Selected Papers of #AoIR2019:
The 20th Annual Conference of the
Association of Internet Researchers
Brisbane, Australia / 2-5 October 2019

TRUSTING SMART SPEAKERS: A TYPOLOGY OF INVOCATIONARY ACTS

Chris Chesher
The University of Sydney

Smart speakers such as the Google Home and the Amazon Echo have recently become popular internet-connected consumer devices with the seemingly magical capacity to hear user 'invocations' and provide intelligent responses in natural language. This paper analyses and categorises popular smart speaker commands and extends Austin (1964) and Searle's (1976) analysis and classifications of speech acts to develop a typology of what I call *invocationary acts*. I argue that a query or command to a smart speaker is a modern translation of the ancient ritual invocation. Invocation is form of supplication to a sublime non-human other, following protocols, seeking guidance or support at a moment of crisis. But today's invocations are not to a deity, but to servers and databases in the cloud, and most crises are trivial. Contemporary invocation is a convenient form of power for users, but also a supplication to the asymmetrical power of corporations.

In order to better understand invocation, I will turn to Austin's (1964) speech act theory which provides the basis for analysing **invocationary acts**. Austin explained the dynamics of speech acts by distinguishing between three elements: the locutionary, the illocutionary, and the perlocutionary. The act of speaking a grammatically sensible speech act, such as 'What is the capital of Iraq?' is the **locutionary act**. The **illocutionary act** is what is performed *in saying* this question. In this case, it is the intersubjective force that obliges the person addressed to respond. If the listener responds, this is the **perlocutionary act** that occurs as a consequence, even if this is 'I don't know'.

When a user makes a request to a smart speaker, they are performing an everyday speech act, but they are also initiating an invocationary act that initiates a technical procedure. The user starts the invocationary act with a proprietary 'wake word' such as 'Hey Google'. In linguistic terms, this is a phatic — an interpersonal communication that

Suggested Citation (APA): Chesher, C. (2019, October 2-5). *Trusting voice assistants: a typology of invocationary acts*. Paper presented at AoIR 2019: The 20th Annual Conference of the Association of Internet Researchers. Brisbane, Australia: AoIR. Retrieved from http://spir.aoir.org.

in this case identifies the person (or thing) addressed (Meltzer & Musolf 2000). At a social level, this obliges the 'assistant' persona, as a quasi-social actor, to provide an appropriate response. At the technical level, the sounds 'Hey Google' invoke the device to record the user's invocation and pass the recorded utterance to the cloud for interpretation.

The invocationary act continues when the computing infrastructure uses machine learning models to (1) interpret the locutionary act using speech-to-text conversion and (2) interpret the illocutionary act using artificial intelligence statistical models, and (3) use machine learning algorithms to find Baghdad from a database as the most probable appropriate response. The assistant then responds in a synthesised voice to perform the perlocutionary act, fulfilling the social obligation.

So, what can we say about speech acts have taken place? We can use Searle's (1976) classification of five kinds of human speech acts:

- **directives** attempt to influence another actor's future actions — such as a question;
- **representatives** represent something as true — such as an answer;
- **commissives** make a commitment to take a future action, such as making a promise;
- **expressives** communicate a psychological state; and
- **declarations** do something in the act of saying it, such agreeing to a marriage proposition, or when a judge passes sentence on a convicted criminal).

In the example above, my invocation is a question — a kind of directive speech act. The assistant's response is an answer — a representative speech act that identifies Baghdad as Iraq's capital. But the interaction has been mediated as an invocationary act.

To develop a typology of speech acts in common smart speaker invocations I found 300 recommended Google Home commands from *CNet*, *Lifewire*, *Android Authority*, *Tech Ranker*, *Tom's Guide* and *Lifehacker*. I also drew from the record of hundreds of invocations made by me and my family. I tested each of these invocations and investigated their services.

The users' invocationary acts are almost all directive speech acts: questions (n.98) or commands (n.166). Non-directive acts were much rarer (n.10), and usually invoked scripted responses. For example, when I said, 'I am your father' (representative) the assistant made a *Star Wars* reference by answering, 'I'm sorry I'm not Luke' (representative)… 'This is kind of awkward' (expressive). When I said, 'It's my birthday' (representative) it gave the expressive response, localised for Australia, 'G'Day and happy birthday. I hope you have a cracker'. When I performed an expressive act by saying 'That's disgusting', the assistant responded with another expressive — 'I didn't mean to gross you out, sorry'. I even performed a declaration by saying 'I am Chris', which prompted the assistant to say 'You'd like me to call you Chris. Is that right? I'll call you Chris from now on (a commissive). I then asked, 'Who am I' and the assistant used speaker recognition to identify me uniquely as Chris (representative). If the assistant

could not recognise the invocation, it responded with an error such as ‘I’m sorry I don’t understand’ — indicating that my speech act has failed, and expressing the psychological state of incomprehension (expressive).

Where user invocations were almost always directives, the smart speaker speech acts were mostly representatives providing facts. However, responses can take many other forms (See Table A). They can make promises, ask the user to do things, make declarations and express emotions.

To speaker’s invocation:	**Smart speaker speech act**	**Speech act type**
What is the capital of Tanzania?	‘Dodoma is the capital of Tanzania’	Representative
Set a timer for five minutes	‘Alright. Five Minutes. And that’s starting now. ‘	Commissive
Play the trivia game	‘…Welcome to “Are you feeling lucky”… I’m the host of this silly show… How many are playing this time?’	Directive
(At the beginning of the quiz and at the end of the quiz)	‘Player one. I’ll call you “dingo”’… ‘And now for your score. Not bad at all. You got four right…’	Declaration
Do you love me?	‘Love. I knew the way I felt about you had a name.’	Expressive

Table A

With some reverse engineering I identified the operations in play with a number of invocationary acts (see Table B). Many invocationary acts **search** internal or external databases, or **lookup** data from a service, such as the weather forecast. Others perform mathematical **calculations**. Many **play streaming media** such as music or radio. Some invoke **scripted responses** or generate **random responses**. Some create more complex interactions such as **tutorials** or **games**, Some commands **control devices** like smart lights or thermostats.

Invocation (User’s Locutionary act)	**Evocation (machine’s locutionary act)**	**Invocationary act**	**Classification**
What is the capital of Iraq	Baghdad is the capital of France (representative)	Searches Google database for answers	Search

What is the weather?	Currently in Newtown it is 24 and cloudy… (representative; commissive)	Looks up information from an established authority	Lookup
Play 'Lust for life' by Iggy Pop	Sure. Lust for life by Iggy Pop. Playing on Spotify (Commissive) [Plays song]	Looks up song in media database and starts streaming	Media
What is spelunking?	According to Wikipedia caving, also known as spelunking in the United States… (representative)	Looks up brief Wikipedia entry	Third party search
I'm talking nonsense	My apologies. I don't understand	Error message	Error
Roll a dice	(sound) It's a five (declarative)	Chooses a random number	Random
Are you Skynet?	No way. I like people. Skynet hates people. I rest my case. (expressive)	Responds with a response scripted for a defined invocation	Scripted response (often randomly selected from multiple answers)
How do you make devilled eggs?	OK I've got a recipe from Food network… (representative) Would you like to hear the ingredients or skip to the instructions? (directive)	Accesses recipe information and steps through ingredients and method. Users must invoke each step.	Interaction (tutorial)
Turn on the light	Directive [light turns on]	Turns on smart home lights	Device
Set a timer for 10 minutes	Got it. Ten minutes, starting now. (commisive)	Sets timer	Clock

Table A.

It is from variations on this repertoire of invocationary acts that users like me are able to get an impression of intelligence or even companionship (Andreallo & Chesher 2019). Exchanges of invocationary speech acts mimic the dynamics of conversation, operating within certain acceptable thresholds of space (what is audible and apparently present) and time (the average 200 milliseconds gap between conversational turns) (Enfield 2017)). With Continued Conversation (Gebhart 2018) it becomes possible for users to respond within 8 seconds without the wake word. The interactivity is particularly interesting when using voice assistants in a social context with the experience of mixed human and non-human conversation partners. In mediating invocationary acts, voice assistants have become a distinctive media form whose implications are only becoming

apparent. But unlike everyday conversations, invocations are supplications to corporations with **monopolies of invocation**.

References

Andreallo, F. & Chesher, C. (2019) Prosthetic Soul Mates: Sex Robots as Media for Companionship *M/C* Vol 22, No 5.

Austin, J. L. (1975). *How to do things with words* (2d ed.). Oxford: Clarendon Press.

Gebhart G. (2018) 'Google Home's new continued conversation setting keeps the mic hot for a smoother chat' *C-net*. June 21, 2018. Available at: https://www.cnet.com/news/google-home-continued-conversation-setting-keeps-the-mic-hot/

Heidegger, M. (1977). *The question concerning technology, and other essays*. New York: Garland Pub.

Meltzer, B.N. and Musolf, G.R. (2000) '"Have a nice day": Phatic Communion and Everyday Life. *Studies in Symbolic Interaction*, Volume 23, pages 95-111.

Phan, T (2017). 'The materiality of the digital and the gendered voice of Siri'. *Transformations* issue 29. Available at: http://www.transformationsjournal.org (accessed 1 October 2019).

Searle, John R. (1976). "A Classification of Illocutionary Acts." *Language in Society* 5, no. 01 (April 1976).

Zuboff S (2018) *Surveillance capitalism*. London: Profile Books.