THE A-Z OF SPEECH
ANALYSIS AND ITS ROLE
IN UNDERSTANDING
CONSUMER BEHAVIOR

Sandeep Arora
Executive Vice President & Head Consulting & Advisory Practice

So speech is a very important element. If we have to understand consumer behaviour, of course there are other variables through which you're going to understand consumer behaviour too. But speech offers a complete package and therefore you cannot ignore it if you really want to understand consumer behaviour.

The reasons for that are the fact that we as Homo sapiens, from the time we have evolved, we have been communicating with ourselves largely through speech.

We would actually express ourselves, express our emotions, express or talk about our needs and wants with each other. Why a speech? Because speech is not a uni dimensional information. It actually carries a whole package with itself. The way I speak, the content

part, the text part, along with the expressions, the emotions, the silences, the pitch, the speed at which I talk, the entire package contains a lot of information as to how I'm feeling right now. And thus, if I was to ignore that, I would actually lose out on a lot of information about a particular person or a particular consumer.

And probably that's why when we want to interact or hear about a consumer directly, we always refer to it as a voice of a customer, not as a text of a customer or SMS of a customer and et cetera.

All of this explains as to why speech, therefore, is not an element that you can ignore when you are wanting to understand consumer behaviour. So the very obvious question that any one would have is why we haven't been using speech all this while if it

were so important to analyze the consumer behaviour. Well, one of the key challenges that one has with speech as a form of data is the fact that the moment it is spoken, it no longer is there.

It's only in late 19th century that we develop the technology to be able to store on voices and our speech. And once we have managed to store it, is really when we can analyze it. So therefore, the entire phenomenon of storage has been very recent. Now, after we have stored it, it doesn't still mean that we can analyze it, you know. Storage is a necessary condition, but not a sufficient one for us to be able to make meaningful sense of what that data is saying. We have to have a very good understanding of different components, which going to the multi dimension information that comes via speech. So as humans, we are programmed naturally to sort

of decode it. And therefore, all this while whatever speech analysis we have been doing head has largely been manual, which is quite accurate and more sense of the ways. But the restriction there is the fact that we can only analyze so much. We cannot do a massive scale based processing manually. The technology is still evolving for us to be able to teach our machines in terms of how to deconstruct the entire speech mechanisms, in terms of the context, in terms of the content, emotions, cultural veracity, etc. for us to be able to make some intuitive sense off of what a person is trying to express sort of very broad level if we want to analyze the speech it has to follow two broad steps.

First is being able to understand the speech. Second is being able to interpret the speech and both have to actually run in tandem for us to be able to get good results. So for the purpose of technology, the two have been separated where the challenge of understanding is handled through speech to text technology. And the second challenge of decoding or interpreting what the content is is handled through text analytics, which deploys technologies like NLP via NER and things like that.

So the two parts have to sort of run in tandem for us to be able to interpret and analyze the whole speech in a meaningful way at a massive scale level. So while we have,

as a fraternity, moved up significantly ahead in terms of being able to understand, interpret through a machine of what people are trying to say. There still remain a lot of challenges. Let me highlight a few. So if we look at the understand pillar, the challenges that one faces is that, of course, the platform that you have working on needs to understand multiple languages. It needs to understand multiple accents. It needs to understand multiple people speaking together, for example. It also is a function of how good the quality of recording is. It's a function of us using multiple languages simultaneously. It can also pertain to the fact that I may be speaking something which is extremely evolved, extremely domain orientated or scientific and stuff like that, which a normal speech to text engine may not be able to decipher.

Okay, so now let's come to the second layer, which is the interpretation layer and let's understand the kind of task that it requires. And the challenges thereof. Interpretation really is about understanding what people are saying and what they're meaning. Essentially that is interpretation. Now, it involves a couple of things for us to be able to basically bridge this gap. But these the gaps are so wide sometimes. And as humans, we know that even we ourselves have not managed to master this particular art and therefore, to teach all of these algorithms to a machine to sort of

decipher, is an equally challenging task. What sort of goes underneath the interpretation layer are these two or three components. The first and foremost one is the content. The content is the choice of words, the figures of speech, the complexity of topic and various other things which sort of put together the core, the content in terms of what is it that I'm saying, then what exactly could possibly be meaning the results out of context that surrounds the whole content. And that comes from cultural sensitivities. It can come from the tonality of my voice. It can come from the pronouns that I have used in the entire conversation. It could be a function of which level are we really having the conversation or the discussion at, and the ability of the analyzer to be at that particular level to sort of understand the real meaning. It also has to do with a lot of mannerisms of delivering the speech to which can add to a whole lot of complexity. So within all of these things is where we are still struggling. Of course, there is a light at the end of the tunnel that we see now and we have as an industry progressed significantly towards that.

But these are some of the elements which are in the way that make the entire journey lot more interesting and a lot more challenging for us to be able to use the whole speech as an integral data to analyze the consumer behaviour.