

Title: Speech Processing Ontology

Author: Razieh Zare

(<mailto:Zare.Razieh98ir@gmail.com>)

Abstract

The field of speech science is rapidly advancing, necessitating a structured framework to categorize and relate its diverse tasks. This article introduces a pioneering ontology aimed at encapsulating all speech processing tasks, providing a comprehensive resource for researchers and practitioners. Following Bermejo's methodology, this ontology was developed from scratch, as existing resources failed to cover the breadth of speech processing tasks. The ontology is informed by seminal books and articles in the field, as well as cutting-edge research on emerging. Personal expertise in speech recognition further enriches the ontology's development. The article details the particular process of extracting terminologies, creating classes, and designing a hierarchy that optimizes the ontology without compromising its integrity. With 79 main concepts, the ontology presents a robust structure, excluding instances and attributes to maintain focus on core concepts. The most challenging aspect was designing a hierarchy that accurately reflects the interrelatedness of classes. The ontology also defines relationships and properties, ensuring a dynamic and adaptable tool that evolves with the field. This work stands as a significant contribution to speech science, offering a foundational tool that enhances research and application, and sets a precedent for future ontological endeavours in the field.

Introduction

In the dynamic and interdisciplinary field of speech science, the development of a comprehensive ontology is not merely an academic exercise but a foundational tool that can significantly enhance research and application. Speech science encompasses a multitude of tasks, each with its own set of methodologies, terminologies, and challenges. From the nuanced complexities of speech recognition to the intricate

designs of speech synthesis, the domain requires a systematic approach to categorize and relate various concepts and tasks. This is where the ontology comes into play.

The primary purpose of this ontology is to serve as a centralized repository that encapsulates the entire spectrum of speech processing tasks. It aims to provide a structured and coherent framework that researchers and practitioners can reference to understand the field's breadth and depth. By doing so, it facilitates a more efficient exchange of knowledge, promotes consistency in terminology, and fosters collaborative research efforts across different sub-disciplines.

Moreover, the ontology is designed to be inclusive and adaptable, accommodating the rapid advancements in speech science. As new technologies emerge and existing techniques evolve, the ontology can be updated to reflect these changes, ensuring that it remains a relevant and valuable resource for the community.

For researchers, the ontology offers a clear and organized overview of the field, enabling them to identify potential areas of study and gaps in the current body of knowledge. For practitioners, it provides a practical guide to the various speech processing tasks, helping them to select the most appropriate methods and tools for their specific needs.

Furthermore, the ontology is not just a static collection of terms and definitions; it is a living document that captures the dynamic nature of speech science. It is constructed with the flexibility to grow and adapt, ensuring that it can accommodate future discoveries and innovations in the field.

In essence, the ontology is envisioned as a cornerstone for the speech science community, supporting and guiding both seasoned experts and newcomers alike. It stands as a testament to the collaborative spirit of the field, embodying the collective knowledge and experience of its members. With this ontology, we unlock the potential for a more integrated and comprehensive understanding of speech processing.

Methodology

The methodology employed in this ontology project is rooted in the structured approach proposed by Bermejo; 2007, which provides a systematic framework for the construction of knowledge-based applications. This approach was meticulously

followed, with adaptations made to address the unique challenges presented by the domain of speech science.

Domain Scope and Final Users:

The ontology's domain scope is firmly rooted in speech science, specifically targeting all speech processing tasks. This broad scope ensures that the ontology encompasses a comprehensive range of activities within the field, from speech recognition to speech synthesis, and everything in between. The final users of this ontology are as diverse as its applications; they include researchers in the field of speech science, engineers working on speech processing technologies, and even students or newcomers seeking to understand the landscape of speech-related tasks.

Resource Identification and Analysis:

The first step involved a comprehensive search for existing ontologies and resources within the field. This search revealed a significant gap: while there were ontologies related to speech, none focused specifically on speech processing tasks. The absence of such a resource underscored the necessity of creating an ontology from the ground up, marking this endeavour as the first of its kind in the field.

Terminology Extraction:

The process of terminology extraction was exhaustive and multifaceted, delving into the rich areas of speech science literature. It involved a meticulous review of main books and articles that have long been considered foundational in the field, ensuring that established concepts were fully integrated into the ontology. These seminal works provided a solid base of traditional terms that are essential for any comprehensive resource in speech processing.

In addition to these cornerstone texts, the exploration extended to the cutting edge of speech science, incorporating terms from the newest articles on emerging topics. One such term is "speech transmission over network," a concept that has gained prominence with the advent of advanced networking technologies and their application in speech processing. This term, among others, could not be found in basic books, as it represents a newer area of study that has emerged with technological advancements.

Drawing from my personal expertise in speaker recognition, I ensured that the ontology reflected the current landscape of the field, including the nuances and specialized knowledge that come with years of focused research. This personal touch

was crucial in capturing the full breadth of the field, from its historical roots to the innovative frontiers that continue to propel speech science forward. This comprehensive approach ensures that the ontology remains relevant and useful for both established researchers and those on the forefront of speech science.

Class Creation and Designing the Hierarchy:

The ontology comprises 79 main concepts, each carefully selected to represent a core aspect of speech processing. These concepts range from 'Speech Transmission' to 'speaker Recognition', covering the full spectrum of tasks within the field.

The inclusion of each concept was justified based on its prevalence in the literature, its relevance to current speech processing tasks, and its significance to the field's future. Concepts that did not meet these criteria were either reclassified or excluded to maintain the ontology's focus and utility.

Instances and attributes were excluded from these main concepts to prevent the ontology from becoming unwieldy. Instead, they were categorized separately, allowing for a cleaner and more navigable structure. This decision also facilitates the ontology's scalability, as new instances and attributes can be added without disrupting the established hierarchy of main concepts.

The ontology was optimized by ensuring that each class served a distinct purpose and contributed to the ontology's overall comprehensiveness. Redundant or overly granular classes were merged or redefined, streamlining the structure without compromising the depth of information. This optimization was crucial in maintaining the ontology's integrity and facilitating its practical application.

Defining Relationships and Properties:

In the ontology, relationships and properties serve as the connective tissue between classes, providing depth and functionality. Relationships such as 'followed by' and 'precedes' were meticulously defined to illustrate the delay and advance sequence, such as the link where 'Speech Coding' precedes 'Speech Transmission'. Properties, on the other hand, were assigned to classes to specify characteristics and constraints. For instance, 'Accuracy' was defined as a property with a value less than 100, adding a quantitative measure to evaluate performance. These elements were crucial in transforming the ontology from a static list of terms into a dynamic and interconnected framework, enhancing its utility for the speech science community.

Addressing Overlaps in Terminology:

A notable challenge arose when defining the classes for speaker verification and voice discrimination. These terms, while often used interchangeably, have distinct connotations within different subfields. Speaker verification is commonly associated with machine learning and automated systems, whereas voice discrimination is more aligned with humanistic studies and speech perception.

To resolve this overlap, the terms were initially defined as separate classes. However, recognizing the inherent relationship between the two, an "IsA" relation was established, positing that speaker verification is a form of voice discrimination. This hierarchical relationship acknowledges the shared essence of the tasks while preserving their distinct applications. Further clarity was achieved by defining data properties for the speaker verification class, such as "text-dependent" and "text-independent," to delineate the specific nature of the verification tasks. These properties serve to differentiate speaker verification processes and highlight their unique characteristics within the broader context of voice discrimination.

Conclusion

This ontology project represents a significant stride in the field of speech science, providing a structured and comprehensive framework that encapsulates the full spectrum of speech processing tasks. Through meticulous research and careful analysis, we have established an ontology that not only serves as a theoretical resource but also as a practical tool for researchers, engineers, and students alike. The 79 main concepts, along with their defined relationships and properties, offer a clear and navigable guide through the complexities of speech science. As the first ontology of its kind, it sets a precedent for future work in the domain and invites ongoing contributions and refinements. Ultimately, this ontology stands as a testament to the collaborative and ever-evolving nature of speech science, reflecting both its rich history and its innovative future.

References

Bermejo, J. (2007). A simplified guide to create an ontology. *Madrid University*, 1-12.

Jurafsky, D., & Martin, J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.