترس، لذت، غم و تعجب
مجموعه‌دادگان گوش کودک (سلنی و مک رابرتز، ۱۹۹۸)	انگلیسی	۵۰۹	گفتار طبیعی مادرها و پدرها	تأیید، توجه و نهی
مجموعه‌دادگان طبیعی مرکز تماس (موریسون و همکاران، ۲۰۰۷)	چینی	۳۸۸	گفتار طبیعی مراجعان مرکز تماس	خشم و خنثی

1. Danish emotional speech database
2. I. Enberg
3. A. Hanssen
4. Emotional Speech of Mandarin and Burmese Speakers
5. T. L. Nwe
6. Baby Ears
7. M. Slaney
8. G. McRoberts
9. D. Morrison

اکثر مجموعه‌دادگان گفتار هیجانی استاندارد به زبان‌های انگلیسی، آلمانی و چینی تهیه و تولید می‌شوند. پایگاه‌داده گفتار هیجانی برلین با توجه به استانداردهای در نظر گرفته شده در تمامی مراحل تولید آن، یکی از بهترین مجموعه‌دادگان گفتار هیجانی است که از سال ۲۰۰۵ تا کنون در اکثر پژوهش‌های بازشناسی گفتار هیجانی مورد استفاده قرار گرفته است.

شایان ذکر است که در دو دهه گذشته شاهد ارائه مجموعه‌دادگان گفتار هیجانی به زبان فارسی مانند پیکره هیجانی درام رادیویی فارسی^۱ (اسماعیلیان و مروی، ۲۰۱۴)، پایگاه‌داده گفتار هیجانی فارسی منصوری‌زاده و حمیدی (حمیدی و منصوری‌زاده، ۲۰۱۲)، پایگاه‌داده گفتار هیجانی سهند^۲ (صدیقی، ۲۰۰۸)، پایگاه‌داده گفتار هیجانی فارسی^۳ (کشتیاری و همکاران، ۲۰۱۵) و پایگاه‌داده گفتار هیجانی شریف^۴ (محمدنظامی و همکاران، ۲۰۱۹) بوده‌ایم. در جدول ۲ این مجموعه‌دادگان با جزئیات معرفی شده‌اند. متأسفانه این مجموعه‌دادگان به این دلیل که فاقد کیفیت و استاندارد لازم هستند حتی توسط پژوهشگران فارسی‌زبان بسیار اندک مورد استفاده قرار می‌گیرند. منظور از کیفیت و استاندارد به تعداد شرکت‌کنندگان، پاره‌گفتارهای بیان شده، تعداد حالات هیجانی و بیشتر بسنده کردن به استفاده از دیالوگ‌های بیان شده توسط هنرپیشگان و یا افراد که دوره‌های بازیگری و فن‌بیان را گذارنده‌اند اشاره دارد. این موضوع سبب شده است تا دانش‌اندکی پیرامون بازشناسی حالت هیجانی در گفتار سخنوران فارسی زبان داشته باشیم. در ادامه این فقدان دانش باعث می‌شود که در مطالعات زبان‌شناختی، کاربردهای هوش مصنوعی، تعامل انسان و ماشین و سامانه‌های خدمات‌رسانی خودکار فارسی‌زبانان با محدودیت و مشکلات جدی مواجه شوند. این محدودیت و مشکلات نیاز مبرم و جدی به وجود مجموعه‌دادگان استاندارد به زبان فارسی را نشان می‌دهد.

پیشتر ذکر شده که پایگاه‌داده گفتار هیجانی برلین از جمله بهترین انواع مجموعه‌دادگان گفتار هیجانی است که مکرراً توسط پژوهشگران مرود استفاده قرار می‌گیرد در ادامه با ویژگی‌ها و مراحل تولید آن بیشتر آشنا می‌شویم.

-
1. Persian Drama Radio Emotional Corpus (PDREC)
 2. Sahand Emotional Speech (SES) Database
 3. Persian Emotional Speech Database (Persian ESD)
 4. Sharif Emotional Speech Database (ShEMO)

جدول ۲. مشخصات برخی از مجموعه‌دادگان گفتار هیجانی فارسی زبان

نام	تعداد نمونه‌ها	گویندگان گفتار هیجانی	توضیحات	حالات هیجانی بیان شده
پیکره هیجانی درام رادیویی فارسی (اسماعیلیان و مروی، ۲۰۱۴)	۷۴۸	هنرپیشگان حرفه‌ای (۳۳ نفر: ۱۵ زن و ۱۸ مرد)	این نمونه‌ها از برنامه‌های رادیویی ضبط شده و دارای نویز محیطی هستند.	خشم، کسلی، انزجار، ترس، خنثی، غم، تعجب و شادی
پایگاه‌داده گفتار هیجانی فارسی حمیدی و منصوری‌زاده (حمیدی و منصوری‌زاده، ۲۰۱۲)	۲۴۰۰	هنرپیشگان حرفه‌ای (۳۳۰ نفر از زنان و مردان)	این نمونه‌ها از ۶۰ فیلم استخراج شده‌اند.	خشم، انزجار، ترس، شادی، غم و حالت خنثی
پایگاه‌داده گفتار هیجانی سهند (صدیقی، ۲۰۰۸)	۱۲۰۰	دانشجویان (۱۰ نفر: ۵ زن و ۵ مرد)		تعجب، شادی، غم، خشم و حالت خنثی
پایگاه‌داده گفتار هیجانی فارسی کشتیاری و همکاران، (۲۰۱۵)	۴۷۰	هنرپیشگان نیمه حرفه‌ای (۲ نفر: ۱ زن و ۱ مرد)	این پایگاه‌داده از نوع نیمه طبیعی است. چون گویندگان آن افرادی هستند که دوره بازیگری گذرانده‌اند.	خشم، انزجار، ترس، شادی و غم
پایگاه‌داده گفتار هیجانی شریف (محمدنظامی و همکاران، ۲۰۱۹)	۳۰۰۰	فارسی زبان عادی (۸۷ نفر: ۳۱ زن و ۵۶ مرد)		خشم، ترس، شادی، غم، تعجب و خنثی

در گام نخست برای تولید آن دعوت از افراد علاقه‌مند برای شرکت در پروژه است. افراد حاضر در این جامعه آماری از افرادی انتخاب شدند که در طی یک فراخوان به صورت یک آگهی در روزنامه منتشر شد، اعلام آمادگی کردند. سپس از این افراد خواسته شد تا در یک جلسه پیش انتخاب شرکت کنند. در این جلسه هر یک از افراد می‌بایست در یک اتاقک مقابل

یک میکروفون که مستقیماً به یک هارد دیسک^۱ (برای ذخیره‌سازی فایل گفتار) متصل است بایستند و یک پاره‌گفتار را با توجه به هیجان هدف بیان کنند. هیجان‌های هدف در پایگاه‌داده گفتار هیجانی شامل خشم^۲، ترس^۳، شادی^۴، غم^۵، کسلی^۶، انزجار^۷ و حالت خنثی^۸ است. در گام بعد سه نفر شنونده خبره^۹ این فایل‌های نمونه‌های گفتاری را با توجه به معیارهای طبیعی بودن و قابلیت بازشناسی اجراها مورد بررسی قرار دادند و در نهایت ۱۰ نفر را از میان این ۴۰ نفر برگزیدند.

سپس از این ۱۰ نفر خواسته شد تا در مجموع ۱۰ پاره‌گفتار که ۵ تای آنها پاره‌گفتارهایی شامل یک فراز و ۵ تای دیگر پاره‌گفتارهایی شامل دو فراز بودند را با توجه به هیجان‌های هدف خشم، ترس، شادی، غم، کسلی، انزجار و حالت خنثی بیان کنند. این پاره‌گفتارها به گونه‌ای گزینش شده‌اند که کلمات و ساخت و ساز نحوی آنها در زندگی روزمره مورد استفاده قرار می‌گیرند و عنصر متغیر در هنگام بیان آنها همان هیجان‌های هدف است و باقی عناصر ثابت می‌باشند تا میزان دخالت آنها در پژوهش به کمینه مقدار تقلیل یابد. بنابراین جملات استفاده شده در این پایگاه‌داده گفتار هیجانی باید برای حالت‌های هیجانی مختلف قابل تفسیر باشند و حاوی هیچ پیش‌فرض حالت هیجانی نباشند و البته شامل جملات معمولی باشند که در زندگی روزمره کاربرد دارند.

در هر جلسه ضبط، نمونه گفتار بیان شده توسط هر شرکت‌کننده تحت نظر سه آواشناس انجام می‌شود که تمامی مراحل این پروسه شامل ارائه دستورالعمل‌ها و نظارت بر تجهیزات را بر عهده دارند. سپس برای اطمینان از کیفیت حالت هیجانی و میزان طبیعی بودن نمونه‌های گفتاری بیان شده به صورت آزمون درک مورد بررسی قرار می‌گیرند. در این مرحله فایل‌های گفتار هیجانی به طور تصادفی و هر نمونه تنها یک مرتبه برای ۲۰ نفر پخش می‌شود تا حالت هیجانی فرد را بازشناسی کرده و بگویند که عملکرد آن برای بیان حالت هیجانی چقدر قانع‌کننده است. البته دو آزمون ادراک دیگر علاوه بر این مورد انجام شد، یکی برای قضاوت پیرامون

-
1. hard disk
 2. anger
 3. fear/anxiety
 4. happiness
 5. sadness
 6. boredom
 7. disgust
 8. neutral version
 9. Three Expert Listeners

قدرت یا شدت^۱ ارائه هیجان و دیگری برای قضاوت در مورد بیان تکیه^۲ در هجاهای هر پاره‌گفتار انجام گرفته است.

از دیدگاه فنی برای داشتن نمونه‌های گفتار هیجانی با بالاترین کیفیت، ضبط نمونه‌های گفتاری در یک اتاقک بدون پژواک^۳ که در دانشگاه فنی برلین^۴ قرار دارد، انجام شد. برای ضبط صدا از یک میکروفون Sennheiser MKH 40 P 48 و یک ضبط‌کننده قابل حمل Tascam DA-P1 استفاده شد.

در مجموع ۸۰۰ نمونه گفتار هیجانی تهیه شده است، که در کل حدود ۵۰۰ نمونه آن دارای نرخ بازشناسی بیش از ۸۰٪ و طبیعی بودن ۶۰٪ هستند که در دسترس پژوهشگران قرار داده شده است. (این پایگاه داده به صورت رایگان به نشانی داده شده قابل دسترسی است.^۵)

۳. روش

همانگونه که پیشتر بیان شد، در اینجا قصد داریم مراحل مختلف طراحی، تولید و آماده‌سازی یک مجموعه‌دادگان گفتار هیجانی مشابه با استانداردها، ویژگی‌ها و کیفیت پایگاه‌داده گفتار هیجانی برلین به زبان فارسی معیار و البته مقیاس کلان‌تر را به صورت مرحله به مرحله شرح دهیم.

مراحل مختلف طراحی، تولید و جمع‌آوری مجموعه‌دادگان گفتار هیجانی فارسی معیار منطبق بر مراحل انجام شده برای تولید پایگاه‌داده گفتار هیجانی برلین است. زبان این مجموعه‌دادگان گفتار هیجانی زبان فارسی معیار یا همان زبان رسمی کشور ایران است. به لحاظ فنی فرآیند تولید مجموعه‌دادگان گفتار هیجانی شامل چند گام است که عبارتند از:

۱. در گام نخست پاره‌گفتارهایی که قرار است با حالات هیجانی گوناگون بیان شوند، طراحی و تهیه می‌شوند. این پاره‌گفتارها به زبان فارسی معیار هستند. هر پاره‌گفتار باید از نظر واژگان و همچنین واجی و آوایی فاقد هرگونه بار هیجانی باشد. حالت هیجانی هر نمونه تنها باید در لحن گفتار توسط گوینده آن بیان شود.

۲. انتخاب یک آزمایشگاه یا استودیو ضبط صدا که دارای بیشترین شباهت به استودیو ضبط نمونه‌های مجموعه‌دادگان گفتار هیجانی در دانشگاه فنی برلین باشد. این نکته از این جهت دارای اهمیت است که بتوانیم استانداردهای در نظر گرفته شده برای تولید پایگاه‌داده گفتار هیجانی برلین را برای نمونه فارسی زبان آن رعایت کنیم.

1. strength

2. stress

3. anechoic chamber

4. Technical University Berlin

5. Berlin Database of Emotional Speech: <http://emodb.bilderbar.info/docu/>

۳. در این گام، از طریق یک فراخوان عمومی از علاقه‌مندان برای همکاری دعوت بعمل آمد. این افراد باید در بازه سنی ۲۰ تا ۵۰ باشند. زبان فارسی، زبان نخست آنها باشد. از نظر میزان تحصیلات در حد میانگین جامعه باشند. همچنین افراد به لحاظ سلامت جسمی، ذهنی و روانی هم مورد بررسی قرار می‌گیرند بگونه‌ای که شرایط آنها نمونه‌های تولید شده را تحت تأثیر قرار ندهد.

۴. در گام چهارم، از میان افرادی که علاقه‌مند به همکاری برای تولید مجموعه‌دادگان گفتار هیجانی هستند و همچنین از شرایط اولیه برخوردار هستند خواسته شد که در شرایط آزمایشگاهی پاره‌گفتارهایی را با حالت هیجانی خواسته شده تولید کنند. پس از بررسی این نمونه‌ها از میان آنها به تعداد برابر شرکت‌کننده زن و مرد انتخاب خواهد شد. در انتخاب این افراد این نکته در نظر گرفته می‌شود که افراد منتخب سخنوران عادی زبان فارسی باشند و هیچگونه آموزش حرفه‌ای برای فن بیان یا بازیگری نداشته باشند. همچنین نمونه‌هایی نیز از افرادی که به صورت حرفه‌ای دوره‌های فن بیان و بازیگری را نیز گذرانده‌اند را نیز تهیه خواهد شد (در اینجا نیز تعداد برابر شرکت‌کنندگان زن و مرد در نظر گرفته می‌شود). از طریق بخشی از برچسبی که برای هر نمونه در نظر گرفته می‌شود عادی بودن و حرفه‌ای بودن فرد را مشخص می‌کنیم.

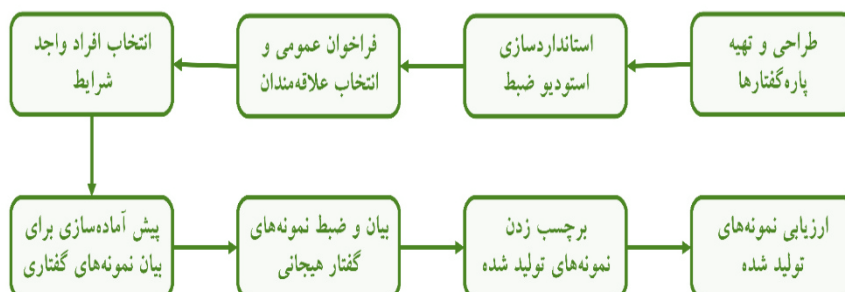
۵. در این گام، افراد منتخب با روند انجام کار که شامل چگونگی ایستادن مقابل میکروفون و آمادگی ذهنی برای ایجاد حالت هیجانی است آشنا می‌شوند. اما این آشنایی در حدی خواهد بود بیان افراد عادی از طبیعی بودن خارج نشود و مانند افراد حرفه‌ای آموزش دیده عمل نکنند. ۶. در این گام از افراد خواسته می‌شود تا پاره‌گفتارهای طراحی شده را با حالات هیجانی هدف بیان کنند. در اینجا حالات هیجانی هدف شامل حالات هیجانی خشم، شادی، غم، ترس، انزجار، کسلی و حالت خنثی است.

۷. هر نمونه تولید شده برچسب خاصی دارد که معرف گوینده آن، پاره‌گفتار بیان شده، حالت هیجانی هدف، جنسیت و حرفه‌ای یا غیرحرفه‌ای بودن گوینده است.

۸. پس از ضبط و برچسب‌گذاری، نمونه‌های گفتار هیجانی، تمامی آنها به صورت تصادفی و تنها یکبار برای حداقل ۵ نفر پخش خواهد شد. این افراد باید حالت هیجانی نمونه را بازشناسی کنند و عملکرد گوینده گفتار را ارزیابی کنند که عملکرد او قانع‌کننده هست یا خیر. بر اساس امتیاز آنها نمونه مد نظر در مجموعه‌دادگان گفتار هیجانی باقی می‌ماند و یا حذف می‌شود. پایگاه داده گفتار هیجانی برلین دارای نرخ بازشناسی بیش از ۸۰٪ و نرخ طبیعی بودن ۶۰٪ است.

مراحل فوق در شکل ۱ به صورت شماتیک نشان داده می‌شود.

شکل ۱. مراحل مختلف طراحی، تولید و ارزیابی مجموعه‌دادگان گفتار هیجانی فارسی معیار



۳-۱. طراحی پاره‌گفتارها

برای تولید نمونه‌های گفتار هیجانی در گام نخست نیاز به مجموعه‌ای از پاره‌گفتارها است که از نظر هیجانی فاقد هر گونه بار هیجانی باشد. فاقد بار هیجانی بودن در اینجا به این معناست که واژگان و نحو پاره‌گفتارها نباید برانگیزاننده و تغییر دهنده حالت هیجانی باشند و تنها لحن گفتار عامل تغییر بار هیجانی باشد. برای دستیابی به این هدف، در گام اول ۳۱ پاره‌گفتار به زبان فارسی معیار طراحی شد که از این میان ۲۰ پاره‌گفتار فاقد بار هیجانی هستند مانند "کفشم را پوشیدم." و بقیه پاره‌گفتارها دارای بار هیجانی هستند مانند "غم از دست دادن فرزند." پاره‌گفتارهای دارای بار هیجانی به منظور جلوگیری از سوگیری در ارزش‌گذاری پاره‌گفتارهای فاقد بار هیجانی طراحی شده بودند. این پاره‌گفتارها بر اساس حقایق علمی و فعالیت‌ها و گفتگوهای روزانه طراحی شده بودند. در گام بعدی پاره‌گفتارها تحت قالب یک فرم گوگل به صورت الکترونیکی از طریق پست الکترونیک در اختیار ۸ نفر برای ارزشیابی و ارزش‌گذاری قرار گرفت. این افراد شامل متخصصان روان‌شناسی، زبان‌شناسی، علوم‌شناختی-زبان، علوم‌شناختی-روان‌شناسی و ۲ نفر غیر مرتبط با این حوزه‌ها بودند. چیدمان و ترتیب پاره‌گفتارها برای هر یک افراد به صورت تصادفی و متفاوت با سایر افراد بود.

در مرحله ارزش‌گذاری افراد می‌بایست تمامی پاره‌گفتارها را ابتدا خوانده و سپس از میان هفت گزینه که هر گزینه بیان‌کننده یک حالت هیجانی خاص بودند، یک گزینه را انتخاب می‌کرد. هفت حالت هیجانی مورد نظر در اینجا عبارتند از خشم، ترس، شادی، غم، کسلی، انزجار و حالت خنثی.

در گام بعد بر اساس انتخاب افراد ۱۷ پاره‌گفتار از میان ۲۰ پاره‌گفتار فاقد بار هیجانی که گزینه حالت خنثی برای آنها به صورت کامل انتخاب شده بودند برای بیان با حالت‌های هیجانی انتخاب شدند. در جدول ۳ چند مورد از این پاره‌گفتارها را نشان می‌دهد.

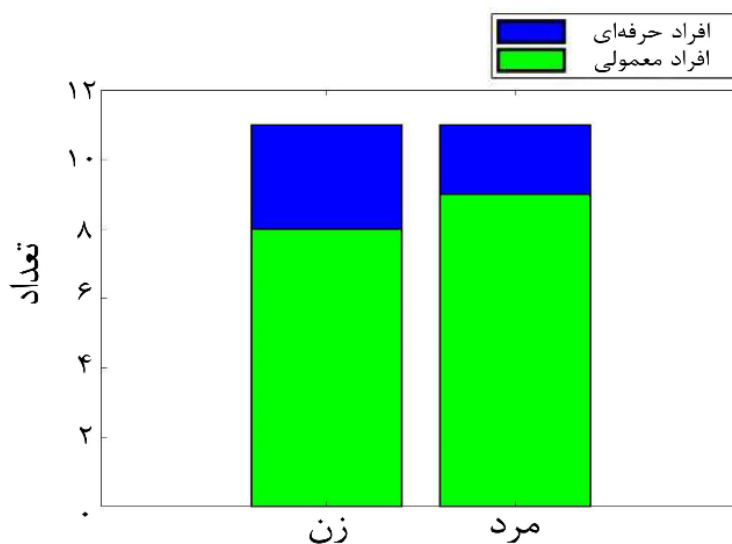
جدول ۳. چند نمونه از پاره‌گفتارهای فاقد بار هیجانی و دارای بار هیجانی

پاره‌گفتار	بار هیجانی
کفشم را پوشیدم.	فاقد بار هیجانی
هر سال چهار فصل دارد.	فاقد بار هیجانی
صبح کتری را آب کردم.	فاقد بار هیجانی
بدن به هشت ساعت خواب نیاز دارد.	فاقد بار هیجانی
غم از دست دادن فرزند	دارای بار هیجانی
صحنه جرم پر از خون بود.	دارای بار هیجانی

۳-۲. ویژگی‌های شرکت‌کنندگان

در این طرح از افراد با بازه سنی ۲۰ تا ۵۰ سال دعوت به مشارکت شد. این افراد برای انتخاب شدند می‌بایست فارسی زبان باشند (زبان اول آنها فارسی باشد) و قادر به خواندن و نوشتن باشند. افراد پس از انتخاب شدن برای شرکت در فرآیند تولید نمونه‌های گفتار هیجانی فرم رضایتنامه را مطالعه و امضاء کردند. در مجموع ۲۲ نفر شرکت‌کننده برای تولید نمونه‌های گفتار هیجانی شرکت کردند که ۵۰٪ شرکت‌کنندگان زن و ۵۰٪ دیگر شرکت‌کنندگان مرد بود. در میان این افراد ۵ نفر شامل ۳ شرکت‌کننده زن و ۲ شرکت‌کننده مرد بازیگر حرفه‌ای یا افرادی بودند که دوره بازیگری و فن‌بیان گذرانده بودند. در شکل ۲ توزیع جنسیتی شرکت‌کنندگان را نشان می‌دهد.

شکل ۲. توزیع جنسیتی شرکت‌کنندگان در فرآیند ضبط نمونه‌های گفتار هیجانی



۳-۳. ویژگی‌های فنی فرآیند ضبط صدا

ضبط نمونه‌های گفتار هیجانی به زبان فارسی در اتاقک آکوستیک گروه گفتاردرمانی دانشکده علوم توانبخشی دانشگاه علوم پزشکی ایران انجام شد. برای ضبط نمونه‌ها از یک میکروفون یقه‌ای بی‌سیم برند بویا^۱ مدل BY-V20 ۲.۴ گیگاهرتزی استفاده شد. نمونه‌های تولید شده با استفاده از نرم‌افزار Adobe Audition با نرخ نمونه‌برداری ۱۹۲۰۰۰ هرتز ضبط شدند.

۳-۴. چگونگی و حالت افراد در هنگام فرآیند ضبط صدا

تمامی شرکت‌کنندگان برای تولید نمونه‌های گفتار هیجانی آزاد بودند که به صورت ایستاده یا نشسته که بیشتر احساس راحتی می‌کردند در مقابل نمایشگر قرار بگیرند. به یقه آنها یک میکروفون بی‌سیم متصل می‌شد تا به لحاظ حرکتی هیچگونه مانعی برای آنها ایجاد نکند. سپس یکی پس دیگری با توجه به انتخاب خودشان یکی از حالات هفتگانه هیجانی هدف در این طرح را برمی‌گزیدند. در گام بعد، هر پاره‌گفتارها را که در نمایشگر به آنها نشان داده می‌شد سه بار با آن حالت هیجانی انتخابی بیان می‌کردند. شایان ذکر است که ترتیب نمایش پاره‌گفتارها در هر حالت هیجانی متفاوت از سایر حالات بود تا از سوگیری شرکت‌کنندگان ممانعت به عمل آید. شکل ۳ نمایی از استقرار یکی از شرکت‌کنندگان را نشان می‌دهد.

شکل ۳. نمایی از استقرار یکی از شرکت‌کنندگان در طی فرآیند ضبط نمونه‌های گفتار هیجانی



۴. یافته‌ها

در طی فرآیند تولید نمونه‌های گفتار هیجانی به زبان فارسی معیار در مجموع ۲۲ شرکت‌کننده داشتیم که با توجه به هفت حالت هیجانی هدف خشم، ترس، شادی، غم، کسلی، انزجار و حالت خنثی، هر یک از ۱۷ پاره‌گفتار فاقد بار هیجانی را به طور متوسط ۳ بار بیان کردند و در مجموع ۷۸۵۴ نمونه اولیه خام را تولید شده است. شایان ذکر است که این گزارش ارائه شده مربوط به فاز طراحی و جمع‌آوری نمونه‌های گفتار هیجانی است. فاز بعدی شامل برچسب‌گذاری و ارزیابی نمونه‌های گفتار هیجانی تولید شده توسط متخصصان این حوزه است تا نمونه‌هایی که دارای نرخ بازشناسی بیش از ۸۰٪ و طبیعی بودن ۶۰٪ هستند از میان تمامی نمونه‌ها انتخاب شوند و تحت یک پایگاه داده استاندارد در اختیار پژوهشگران و علاقه‌مندان این حوزه قرار گیرد.

۵. بحث و نتیجه‌گیری

الگوهای گفتاری حامل‌های اصلی پیام‌های ارتباطی می‌باشند. هر الگو از دو بخش نحو (قالب) و معنا (محتوا) برخوردار است. از منظر زبان‌شناسی، نحو^۱ الگوی گفتاری مشخصات ساختاری پیام را دربرگرفته و معنای^۲ آن دربرگیرنده معنی و مشخصات شناختی الگو است. بنابراین، الگوی قالب، ممزوج بالحنی است که محتوای آنرا تحت تأثیر قرار می‌دهد و می‌تواند هیجان را به عنوان یک ویژگی شناختی حمل نموده و منشاء تولید یک پدیده ارتباطی گردد.

زبان‌های مختلف به لحاظ ساختاری در فرآیند ارتباط، حامل ویژگی‌های هیجانی متفاوتی هستند. برخی از این ویژگی‌ها، حامل خصوصیات هیجانی نژادها و ملل مختلف هستند که در هر یک رفتاری متناظر نیز رقم خورده است. فقدان مجموعه‌دادگان به زبان فارسی که از استانداردها، کیفیت و ویژگی‌های مناسب برخوردار باشد سبب شده تا دانش ما پیرامون الگوهای هیجانی و تأثیر آن در زبان فارسی به وضوح روشن و شفاف نباشد. همین نکته باعث شد تا اقدام به طراحی، تهیه و تولید مجموعه‌دادگان گفتار هیجانی به زبان فارسی معیار کنیم. مجموعه‌دادگان گفتار هیجانی به زبان فارسی در مطالعات پردازش گفتار و بازشناسی هیجان گفتار در سطح دانشگاهی، صنعتی و درمانی بسیار حائز اهمیت است. نتایج حاصل از این مطالعات دانش ما پیرامون الگوهای هیجانی و تأثیر آن در زبان فارسی را افزایش می‌دهد که در کاربردهای هوش مصنوعی، تعامل انسان و ماشین، سامانه‌های خدمات رسانی خودکار، کاربردهای آموزشی و درمانی بسیار مفید و کارآمد است. به عبارت دیگر نتایج پژوهش روی

1. syntax

2. semantic

این دادگان، برای طراحی واسطه‌های انسان و ماشین، طراحی نرم‌افزارها، برنامه‌های کاربردی و پروتکل‌های آموزشی، درمانی و توانبخشی بسیار سودمند و ضروری است. در این مقاله گزارشی از مراحل مختلف طراحی، تهیه و تولید پایگاه‌داده گفتار هیجانی به زبان فارسی معیار ارائه شد. کشور ایران از جمله غنی‌ترین کشورها از نظر زبان‌ها و گویش‌ها در سطح جهان محسوب می‌شود. امید است ارائه این گزارش سبب شود تا علاقه‌مندان به حوزه بازشناسی گفتار هیجانی و زبان‌شناسی رایانشی در سطح کشور اقدام به تهیه و تولید پایگاه‌داده‌های گفتار هیجانی با زبان‌ها و گویش‌های گوناگون در سراسر کشورمان ایران نمایند تا دانش‌مان پیرامون هیجان و الگوهای هیجانی به عنوان عوامل بسیار بنیادین و تأثیرگذار در مطالعه زبان افزایش یابد.

تعارض منافع

تعارض منافی وجود ندارد.

سپاسگزاری

نخست از استاد ارجمند سرکار خانم دکتر شهین نعمت‌زاده برای تمامی راهنمایی‌ها و همراهیشان کمال قدردانی و تشکر را داریم. همچنین از تمامی افرادی که در ارزیابی پاره‌گفتارها شرکت کردند و از شرکت‌کنندگان محترم برای تولید نمونه‌های گفتار هیجانی بسیار سپاسگزاریم. همچنین از مؤسسه آموزش علوم شناختی و ستاد علوم شناختی برای تمامی حمایت‌ها قدردان و سپاسگزاریم.

منابع

- علی‌نژاد، بتول (۱۳۸۴). بررسی رابطه بین ویژگی‌های عروضی گفتار فارسی و نقش‌های کاربرد شناختی آنها (مبتنی بر مطالعه موردی)، *مجله علمی- پژوهشی دانشکده ادبیات و علوم انسانی دانشگاه اصفهان*، دوره دوم، شماره ۴۲ و ۴۳ (۲۲۶-۱۹۹).
- علی‌نژاد، بتول؛ حسینی بالام، فهیمه (۱۳۹۲). *مبانی آواشناسی آکوستیکی*، انتشارات دانشگاه اصفهان، اصفهان.
- علی‌نژاد، بتول؛ ویسی، الخاص (۱۳۸۶). بررسی رابطه بین ویژگی‌های کاربردشناختی آوایی و بیان عواطف در زبان فارسی (مطالعه موردی)، *هفتمین همایش زیان‌شناسی ایران*، (۱۶۴-۱۴۳).
- قاسم‌زاده، حبیب‌الله (۱۳۸۵). *مقدمه‌ای بر علم شناختی*، نشریه درون‌سازمانی.

- Alinezhad, B (2005). A study of the relationship between the prosodic features of Persian speech and their cognitive functional roles (based on a case study), Scientific-Research Journal of the Faculty of Literature and Humanities, University of Isfahan, 2(42 & 43), pp. 226-199. [In Persian]
- Alinajad, B.; Vays, E. (2007). Investigating the relationship between phonetic cognitive features and the expression of emotions in Persian (case study), *7th Conference of Linguistics in Iran*, Iran, pp. 164-143. [In Persian]
- Alinezhad, B. (2010), A Study of the Relationship between Acoustic Features of “bæle” and the Paralinguistic Information, *Journal of Teaching Language Skills*. Shiraz, 2(1), pp. 1–26.
- Alinezhad, B. and Hosseinibalam, F. (2013). *Fundamentals of acoustic phonetics*. University of Isfahan Press. Isfahan, Iran. [In Persian]
- Bishop, C. M. (2006). *Machine Learning and Pattern Recognition*. Springer.
- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. and Weiss, B. (2005). A Database of German Emotional Speech, in *Proc. Interspeech, Lisbon*, pp. 1517–1520.
- Damasio, A. R. (1999), *The feeling of what happens: body and emotion in the making of consciousness*, Harcourt Brace, New York.
- Deller, Jr. John R., Hansen, John H.L. and Proakis, John G. (1999). *Discrete-Time Processing of Speech Signals*, Wiley-IEEE Press, Classic Reissue.
- Edelman, G. M. (2004). *Wider than the sky: The phenomenal gift of consciousness*. Yale University Press, USA.
- Ekman, P. (2007). *Emotions Revealed_ Recognizing Faces and Feelings to Improve Communication and Emotional Life*. 2nd edition, Holt Paperbacks, New York, USA.
- Engberg, I. S., & Hansen, A. V. (1996). Documentation of the Danish emotional speech database des. *Internal AAU Report, Center for Person Kommunikation, Denmark*, 22.
- Esmailyan z., and Marvi H.(2014). “A Database for Automatic Persian Speech Emotion Recognition: Collection, Processing and Evaluation.” *International Journal of Engineering, Transactions A: Basics* 27(1):79–90.
- Franti, E., Ispas, I., Dragomir, V., Dasc, M., Alu, Zoltan, E., & Stoica, I. C. (2017). Voice Based Emotion Recognition with Convolutional Neural Networks for Companion Robots. *Romanian Journal of Information Science and Technology*, 20(3), 222–240. Retrieved from <http://www.romjst.ro/full-texts/paper562.pdf>

- Friedenberg, J. and Silverman, G. (2006). *Cognitive science: an introduction to the study of mind*, Sage Publications.
- Ghasemzadeh, H. (2006). An Introduction to Cognitive Science, *Internal Publication*, Iran. [In Persian]
- Hamidi M., and Mansoorizade M. (2012). "Emotion Recognition from Persian Speech with Neural Network." *International Journal of Artificial Intelligence & Applications* 3(5):107–12.
- Hasrul, M., Hariharan, M. and Sazali, Y. (2011). Human Affective (Emotion) Behavior Analysis using Speech Signals: A Review, *Proceedings of the 2012 International Conference on Biomedical Engineering*, ICoBE.
- Keshtiari, N., Kuhlmann, M., Eslami, M., & Klann-Delius, G. (2015). Recognizing emotional speech in Persian: A validated database of Persian emotional speech (Persian ESD). *Behavior Research Methods*, 47(1), 275–294. <https://doi.org/10.3758/s13428-014-0467-x>
- Kleinginna, P. R. & Kleinginna, A. M. (1981). A categorized list of motivation definitions, with a suggestion for a consensual definition. *Motiv Emot* Volume 5, Issue 3, pp 263–291. <https://doi.org/10.1007/BF00992553>
- Luria, A. R. (1982). *Language and Cognition*. Wiley.
- Mohamad Nezami, O., Jamshid Lou, P., & Karami, M. (2019). ShEMO: a large-scale validated database for Persian speech emotion detection. *Language Resources and Evaluation*, 53(1), 1–16. <https://doi.org/10.1007/s10579-018-9427-x>
- Morrison, D., Wang, R., & De Silva, L. C. (2007). Ensemble methods for spoken emotion recognition in call-centres. *Speech Communication*, 49(2), 98–112. <https://doi.org/10.1016/j.specom.2006.11.004>
- Nwe, T. L., Foo, S. W., & De Silva, L. C. (2003). Speech emotion recognition using hidden Markov models. *Speech Communication*, 41(4), 603–623. [https://doi.org/10.1016/S0167-6393\(03\)00099-2](https://doi.org/10.1016/S0167-6393(03)00099-2)
- Picard, R. W. (1995). Affective Computing. *MIT Media Laboratory Perceptual Computing Section Technical Report*, (321), 26. Retrieved from <http://www.media.mit.edu/~picard/>
- Sedaaghi, M. (2008). *Documentation of the Sahand Emotional Speech Database (SES)*. Technical report, Department of engineering, Sahand University of Technology.
- Slaney, M., & McRoberts, G. (1998). Baby Ears: a recognition system for affective vocalizations. *Proceedings of the 1998 IEEE International Conference on*

Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181), 2, 985-988 vol.2.

Thagard P. (2005). *Mind: Introduction to cognitive science*. Cambridge, MA: MIT Press.