July 16, 2018

# Automated Speech Pattern Generator for Natural Language Output

Hemamalini Manickavasagam

David Yu Chen

Alissa Scherchen

# Automated speech pattern generator for natural language output

## ABSTRACT

Computer-generated text or speech, e.g., as utilized by assistant applications and bots with voice capabilities, utilizes words and phrases that are chosen carefully by the underlying language generation functions. Such text/speech usually lacks the nuance of real-world human speech that often provides rich indications of the personality, sociocultural characteristics, feelings, and thoughts of a human speaker. The techniques of this disclosure are usable to generate text/speech that expresses personal characteristics akin to a human speaker, and support representing a wide variety of personalities. The techniques involve preparing speech models of human characters by analyzing language use in existing conversation texts to cluster speech with similar word choices and speech patterns into character templates. When an assistant application or bot delivers output via text or voice, an appropriate character is chosen from among the templates. The characteristics of the template are applied to revise the originally generated response content to fit the character template.

## KEYWORDS

- Speech pattern generator
- Natural language synthesis
- Voice interface
- Speech variation
- Voice character model
- Smart speaker
- Virtual assistant

## BACKGROUND

Computer-generated text or speech, e.g., utilized by programs such as assistant applications and bots with voice capabilities, utilizes words that are chosen carefully by the underlying language generation functions. Such text/speech typically lacks the nuance of real-world human conversation that often provides rich indications of the personality, sociocultural characteristics, feelings, and thoughts of a human speaker. People interacting with such programs and devices such as smartphones, wearable devices, home speakers, appliances, etc. that provide voice interaction expect these interactions to mimic real-world conversations with other humans. Owing to the lack of personal expression in computer-synthesized language voiced by such programs, these interactions often come across as artificial and awkward.

Designers sometimes mitigate this issue by manually selecting specific personal characteristics for a program such as a bot or assistant application that utilizes computer-generated speech and adjusting the generated dialog to fit the selection. However, such an approach requires substantial time and effort and is difficult to scale in order to support a large number of personalities for the generated text or voice. As a result, such approaches typically analyze limited sets of past responses, cover only known speech patterns, and support only a small set of variations.

## DESCRIPTION

The techniques of this disclosure are usable to generate text and/or speech that expresses personal characteristics akin to a human speaker, and support representing a wide variety of personalities. The techniques involve preparing speech models of human characters based on analyzing language use in existing conversation texts, such as books, movies, online posts, etc.

The input text and associated metadata are analyzed to identify and classify various relevant characteristics within the text, such as location, time, attributes of the speaker, attributes of the listeners, emotions, etc. For instance, analysis of the text of a movie script includes examining content of the speech, description of the characters, and metadata for the movie from sources such as the Internet Movie Database (IMDB). The analysis includes clustering speech with similar word choices and speech patterns into character templates.

The templates are further refined based on emotions and listener attributes. For each of the refined templates, each exemplar standard word is associated with a list of corresponding words and phrases. For example, the exemplar word “yes” may be associated with a list of words such as “yeah,” “OK,” “alright,” etc. Similarly, common grammar structures, such as nouns, verbs, etc., are generated for each of the refined templates.

Upon request to deliver output via voice, e.g., by an assistant application or a bot, the character templates are searched for an appropriate match based on relevant contextual and personal characteristics, such as location, time, age, gender, emotional context, etc., if the user permits the use of contextual information for such personalization. If the user does not permit the use of contextual information, alternate selection mechanisms, such as manual input, developer specified choices, default settings, are used.

An appropriate character is chosen from among the matching templates. The template includes the associated character traits, such as emotions, attributes, etc. The characteristics of the template are applied to revise the originally generated response content to adjust the word and grammar choices to fit the character template. The revised content is delivered as voice output, or based on the context, text output.

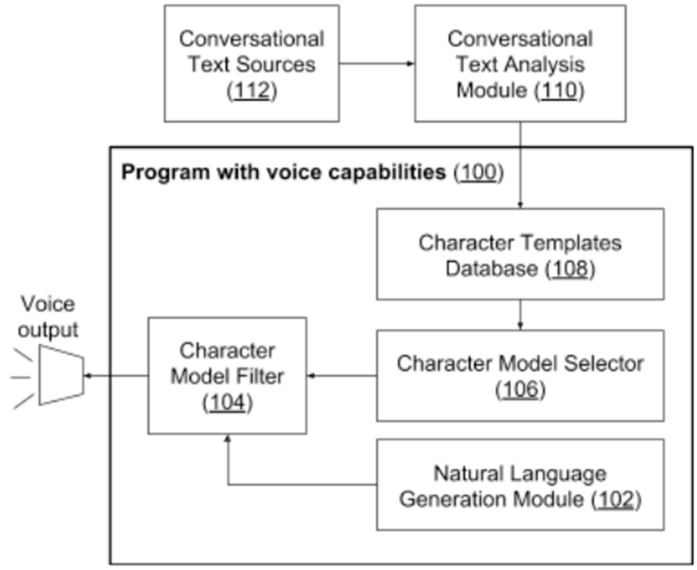
Conversational Text Sources (112)
Conversational Text Analysis Module (110)
Program with voice capabilities (100)
Character Templates Database (108)
Voice output
Character Model Filter (104)
Character Model Selector (106)
Natural Language Generation Module (102)

**Fig. 1: Adjusting natural language speech based on a character model**

Fig. 1 illustrates adjusting natural language speech based on a character model, per techniques of this disclosure. A natural language generation module (102), e.g., within an assistant program or bot with voice capabilities (100), is provided to generate the output speech for interactive dialog with a user. The generated text of the intended output speech is passed to a character model filter (104).

Simultaneously, with user permission, contextual aspects and suitable personal voice characteristics are examined by a character model selector (106) to select an appropriate matching character model from the character templates database (108) generated by the conversational text analysis module (110) that analyzes and classifies conversational text (110) from a variety of sources. The originally synthesized speech is adjusted by the character model filter according to the linguistic characteristics of the template selected by the character model filter. For example, “yes” may be replaced by “alright” if such a replacement results in a better fit with the speech choices of the selected character model. The revised output of the character model filter is then delivered as the voice output.

The techniques of this disclosure enable separation of the generation of the knowledge content of the natural language output to be delivered via speech from the nature and characteristics of the voice delivery of the content, e.g., that mimics a human character. Moreover, the techniques enable adjusting the characteristics of the speech output in an automated and scalable way. Thus, the techniques can support a large variety of voice character models.

The operation allows personalizing content delivery with varying voice character models akin to the tailoring the appearance of graphical avatars via various skins, thus enabling creative voice-based delivery and interaction that can be utilized for storytelling and branding. In addition to automatically generated content, the techniques of this disclosure can be extended to support manually specified dialog. The speech output techniques can also be provided as a web service or an Application Programming Interface (API) that enables developers to share and embed the features across applications and sites while providing fast performance and user experience comparable to native applications.

While the foregoing discussion refers to voice output from a software program, the described character templates and speech synthesis can also be used by writers, e.g., of books, movies, plays, video games, etc.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user’s social network, social actions or activities, profession, a user’s preferences, or a user’s current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

## CONCLUSION

Computer-generated text or speech, e.g., as utilized by assistant applications and bots with voice capabilities, utilizes words and phrases that are chosen carefully by the underlying language generation functions. Such text/speech usually lacks the nuance of real-world human speech that often provides rich indications of the personality, sociocultural characteristics, feelings, and thoughts of a human speaker. The techniques of this disclosure are usable to generate text/speech that expresses personal characteristics akin to a human speaker, and support representing a wide variety of personalities. The techniques involve preparing speech models of human characters by analyzing language use in existing conversation texts to cluster speech with similar word choices and speech patterns into character templates. When an assistant application or bot delivers output via text or voice, an appropriate character is chosen from among the templates. The characteristics of the template are applied to revise the originally generated response content to fit the character template.