CALLER ACOUSTIC PUSH TO ANSWERS

Amit Agnihotri
Nivedha Thangaraju
Prabhakaran Suyambu
Gowtham J. A

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Agnihotri, Amit; Thangaraju, Nivedha; Suyambu, Prabhakaran; and A, Gowtham J., "CALLER ACOUSTIC PUSH TO ANSWERS", Technical Disclosure Commons, (July 01, 2020)
https://www.tdcommons.org/dpubs_series/3393

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
CALLER ACOUSTIC PUSH TO ANSWERS

AUTHORS:
Amit Agnihotri
Nivedha Thangaraju
Prabhakaran Suyambu
Gowtham J A

ABSTRACT
Techniques are described herein for caller acoustic properties extraction by sampling an initial duration of the connected Real-time Transport Protocol (RTP) caller stream and analyzing a received transcript of the ongoing live stream. A Speech Synthesis Markup Language (SSML) mapping of the caller acoustic information is generated in the most suitable customer-understandable format. These techniques may provide automated identification and enable Push To Answer when the caller is not able to understand what is being communicated by the agent.

DETAILED DESCRIPTION
Contact centers deliver accurate and efficient information to solve customer issues. However, oftentimes the caller is unable to understand what is being communicated to them, in spite of multiple attempts by the agent, even though that information is correct and accurate and can resolve the caller-customer issue. This can lead to unsatisfied customers and even questioning of the capabilities of the agent.

This scenario can arise because of caller limitations in hearing capabilities, agent limitations in speaking capabilities, or a compatibility issue between agent speech and caller hearing/listening. There can also be limitations regarding the number of agents and agent capabilities. The day in and day out of the number of cases to be handled might be too high, and sharing the same information to multiple callers may end up over-exhausting the agent. Furthermore, the contact center may have a noisy environment for any number of reasons.

Accordingly, techniques are provided herein relating to caller acoustic Push To Answer. This can empower the efficient and effective communication of accurate information to resolve customer issues. In one example, an agent desktop "Caller Acoustic
Push To Answer" may identify caller acoustic properties at run time from caller stream analysis.

The "Caller Acoustic Push To Answer" may sample an initial duration of the connected Real-time Transport Protocol (RTP) caller stream to obtain the voice pattern of the caller using waveform and acoustic features. The caller pitch (frequency) and pauses (silence between words) may be fetched from the sampled data, as the easiest way for the caller to understand the information will be in a manner that is similar to their own style of speaking.

The "Caller Acoustic Push To Answer" may analyze a transcript of the ongoing live RTP stream for different keyword patterns, which exposes the non-clarity of the information received at the caller side. In spite of having a similar style of speaking, the caller may not be able to understand the information being communicated, which further requires proper tuning. Keywords from the caller such as "say it slowly", "speak louder", "I didn't get the amount," etc., may be modeled.

Sampling the RTP and analyzing the transcript may help to set the optimal communication values for pitch, pace, pauses, etc. for the Caller. These properties are used to map and set different Speech Synthesis Markup Language (SSML) properties such as emphasis, break, prosody, etc. This solution may use any suitable local model or transcribe feature.

Table 1 below illustrates an example mapping of the caller acoustic signal to SSML.

<table>
<thead>
<tr>
<th>Transcribe-RTP</th>
<th>SSML Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Say it slowly&quot;</td>
<td>prosody rate='x-slow'</td>
</tr>
<tr>
<td>&quot;Say it louder&quot;</td>
<td>prosody pitch='high'</td>
</tr>
<tr>
<td>&quot;I missed the details&quot; or &quot;I didn't get the amount&quot;</td>
<td>emphasis level='strong'</td>
</tr>
<tr>
<td>RTP pauses in caller speech from sampled packets</td>
<td>break strength='medium'</td>
</tr>
<tr>
<td>RTP caller pitch from sampled packets</td>
<td>prosody pitch='medium'</td>
</tr>
</tbody>
</table>

This enables easy identification and notification if the information being communicated from the agent to the caller (i.e., the unsatisfied customer) is not clear.
Keywords from caller such as "repeat it," "say it again," "I didn't hear what you said," etc., can be modeled for feature activation notification on the agent desktop.

Analyzing ongoing live transcripts may help to automatically enable this feature functionality on the agent desktop. An agent might be reluctant to accept that the caller does not understand the information being communicated by the agent.

Figure 1 below illustrates an example flow for sampling using a connected RTP caller stream and activation using transcription.
Figure 2 below illustrates an example system configured for live transcript analysis for playback properties and activation using transcription.

![Figure 2](https://www.tdcommons.org/dpubs_series/3393)

Figure 2

Figure 3 below illustrates an example system wherein the agent can also manually trigger the Push to Answer feature.

![Figure 3](https://www.tdcommons.org/dpubs_series/3393)

Figure 3

This solution empowers the agent to have sample reference templates as well. In one example, only information crucial for the customer is played back in a more understandable format. Playback parameters may be automatically set based on the RTP sampling and transcript analyses. The agent can also reset these parameters as needed/wanted.
Figure 4 below illustrates a sample back-end template. As shown, the appropriate level of emphasis and other suitable parameters are determined for the information desired by the customer.

```
<prosody rate='slow' volume='loud'>
Thanks for contacting the bank. <break strength='x-weak'/>
We are happy to answer your query on your account balance.
<emphasis level='strong'>
Balance in your account is
</emphasis>
<Say-as Type="Currency">$250.32</Say-as>
</prosody>
```

*Figure 4*

The agent desktop Push to Answer may also have an improved control drop down to fine tune these SSML parameters. They can optimize the properties to match the caller to their most comfortable/understandable format.
Figure 5 below illustrates an example call flow diagram.
In summary, techniques are described herein for caller acoustic properties extraction by sampling an initial duration of the connected RTP caller stream and analyzing a received transcript of the ongoing live stream. A SSML mapping of the caller acoustic information is generated in the most suitable customer-understandable format. These techniques may provide automated identification and enable Push To Answer when the caller is not able to understand what is being communicated by the agent.