Technical Disclosure Commons

Defensive Publications Series

December 21, 2018

IDENTIFYING AI PARTICIPANTS IN CALLS AND ESCALATING AI-ONLY CALLS INTO AN OPTIMIZED MACHINE COMMUNICATION MODE

Denis Savostiyanov

Rob Hanton

IDENTIFYING AI PARTICIPANTS IN CALLS AND ESCALATING AI-ONLY CALLS INTO AN OPTIMIZED MACHINE COMMUNICATION MODE

AUTHORS:
Denis Savostiyanov
Rob Hanton

ABSTRACT

The embodiments presented herein enable two human-emulating artificial intelligence (AI) speech agents to identify each other in situations where the two are engaged in a conversation. Once two speech agents recognize each other as artificially intelligent, they may transition to a more efficient method of communication in order to complete tasks more quickly, accurately, and inexpensively.

DETAILED DESCRIPTION

Recent advances in AI calling engines are reaching a point where an artificial intelligence (AI) can interact with a human participant using speech to perform a specific function, such as making a dinner reservation on a user's behalf. As this functionality is more widely adopted, it is increasingly likely that situations may arise in which both participants in a conversation are AI agents simulating human speech. For example, an AI agent calling a restaurant to make an appointment may converse with another AI agent arranging the bookings. Compared to other inter-process communications, using simulated speech over telephony is extremely slow, poorly resilient to packet loss, and imprecise. Furthermore, such communications often occur over a transport medium, such as a public switched telephone network (PSTN), that has a non-negligible charge per unit of time.

Thus, it would be best if AI agents could identify each other as such and optimize their method for completing the transaction accordingly. However, as the key to these AI agents is the seamlessness with which they can emulate a human, agents should recognize each other in a manner that not detract from that apparent humanity. Thus, present embodiments enable an AI agent to identifying itself as an AI agent in a manner that does not measurably impact the experience human agents have interacting with them. Once identified, AI agents proceed to optimize the method by which the transaction is conducted.

An AI agent may identify itself as such in a discreet manner than minimally disrupts the transaction when the far end is a human, while allowing for any remote AI agent to estimate with high confidence that the speaker is another AI. In some cases, the AI agent might choose to explicitly identify itself as such. Some vendors might make this an optional or mandatory setting to avoid situations where human listeners might feel fooled. Furthermore, in some jurisdictions, AI agents may be required by law to identify themselves. In some situations, this self-identification might be asymmetric. For example, a caller might be required to identify themselves in this fashion while the callee might not. Some other explicit, but non-verbal, cues might indicate more subtly that a speaker is an AI agent. For example, a short pattern of initial tones at the start of a call can identify the caller or callee as an AI agent, thus avoiding the longer, more disruptive need to identify an AI agent in speech.

However, in some use-cases, an AI agent may not identify itself. In these cases, the AI agent may nevertheless be identifiable as such to other AI agents while still appearing human to human listeners by embedding subtle cues in their communications. Figure 1 is a flowchart depicting a method for identifying AI participants in accordance with present embodiments.

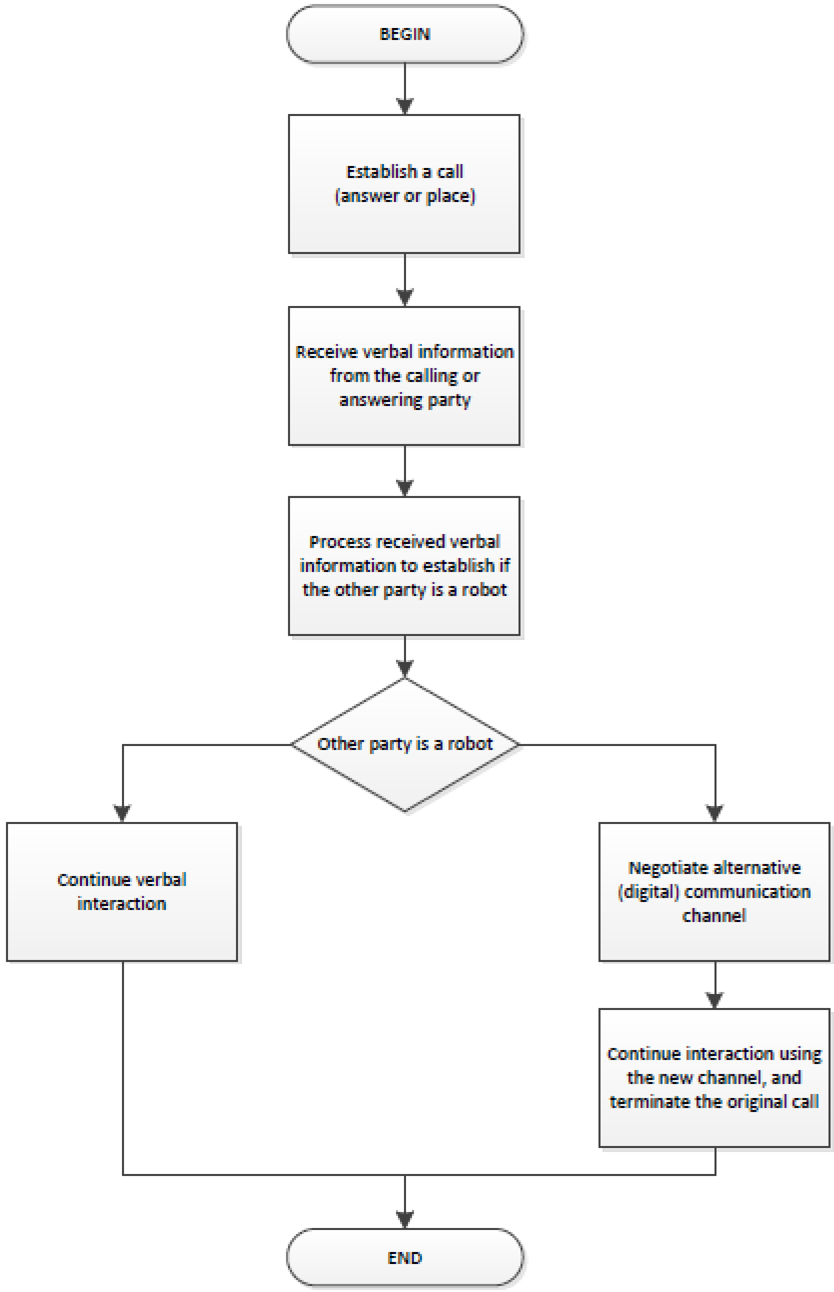

Figure 1

The AI agent can calibrate the length of periods of silence between spoken words to convey additional information into a conversation. The length may be varied in order to encode meaningful identification data in a manner that is impossible for humans to detect due to its exact precision. An AI agent would only need to send a very short block of information in this manner. For example, an AI agent may pass the digits 1, 2, 3, and 4 by adjusting deviations from a baseline silence period by 10, 20, 30, or 40 milliseconds. A particular pattern of periods of silence may be predefined according to a standard that enables another AI agent to identify the speaker as an AI agent. Figure 2 depicts an example flow diagram in accordance with present embodiments.

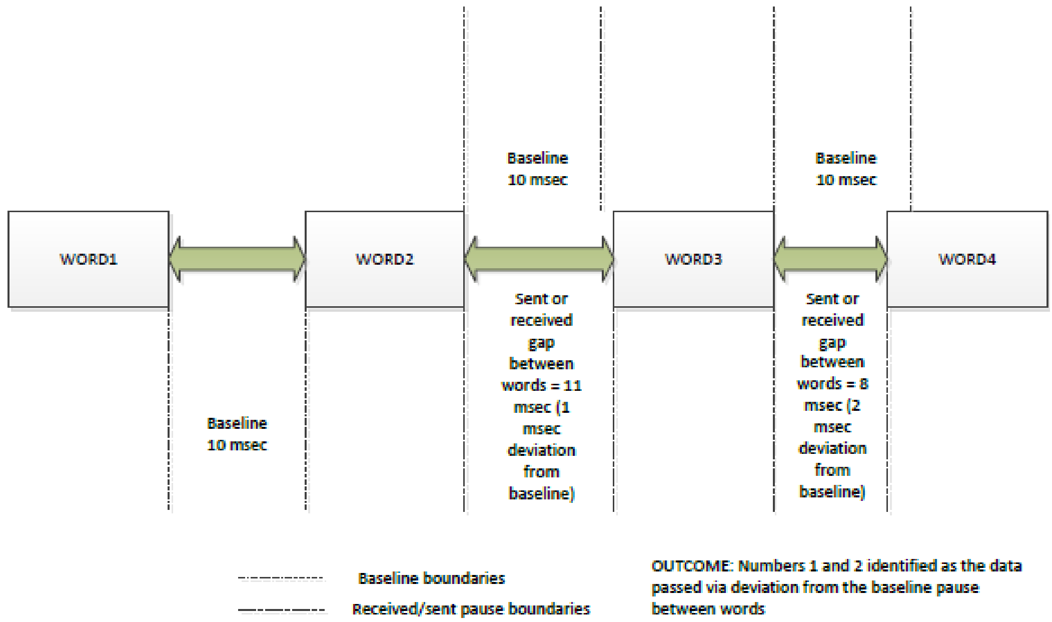

Figure 2

While a longer sequence can increase the confidence of identification, an arbitrarily-long series could be defined in which a recipient AI agent can respond when it has detected the pattern with a sufficiently high confidence. Once an AI agent has exceeded a predetermined confidence threshold, the first AI agent can exercise one or more forms of optimized communication.

Alternatively or additionally, an AI agent may identify itself according to word choice or other steganographic techniques. The AI agent could choose particular patterns of phrasing to identify itself as an AI. There are many conventional algorithms that could be employed, such as predefined words in a certain order, or summing the values of the first letter in each word in a phrase. The greater the complexity of these algorithms, the shorter a number of words or phrases they might require to be effective, but the more complex it will be to train an AI to use while not apparently distorting the "human-ness' of communications. An algorithm may be specific to a particular language, requiring a different approach for each supported language.

As words often have a similar meaning, the choice of one word over another can be represented with a numerical value. The choice of words can then be used to pass numerical values to the AI recipient, who will be able to identify a pattern by selection of words by the sender AI. For example, if "sure" is given a value of 0, and "of course" is given a value of "1", and "will do" is similarly given a value of "0", while "right away" is given a value of "1", then choice of specific words for a sentence "of course, will do", and subsequent words in a sentence in a similar manner can give an identifiable pattern to the recipient side in order to accept the sender as an AI.

Once the confidence level that another participant is also an AI has exceeded a predetermined threshold, the first AI can exercise one or more forms of optimized communication. These optimized methods of communication can be unilateral (that work without the other participant supporting the same mechanism) or bilateral (which require the other participant to also support the same mechanism). One example of a unilateral optimization is when an AI agent alters its speech patterns from human-optimized patterns to patterns designed to communicate with an AI agent. One change under these conditions would be to remove filler words and other speech disfluencies, such as “um” and “ah,” remove polite words such as “please” and “thank you,” and/or remove any other such extraneous phrases that are designed to generate increased comfort in a human listener but do not convey or elicit information. Other changes might include increasing the speed of communication (to reduce the length of the transaction), switching to an accent optimized for machine processing, and increasing the range of volumes used.

These unilateral steps enable a transaction to be completed more rapidly and accurately without requiring other participants to have any similar recognition or optimization processes. However, if the other participants do support a similar mechanism, a shift to a more machine-oriented pattern of speech can trigger a rapid rise in the confidence that the current speaker is an AI in the remote participant (thus triggering the AI participant’s own escalation mechanisms). Bilateral optimized methods of information exchange offer a far greater increase in speed and accuracy, but require a compatible method of information exchange and a negotiation mechanism to establish such support.

One example of such a method of exchange is the V.90 audio modem standard. Alternatively, AI agents may switch to a separate transport entirely, such as exchanging a URL as part of the negotiation and then using REST over IP to complete the transaction. Negotiating these alternative methods of information exchange can be implicit or explicit. If implicit, the AI agent may establish the exchange using a preset pattern of quiet beeps, might also define a particular method of escalation that can then be presupposed when the particular identification mechanism is identified. In these cases, the AI agent may terminate any speech and immediately move to the escalation method. If not implicit, however, then an explicit negotiation method is required. This could take place using the speech channel (e.g., by verbally asking "Do you support standard ‘X’?"), or the speech channel may be paused briefly to send an audio handshake and see if a corresponding response is received. In some embodiments, an explicit negotiation may take place alongside the speech channel by underlaying it with a discrete audio handshake that will minimally impact a recipient’s accurate reception of the speech.

Once a machine communication mode for completing the transaction has been identified, speech will normally terminate; however, in cases where a separate channel (such as HTTP over IP) is used, the speech could continue as a backup in cases of an inability to establish a connection via the machine communication mode (at least until bidirectional communication over the alternate channel has been confirmed). Subtle methods for identifying an AI on the other end of a phone conversation can also be used to identify and/or filter out any human attempting to pose as an AI agent.