

デジタル市場競争会議ワーキンググループ（第26回）

1. 開催日時：令和3年10月1日（金）15：30～16：30

2. 開催形式：通信システムを用いた遠隔開催

3. 出席者：

依田 高典 京都大学大学院 経済学研究科 教授

上野山勝也 ㈱PKSHA Technology 代表取締役

川濱 昇 京都大学大学院 法学研究科 教授

伊永 大輔 東京都立大学大学院 法学政治学研究科 教授

塩野 誠 ㈱経営共創基盤 共同経営者/マネージングディレクター

増島 雅和 森・濱田松本法律事務所 パートナー弁護士

山田 香織 フレッシュフィールドズブルックハウスデリンガー法律事務所 パート
ナー弁護士

（デジタル市場競争会議 構成員）

北野 宏明 株式会社ソニーコンピュータサイエンス研究所 代表取締役社長

4. 議事

（1）新たな顧客接点における競争環境について、業界関係者からヒアリング

5. 配布資料

資料 野村総合研究所桑津理事提出資料

○依田座長

ただいまから「デジタル市場競争会議ワーキンググループ」を開催いたします。

本日は、生貝議員、川本議員、森川議員は御欠席です。

競争会議から、北野議員にも御参加いただきます。

なお、伊永議員、山田議員は、途中退席になると伺っております。

それでは、本日の議事や資料の取扱いについて、事務局から説明をお願いいたします。

○成田審議官

ありがとうございます。成田でございます。

本日の議題は1点でございます。新たな顧客接点における競争環境につきまして、本日は野村総合研究所の桑津理事にお越しいただきまして、関係者からのヒアリングということで、ヒアリングをお願いしたいと思っております。

本日の資料でございますけれども、桑津様の資料は一部非公表の部分がございます。それから、議事録につきましては、いつもとおおり発言者の皆様の御確認を経た上で公表する予定でございます。

私からは以上でございます。

○依田座長

ありがとうございます。

それでは議題に入ります。

まず、野村総合研究所の桑津理事から、資料について25分ほどで御説明いただいた後、意見交換を行います。意見交換では、いつもとおおり上野山議員からワーキンググループ名簿順の後、北野議員の順番で、お一人当たり3分ほどで桑津理事への御質問やコメントを伺ってまいりたいと思っております。

それでは、桑津理事、どうぞ御説明をよろしくをお願いいたします。

○桑津様

よろしくをお願いいたします。

ただいま御紹介いただきました、株式会社野村総合研究所未来創発センターの桑津と申します。

それでは、ボイスアシスタントもしくはインテリジェントスマートスピーカーというのが、今はマーケットがどういう状況で、競争環境的にどのように認識されているかということについて簡単に御説明をさせていただきたいと思っております。

画面を共有させていただきます。共有できておりますでしょうか。

それでは、御説明をさせていただきます。

全体の流れを簡単に御説明させていただきますが、まずボイスアシスタント、スマートスピーカーが代表例なのですけれども、この仕組みについて。これはスマホと何が違うのかという言い方をしますと、キーボードがない、ディスプレイがない、無線がない

ということで、要は音声認識で、主たる処理機能の大半はネットワーク経由でクラウド側で行うということで、見た目は例えばスマートアシスタントボイススピーカーを想定しますと、いわゆるハードウェアが非常に目につくのですけれども、ハードウェアとしては非常に軽量でございます。最先端のプロセッサが必要なものではございませんで、あくまで聞き出した音声をネットワーク側に流して、その結果をスピーカーで返すという仕組みでございます。そういう面で、これは認識の精度と、後ろにあるプラットフォームの優秀さ、それを支えているクラウドという3点で取り上げるべきものです。

本日のテーマが、スマートスピーカーではなくてボイスアシスタントになっております。マーケットとしてはスマートスピーカーが8割方を占めております。ただ、実はスマートスピーカー自体はもう2020年ぐらいからスピーカー以外に展開しようよということで、特にスマホとの関係にある意味では見直す段階に来ております。

実は2～3年ぐらい前ですと車が大本命で、BMWなんかは実際に設置されたと聞いておりますけれども、現状、車は展開が遅れて止まっております。これはある意味、事業者自体がアマゾンに依存することを避けたという状況です。

スマホを持たない事業者さん、これは何の意味かと言いますと、アマゾンとかアリババが比較的強いのですが、両者に共通していますのはスマホのプラットフォームは実はそこまで強くない、クラウドは非常に強いということで、携帯を持たないという立場から、プラットフォーム展開をするに当たって、この音声クラウドプラットフォームという組合せがある意味非常に有効であると評価して参入してきたものと見ています。

繰り返しですが、クラウドと連携したことで、本来のコストが極めて引き下がりました。ブラックボックス化しています。ブラックボックス化といいますのは、例えばOSがどうかという議論は基本的にはなくて、後ろ側で一体どのような処理がされているのか、どこで処理がされているのか。もうちょっと非常に素朴な議論で言いますと、そもそもスマートスピーカーというのは人間が後ろで聞いているのではないかとかですね、結構こちら側から分からないブラックボックス化が非常に進んでおります。

同時に、データ収集の強みがございます。これはクラウドを介して端末のコストが低いです。かつ、家の中にあって、いわゆる携帯電話が比較的弱い分野でデータをたくさん取ることができるということでございます。

これも余談でございますけれども、後ろで誰かが聞いているのではないか、盗み聞きしているのではないかというのは昔から言われていたところでありまして、実際、例えばアマゾンとか大手のところに盗み聞きしていましたという議論は基本はないと理解しているのですけれども、それは決して証明されたわけではありませぬし、公式には、例えば私がスマートスピーカーに話しかけた音声は個人が聞くことはないとなっているわけですが、研究開発の部隊がサンプルとして評価したみたいな話は幾つか聞いたことがございまして、このブラックボックス化というのが、取られたデータがどうなっているかというのがよく分からないということでございます

このリスクが物すごく極端に出た例が、ちょっと変な例なのですけれども、アメリカでリカちゃん人形にボイスアシスタントが入っておりまして、女の子が、リカちゃんはどうしたのかと言うのです。お父さんとお母さんはお出かけしてしまったとかというのをずっと子供が話しかけるといふのがありまして、これがブラックボックス化でずっと盗聴されていたということで、一時大問題になりました。

小さい子供が言っていることなのだから別に大した問題ではないかという、それはそんなことは決してなくて、今、お父さんお母さんがいないとか、お父さんはこういう仕事をしているのか、こういった内容が普通に出ていってしまったことがございまして、これは非常に極端なケースですけれども、データ収集という議論はあると思います。ポイントは、ずっと置かれているということ。スイッチを入れなくてもデータは取っているということです。

ただ、このマーケットなのですが、実はデカップリングしております。というのは、もともとアメリカで生まれたビジネスなのですけれども、今は数的には中国のほうが圧倒的に多いです。中国においては別にスピーカーにこだわる必要ないということで、今、スピーカーではなくなりつつありまして、普通にディスプレイのついたホーム端末、多機能端末に多様化してきております。ボイスアシスタントが要らなくなったというわけではなくて、ボイスアシスタントもインプットの有力なメディアとしては残っているのですけれども、多機能化というのは、ディスプレイで表示するもしくはタッチパネルで操作するというような議論も出始めています。ですので、私の音声縛りだったボイスアシスタントから、今は多機能端末の中の有力メディアとしての音声認識もしくは音声オートという方向感に多様化してきております。

プレーヤーの動向なのですが、先ほどのデカップリングでして、いろいろな意味でデカップリングしています。というのは、まず米中デカップリングでして、アマゾンが先行して、グーグルが対抗するという図式が2年前まであったのですけれども、ここでアリババ社及びバイドゥが独自経済圏をつくって、自国に投入しています。これは言うまでもありませんけれども、比較的端末が安くて、例えばシャオミとかバイドゥなどはほぼただみたいな端末で配ったと言われていました。それで、独自の経済圏をおつくりになった。

後から入ってこられたアップルさんも、ある意味これは米国の中ですけれども、アマゾンやグーグルとは別のアプローチで独自の経済圏をつくっているということで、現状は覇権争いというよりは、御自身の領域を切り取るというような競争の構図になっております。

言うまでもありませんが、アマゾンのスピーカーとグーグルのスマートスピーカーは中国国内では販売禁止になりますので、基本的にはまさに市場を含めてデカップリングされております。

これらが競争上どのような懸念を持っているかということでございます。1つは量的

なものでございまして、直感的に気づかれるところだと思うのですが、これは家の中に2台置く必要が全くない製品でして、最初に置いてしまう。しかも、スマホに比べて、この機械は比較的lowコストなのです。スピーカーの場合、ディスプレイは要りません。ユーザーインターフェースの部分が音声以外ほとんど要らないのです。処理機能も要らない。つまり、高度な画像処理とかが必要ありませんので、後ろのネット側で処理しますので、非常にlowコストで設置する、もしくはビルトインさせることが可能になるということで、プリインストールやビルトインといったような初期コストを下げた展開が非常に重要であります。

一方で、データがたくさん取れる。これは公式には言われておりませんが、検索と同じでありまして、その人が何に興味を持っているか、あるいはちょっと誤解を恐れずに言えば、いつ出かけるのかとか、いつ帰って来たかとか、そういう情報が取れてしまうわけです。帰ってきてから明かりをつけてくれとか、エアコンのスイッチを入れてくれとかという話が出るわけですし、生活スタイル等が非常に分かります。

一般にスマホはいわゆる朝の8時から夜の8時までと言われているわけですが、このスマートスピーカー系のボイスアシスタントのポイントは、どちらかと言いますと24時間、それもスマホがカバーし切れていない夜の8時から朝の8時まで、寝ている間はあまり関係ないので、その部分のデータも取れてしまうということで、ある意味先行者優位でデータがたくさん取れる構造になっています。

これもスマートフォン、スピーカー同様でございまして、アプリ認証みたいな議論があります。というのは、グーグルもアマゾンもいずれもたくさんのアプリをショップもしくは、厳密にはこの場合はショップではないのですが、サードパーティがアプリを入れることができるよという形になっています。当然ですが、アプリ側としては自分のアプリを採用してほしいので、一生懸命作られる。一方で、この認証等についての商慣習という議論が若干課題になっています。これは明示的に問題になっているということではなくて、そもそもそのアプリはよくてこのアプリは駄目だとか、どういう理由でそうなっているのかとか、非常に中の審査基準等が若干分かりにくい。これが一般的なゲームとか金融決済の場合ですと比較的分かりやすいルールで、例えばこのレスポンスでこれが出せないのだから駄目だみたいなことが言えるのですが、このスマートスピーカー系のアプリに関して言いますと、誤解を恐れずに言うと、処理機能の大半はクラウド側にあるわけですし、ブラックボックスであり、かつ見えないので、言葉が悪いのですが評価、判断基準がいまいち見えにくいなど。もちろん各社ガイドラインを出して、例えば不適切なことは駄目だとか、関係ないときにデータを取ってはいけないとか、いろいろなルールで運用されていますので、認証制度がないわけではありません。認証制度があるのですが、非常に分かりにくい。実態として、なぜこちらが先になったのかなとか、そういったものが見えにくい構造になっているというのが実態だと思います。この辺り、随時細かく御説明をさせていただきます。

まずマーケットの動向なのですが、爆発的に伸びているように見えるのですが、実はできたのが2017年ぐらいで、1回伸びがとどまっています。2019年から中国の市場が急成長して、第2期ブームです。2020年以降、実は中国以外の市場が伸び悩んでいます。中国マーケットは、数はべらぼうに伸びているのですが、何せ単価が低い。普通1,000円、3,000円というレベル感で展開していますので、数は出るのだけれども、市場の大きさそのものはそこまで大きくないということです。

シェアですが、中国が3社、バイドゥ、アリババ、シャオミが大手3社なのですが、ここが35%から始まって今は45%ぐらい、ほぼ半分中国という状況になっています。ただ、中国のマーケットを確認してみると、これもやはり極めて大都市製品らしい。郊外僻地といいますか、山村農村で売れているのかということそうではないのだということで、これはスマホがあまり入っていないところにも入っているという性格のものではなくて、スマホをたくさん持っている、もしくは先端的にパソコンやIT化が進んでいる方々が品ぞろえの一環として持っているという性格のものだと位置づけられています。

プレーヤーの変遷ですが、アマゾンが先行しました。当然です。グーグルが猛追しました。これを中国のバイドゥ、シャオミ、アリババが三国時代で追いかけて、アップルは独自路線ですということで、2019年より前と後で状況が変わっておりまして、2019年まではアマゾンが先行。普及品が低価格化。アマゾンも製品ライン、当初の1万5000円から、今はスマートのスピーカー機能をかなりなくしているもので3,000円ぐらいまで行っています。一方で、高付加価値製品ということで、アプリに代用されるような2万円、3万円のものも出てきております。ただ、当然アマゾンが価格競争力が強いのです。グーグルがこれを猛追していて、品ぞろえは限定的です。アマゾンは低価格化対応したのですが、実態のところ評価として、別に劣っているわけではないのですが、いまいち存在感が高まっております。

一方で、全然別のゲームをやっているというのがアップルでございまして、非常に大きいものを作って、高級化しました、音質も非常にいいですと言っているのですが、あくまで自社ファミリー製品圏内でのステイといいますか現状維持にとどまっております。誤解を恐れずに言いますと、Siriの機能、音声認識ですが、これはどちらかというスマホ側で使えよという趣旨で使われておりますので、Siriだけが乗ったスピーカーというのはアップル製品として若干相性が悪い。結果、高品質のディスプレイをつけているというのが実態なのかなと思います。

2020年以降、Amazon Alexaはソフト化とディスプレイ対応になりました。どういうことかということ、いわゆる今までスピーカーだったのが、スピーカーではなくてソフトウェアだと。ボーズとかBMWとか幾つか有力企業が採用しましたが、ただメジャーになったか、マジョリティーを取ったかということ、そういう状況ではありません。ですので、ソフトになった。同時に、ディスプレイがつきまして、スピーカー機能というのはワ

ン・オブ・ゼムだという方向に変わります。

同時に、中国ではBSAの3社が販促金をつけて販売競争を激化、低単価化が進みました。この時期に赤外線リモコンやディスプレイの動画配信といった家電のほかの分野とのすみ分けが曖昧になっています。つまり中国のバイドゥ、シャオミ、アリババ製品はディスプレイがついていて、動画をこれで見てくださいというような性格のもので、別に音声認識が本当に要るのかということとちょっと怪しいなど。スマホの親戚だというような位置づけが非常になされ、その分、低価格で、例えばアリババなどは家電量販店だけではなくて、普通に文房具屋さんにも置いた。販促としてどんどんただで配ったというようなものがありまして、低単価化が非常に進んだというのが実態でございます。

音声応答スピーカーの製品動向ですけれども、2018年まではスピーカー。低コスト化でUSD50ドルから30ドルに下がって、今、無料配付も一部出ている。19年以降はソフトウェアエンジンのディスプレイ対応が進んで、アップルとボーズは高級スピーカーになっている。2020年は中国でスピーカーからホームのミニ端末、持ち運ばないテレビに移った。音声認識をアピールするのはむしろバイドゥなんかで例外的です。というのは、中国のBSA3社は音声認識に技術者が結構いると言われておりまして、一番認識能力が高いと言われていたのはバイドゥです。私も中国語を上手にしゃべれるわけではないのですけれども、バイドゥの認識は、例えば日本人が中国語でしゃべっても比較的認識してくれるということで、バイドゥの音声認識の技術力は確かに高いというのは割と業界のコンセンサスになっていると思います。

ただ、実態はディスプレイ上で操作したり、動画を視聴したりというような形で、スピーカーでなくてはいけないという性格のものではなくて、ホーム、ミニ端末になっております。

各社の製品の規格等を細かく書いておりますが、今日は商品の細かい説明ではないのですが、何を言わんとしているかと言いますと、左側はどちらかというと初期のアマゾン、初期のグーグルと、低コストのアマゾン、グーグルという理解でございまして、音声認識機能は実はほとんど変わっていないと言われております。変わったのは何かということ、若干スマートホームに寄せましたよねとか、そういったところと、あと、コストが下がったかなという辺りがポイントです。

2019年以降の中国の製品は、音声系もそこそこ、例えばなのですけれども、よくしゃべります。グーグルやアマゾンは比較的必要なことをしゃべったら黙っているのですけれども、中国の製品はやたらしゃべるといふのがありまして、国民性かなと思います。

スマートフォンもそうですし、動画系サービス、つまり、向こうで言うとビリビリとか、そういったものをどんどん見るような状況になっています。価格はどんどん下がっています。音声認識は比較的上がってきたと言われておりまして、馬鹿にならないということで、画面表示も含めて、比較的多機能化が進んでいるということになります。

事業戦略の違いなのですが、アマゾンは音声認識のパイオニアだと。OSのプラットフォーム

ホーム型戦略で、あえて画像を後送りにしました。意図的に最初は画像を乗っけなかった。ただ、商圈が拡大しませんので、画像を乗っけざるを得なくなったという性格のもので。アップルは、自社製品群を住宅向け、バイドゥ、シャオミ、アリババは中国国内でのアマゾン的な位置づけから多機能端末へ移ったのだなということです。

アマゾン、音声入力というのは新しい切り口だとして、携帯やPCとは違うマーケットを開拓するということでOSプラットフォーム戦略を志向された。ただ、商品の位置づけが曖昧になる。画面对応製品を出してしまうと、ほかのパソコンの親戚かということで区別がつかなくなるので、出し遅れた。一方で、家電の連携とかIoTセンサ対応を非常に幅広くやったのですが、現状、まだ不統一感が非常に強い。御自身で家庭に導入されている方もそうですが、例えば照明です。例えばリモコンとかは私も使っているのですけれども、やはり若干心配です。というのは、時々リモコンのネットワーク側のほうが止まっていて使えない。あと、目覚まし、私は昔、アマゾンのほうをずっと使っていたのですけれども、最近は本当の目指し機能だけにしました。というのは、ニュースとかを音声で出すやつは、ネットワークが止まっていたりすると音声が出なくて、目が覚めないとかということがあるのです。明かりもそうです。私は照明とニュースとアラームで朝、起きているのですけれども、やはり1か月に1回ぐらい調子が悪いと、ニュースがないみたいなことがあります。明かりはつきます。アラームは鳴ります。

これは笑い話ですけれども、アマゾンのAlexaのアラームというのはネットワーク連動ではないです。つまり、ネットワークをたたき切ってもアラームはつきますので、それであれば普通の時計と同じではないかということで、結局元へ戻っています。

何を言わんとするかというと、小慣れていません。やはりセットアップが大変で、よく言われるのですけれども、これは恐らく今ある製品を無理やりつなぐのは無理なのだろうなと。住宅を買い替えたときに、家電を1社のメーカーにそろえて、それをつなげるという形で使わないと、家庭の製品と同じ扱いにしないと、後から買ってきていろいろつないだり何なりというのは、結局いろいろところでトラブルが起きているし、完成度は低いなということで、勢いは出ないと思っています。

自動車の展開も、BMWとかを入れたわけですが、正直、すごく伸びたかということ、そういう状況ではないと思っています。

アップルは、音声認識はスマホと同じで、アップルブランドの住宅向け端末だということで、価格帯も安いものは避けて、音声オートよりも高級スピーカーということで、これは別路線。

中国のバイドゥなのですが、位置づ的にアマゾンと類似の戦略でありまして、販路はアリババに劣る。アリババが圧倒的に先行しましたが、一部店舗販売のために補助金を使ったり、低価格品が0円になったりということです。音声オートの精度は比較的高く評価されていますが、各社ともに、結局のところ家でそんなに明かりをつけたり質問したりすることばかり、ニュースを教えてくれというのもいいのですけれども、映像配

信とかゲーム端末で言ったらマルチメディア端末としてレベニューストリームが出せるところへ重視するというので、スマートスピーカーではなく、ボイスアシスタントのほうに移ってきたというのが実態かと思います。

各社のエコシステムなのですけれども、特にアマゾンが先行しまして、先ほど申しましたようにハードウェアの値段が非常に安いので、利益が出る製品ではありません。ですから、サービスをつなぎ込んでいろいろもうけていこうと考えるわけです。当然ですが、全部自分でできませんので、他社デバイスでデジタル経済圏を構成したわけですが、他社へのコントロールがうまくいっていません。これは2つの意味でうまくいっていません。一つは、技術的に各社がアマゾンに提供する機能を限定している。例えばリモコン系の機能、IoTなどで言いますと、アマゾン経由だと機能の7割ぐらいしか使えなくて、自分のところでやると10割使えるとか、完全に連携し切れないということで、小慣れないという議論。

一方で、アマゾン側が自社以外の取引の制限、例えば、よく言われているのですが、アマゾンのお店でグーグルのGoogle Nestは売っていないのです。つまり、そもそも売らない。この辺り、それはやっていいのかという状況が出ていまして、取引制限されている。グーグルは、他社連携以前にデバイスの提供が遅れ気味で、家電はちょっと頑張っていると言いましたが、先ほど言いましたように、家電量販店では割と隅っこに追いやられています。

アップルは完全に独自化していますので、他社とは同じ土俵には立たないというモデルになります。

バイドゥは一般販路に販売支援金がついて、コンテンツサービス事業者にも販路負担支援を行うというように積極的に展開していたわけです。

アリババは販路を支援して、企業販路向けの販売等に拡大して、バイドゥ同様、コンテンツサービス事業者との連携に重点を置いています。

シャオミはどちらかというところ、中国の中ではややアップル寄り、ブランドを形成して、中グレード以上でコンテンツ寄りだということ、いわゆる従来型の多機能端末でブランドの高いものを目指しますよという位置づけになっております。

最後です。競争環境上の懸念なのですけれども、表面上、競争をすごく妨害されるという話はないのです。アマゾン、グーグルから、これはやめておけとか、**特定のアプリに対して、プラットフォーマーが同種のアプリを後から出してひっくり返しましたという**ことは、少なくとも表面上は見られていません。理由の一つは、音声応答スピーカーが普及遅れだということ、市場が停滞しているのではないかということでもあります。

アマゾンのスキルやアレクサ連携を阻害される、もしくは多額の費用負担、つまりこれに乗っけて一番最初に選ばせてやるから、費用をたくさん持てよという状況ではなくて、どちらかというところアマゾン側も、引っ張ってくれるのであれば、いいアプリのためには特別な席は御用意しますよという感じになっています。

サードパーティは、両者の関係を維持しています。どちらか片一方だけしかお付き合いできないという構造は基本ありません。ですので、新しいムーブメントを期待しているということでもあります。

重点分野と目されていた自動車は、BWM以降、動きが乏しくて、あくまでサードパーティかなと思います。

ただ、ここには書いていないのですけれども、さはさりながら、ここで言うスキルもしくはアプリを新しく搭載してほしいというときに、どのような審査基準でやっているのかなということが意味分りにくい構造になっています。もちろんガイドラインがあって、ツールセットがあって、そこを通せばいいよという話になっているわけですが、何でも駄目だったのかというのはよく分かりません。かつ、なぜいいか、もしくはこれを使ってもらうための独自の取組を行いたいみたいなどは非常にしにくい構造になっています。もうちょっと単刀直入に言うと、例えば日本の住宅メーカーさんが自分のほうでアマゾンのアレクサのソフトのエンジンとクラウドの部分だけもらって、うちで自分で作りたいからというのは、アマゾン側はなかなかいい顔はしない。一応、アプリとして使わせてあげてもいいよみたいな話になっているわけですが、ではOEMとして全部任せるといふ議論にはあまりなっておりませんし、特に商品自体もよく分りにくいのと、画面の裏側で何をやっているのかはよく分からないなという議論はあります。

先ほどのリカちゃん人形盗聴事件もそうなのですが、恐らくデータを収集しているはずですが、例えば方言、もしくはくせ、しゃべり方、後ろの雑音、大量にデータを収集しているはずなのですが、この辺が一体どこまでやっているのかとか、どの程度なのかというのはよく分からないというのがポイントで、公式には不適切な情報収集はしていないということで、そのとおりのだろうなと思いますけれども、時々マスコミ等で、実は研究部門が、特定のワードは聞いているというような話は聞いていますし、あるいは不適切な発言等が出ている、もしくは聞いてはいけないような音が出ているみたいなものは、データのサンプルとして採集しているのではないかみたいな議論があって、この辺りはブラックボックスですし、かつ、そのデータはサードパーティ側にはなかなか回ってこないという構造になっております。

そういう面で、完全に公正競争かと言われると、若干そういう状況ではない。日本自体は、実は2019年ぐらいに日本のLINEや2～3の会社は出しているのです。けれども、希望的な数や価格競争力等で全く勝負にならず、全て撤退しておりますので、実質的にここに日本プレイヤーはいない。サードパーティやスキルプレイヤーはいますけれども、プラットフォーム側はいないという構造になっています。

バイドゥ、アリババは、スマホ、PCを実際に持っていませんでしたので、そこを補完していたわけですが、ディスプレイの実装後は、住宅向けのマルチメディア端末になってしまいましたので、シャオミを含めて、中国国内は市場競争が激しくて、サー

ドパーティ、販路側に対して、競争環境制約等の動きは今は見えていないと言われてい
ます。ただ、そうは言いながら、アリババなどは電子決済の際にアリババの電子決済側
にテンセントのアプリを載せるのは駄目だという話になったりしたわけでありまして、
暗黙の了解的に、あまり競合他社のものは載せないよという構造感はある上がっている
と思います。

話をただただと垂れ流してしまいましたが、競争環境上から言いますと、物すごく排
除されているとか、サードパーティの作ったものをアマゾンやグーグルが置き換えたみ
たいな議論もありません。あるいは、参入するなという議論もなされてはいないと思
います。ただ、反面、どこまでデータを取っているのかとか、彼らが取ったデータがサ
ードパーティに健全な形でフィードバックされているのかというところは若干疑問が残
るところでありますし、自動車や住宅向けの展開というのも、彼らはやりたいと言っ
ているのですけれども、あまりいい条件は出ていないと聞いておりまして、一つの理由が、
その際に収集したデータは自分たちのものにしたいというコメントがあったやに聞い
ています。ですので、お付き合いした際にお客様のデータ、それも取っているか取っ
ていないか分からない音声のデータとかが持っていかれてしまうのではないかみたい
なところが若干不透明になっている。少なくとも共有にしろみたいな形になっていたり、
データの所有もしくはスキル、アプリの認証のほうをめぐりまして、若干商慣習的に不
透明と申しますか、コンフリクトと申しますか、そういったフリクションがあると申
しますか、そういった領域が残っているのは否定できないのかなと思う次第であります。

雑駁でございますが、私どもからお伝えしたいのは以上です。

御清聴、どうもありがとうございました。

○依田座長

どうもありがとうございました。

それでは、上野山議員、伊永議員、川濱議員の順に、御質問、コメントを伺い、グ
ループに区切って桑津理事から御回答いただきたく思います。

まずは上野山議員からお願いいたします。

○上野山議員

桑津さん、プレゼンテーションありがとうございました。

いろいろと聞いておる中で、現状、足元には競争観点は大きくは存在しないという
ところ、私もそういう形かなとは思っております。一方で、未来において起こる可能性
が高い領域であるのもまた事実でありまして、未来に先回りして、一定、動ける構
えを想定しておくという意味で、非常に重要な論点だなと思われました。

それに際していろいろ聞く中で、こちら辺の論点が重要なのだろうなということ
をちょっとコメントさせていただいて、もし御意見等があれば、いただければな
と思っております。

大前提として、議論の射程として何年先ぐらいまでの未来を見据えるのかによ
って論

点が変わってしまうのでそこは一定認識合わせたいです。その前提がある上で3つの論点に関してコメント致します。

論点の1つ目は、現状どのような複数間プレイヤーの連係構造があるのか。基本は今、垂直統合なのだと思うのですけれども、これが近い将来どのような形で、どこか一部の例がオープン化して、App Storeみたいな構造になっていく動き等々があるのかどうかというところのプレイヤー構造をどう捉まえるのかというところが1点目でございます。

2点目は、複数プレイヤーの分担が現状行われる中で、独占性の議論が起こるようなレベルのある種のインテリジェンス・知能的な役割がどのレイヤーに宿っていくのかということに対する仮説が非常に重要になってくるかなと思います。これが論点2です。

3つ目、最後ですけれども、どの時間軸でどのレイヤーのインテリジェンスレベルがどれくらい上がってくるのかという、誰も答えが書き切れないことではあるのですけれども、結局この見立てみたいなのがほとんどこの競争の議論の一番大きな要素になるので、少なくとも仮説を持ちながら議論したほうがいいのではないかとということで、大きくこの3つです。これは全部、誰も答えが分かるものでもないと思うのですけれども、一定分かるもの等々、お考えがあれば教えていただければと思います。

以上でございます。

○桑津様

どうもありがとうございます。

非常に本質で、重たいところの質問なので、私もなかなかお答えし切れないところなのですが、現状、垂直統合になっていて、どこのレイヤーが切り出されたら水平オープンになるのかということだと思うのですけれども、自動車屋さんのお話を聞いていると、例えば音声認識エンジンの部分をまず一部切り出す。これは音声認識の機構そのものではなくて、その後ろ側についている単語や認識のくせのデータベースのようです。というのは、車の中で話す会話と家の中の会話は違うので、使うシラバスや雑音環境みたいなものは外へ切り出して、独自でやってくださいみたいな議論はされていたようです。ですので、開放しているというよりは、自分の手の届かないところを任せるみたいな形で連携を図りたい。

ちょっと嫌らしい言い方をしていたのが、その際のデータも全部こっちに下さいねみたいな話をされていたようで、そこが非常に引っかかっていたのかなと思います。

ネットワークは開放されていますし、端末もハードウェアは、誤解を恐れずに言えば、軽めのプロセッサとマイクとスピーカーがあればいいわけですし、誰が作っても同じなので、そこはある意味、簡単に開放されてしまうのだろうなと思っています。

よく言われる音声認識エンジンのところは、先ほども申しましたが、エンジンの機構そのものは開放するのかもしれないのかというのはアマゾンの中でも声が2つぐらいあって、するという人としらないという人がいて、どちらも正しいみたいです。すると言っているのは先ほどの自動車パターンと同じで、エンジンのソフトは提供するのだけれども、結

局そこを解釈したり、どのような観点で評価するかという後ろ側のコンテキストをコントロールすること。夜の話なのか、子供の声なのか、車の中にいるのかみたいなくせ、そこは開放しないよという言い方で、結局それは開放していないではないかという話になるのか、そういうところが、全部開放というよりは、一部機能を分担して、特化したものは作っていいよみたいな議論になってくると思われています。

独占性なのですが、現状、端末が非常に安くて、かつ、圧倒的に強いというレベルで、誰が見ても正しいという独占には正直なっていないなと思っています。どこに独占が発生するのですかと。車なんかは、例えばどこかの大手メーカーが採用したら、それが全部独占になってしまうのではないかみたいな話があるのですが、まだそうになっていません。同じことが住宅に関しても言えるのかなと思っておりまして、先ほどちらっと申しましたが、第1優先順位であるスマートスピーカーは結局、大昔から言われている住宅のIT化に限りなく近い構造になってくるなということで、住宅に組み込みで、要は積水ハウスを買ったら必ずアマゾンが入っているみたいな構造になるのではないかと、というのが恐らく独占に近い形。ただ、日本の住宅メーカーは非常に分散していますので、どこか大手の市場の完全独占には届かないのですけれども、恐らく高級住宅に入るのは間違いのないのだろうなと。例えばエアコンにしても、私もそんな家は住んだことがないのですけれども、結構高級めのマンションや住宅をお伺いしていますと、リモコンの数がやたら多くて大変だみたいな話があって、この部分は統合しようみたいな、あるいは寮、ドミトリーとかマンションみたいな規格化された部分は、全部一まとめでいいのではないかみたいな議論で恐らくタイムフレーム上、独占というのが出てくる。つまり、他社が入れない、面倒くさくてとても入れられない、実質難しいみたいな議論が出てくるのかなと言われていています。ただ、これはタイムフレームで言うとこれから出始めて、ただ、10年先ではなくて、もう要素技術は出来上がっているの、あとは住宅のメーカーさんとかそういう方々が、いつどれぐらい入れるかというそれだけの話だと理解しています。

実際、実は日本の住宅メーカーの大手さんに来ているみたいです。買ったときに、ただで入れてあげますよみたいなことは言っているようです。ただ、その代わりデータを全部よこせと言われていたので、そんな契約は怖くてできないということで止まっているというのが2年間ぐらい続いていると聞いています。

時間軸なのですが、ちょっと言い方は悪いのですが、要素技術はほぼ完成しているのです。音声認識のパターン分けも、ニッチの部分とかで日本語対応とかも特に問題はありませぬので、あとはテクノロジーというよりはインプリメンテーションに徹していて、これをいかに数が大きい人が、ユーザーの目に見えない形で抵抗感なくすーっと入れることができるかどうかになってしまったなということでして、車はその先頭だと。ただ、車の方は、いろいろなトラブル等が起きたり、その部分をインターフェースを取られるのは嫌だという話になっていまして、簡単な操作ならば自分でやるよみたいな感じで戻しています。ですので、住宅もしくは家電、冷蔵庫は違うのですが、

恐らくテレビ、照明、あとはエアコンの辺りに大量に入り始めるのがポイントかなど。特にダイキンさんとかいろいろな方に聞いていると、もう2年ぐらい前からずっとこれを入れようという話はずっとしているということですので、これが入り始めると、彼らからすると複数のメーカーで競争してほしいのだけれども、ネットワークのくせとかデータの収集の都合を考えると、例えばリモコンは複数メーカーに分散して発注してもらっていいのだけれども、データを収集することを考えると、1つの会社に束ねたほうがいいように思うよねという言い方はされていますので、この辺りが潜在的なタイムフレームとして危ないなということで、私の予測としては、2022年ぐらいから、例えばエアコンやテレビのより内蔵本格型が出始めてきて、当初高級住宅、それがドミトリーや寮みみたいな形で広がってくるのが2025年ぐらいかなと思っていますので、このタイミングでは、独占というのは割と真面目に心配しなければいけない。その際に、データとかは結構持っていかれていて、誰がどうなっているのかみたいな、そもそもその家の中に人がいるのかとか、何人いるのかとか、使っている電気の量の評価と、どんな会話がなされているか、何時から音がしているか、そもそも騒音はどのような音がしているのか。

例えば、これも笑い話で出たのですけれども、テレビの音を聞いて、番組を指定して、視聴率を測定できるのではないかみたいな話が実際にあったと聞いていますので、思ったよりいろいろなことができる構造になっていて、そのデータが取れるポテンシャルが不定形ながら存在しているというのが実態だと思います。

以上です。

○上野山議員

ありがとうございました。

○依田座長

ありがとうございました。

それでは、次に伊永議員、川濱議員、塩野議員のお三方まで行って、そこで一区切りつけて、桑津様から御回答いただければと思いますので、次、伊永議員、お願いいたします。

○伊永議員

ありがとうございます。都立大の伊永と申します。経済法、競争法を専門にしております。

私のほうからは、競争法の観点から1点だけ質問なのですが、お話を伺っていて、個人情報保護法上の懸念はいろいろあるということが分かりました。一方で、スマートスピーカーの市場で、午後8時から午前8時までの情報が補完的に得られるのではないかというお話だったのですが、それに対してグーグルなどは文字検索だけではなく、モバイルでの音声情報、位置情報など、いろいろなものを使っており、あくまでもスマートスピーカーはその補完にすぎないと思ったのですが、そういったグーグルに対して、果たしてアマゾンが競争上の脅威になっているのか、そこはどうかというのが、私の

質問になります。

家や自動運転、それからIoTの家電といった個々のマーケットについては、それぞれ問題が生じないようにウオッチしていかなければいけないというのは警鐘と受け止めましたが、一般的なグーグルとの競争関係において、アマゾンというのが脅威になっているかどうかという点を、専門的な観点から教えていただければと思います。

以上です。

○依田座長

では、続きまして、川瀆議員、お願いいたします。

○川瀆議員

どうも興味深いお話をありがとうございます。

私は、京都大学で経済法を専攻している川瀆と申します。

私も今の伊永議員と同じ疑問を持ったのですけれども、同時に、今回のボイスアシストの問題というのは、先ほどから議論がありましたとおり、以前からこういったプラットフォームの持つデータの力の問題が気になっております。音声認識の問題というのは、今までの手の入力で収集される個人情報以上に、我々の個人の生活自身が完全に丸裸にされてしまって、その結果、ある種、潜在的に持つ脅威の力が大きいかも分からないという点と、さらにそのための要素技術はもう完成されたという今のお話がありました。逆の側面、大きな脅威であるけれども、脅威をもたらすほど、我々の日常に接近した利便性ももつはずです。それにもかかわらずこれが進展しないということも問題のような気がします。

もちろんどの程度進展するべきなのかという問題はあるかもしれませんが、今のところうまく連携が進んでいないという点と、日本の状況を見ると、かなりお寒い状況になっている。もともと、日本語に関しての解析が遅れたらどうなるのかなという心配があるところですが、思うに要素技術の点においては障壁がないというのが先ほどの御説明にあったかと思いますが、その問題はないとしても、十分なデータを構築して、日本語に関してそれを遂行しようという企業がどれだけあるかということは問題なのはと気になります。

それに関連して、将来の競争上の観点でお伺いしたいのは、もちろん多量のデータをもって先行して利益を得る者が強い地位をこの音声において獲得する可能性は将来あるかも分からないと思うのですけれども、今ここに入ってこられているプレーヤーたちに限定されているものと考えていいのかをまず確認させてほしいところです。潜在的な参入の技術的な能力と、巨大なクラウドを構築して入っていく点で、すごく大きな規模の経済性があることから、今のところアマゾン、アップル、それからグーグルのようなクラウドも含めたビッグテック以外は入ってこられないということから、それらのうちのいずれかが力を握ると考えられるのか、それともビッグテックが並行して、要するに、この場合に検索のような形でやるのか、それともそれを越えた企業が併存する形で存

在し得るのかという点なんかも含めた御関心をお教えいただければと思います。

どうぞよろしく願いいたします。

○依田座長

あと、もう一方、塩野議員まで行きましたら、桑津理事のほうにお戻ししたいと思いますので、塩野議員、お願いいたします。

○塩野議員

塩野でございます。お話しいただき、ありがとうございました。

私も、2014年前後ぐらいに、日本の家電メーカーが音声認識でこれをやる、やらないという話を結構されていて、そのときも私の感覚としては、スマホを取られてしまったので、スマホより人間に近くなるには、もう声しかないでしょうと。そういう意味では、スマホを取られたので、声を取ると。より近いところを取れるか、取れないかで、機械学習を始めた人が勝つので、始めたほうがいいのではないかという会話をしたことをよく覚えています。

そこまでは進まなかったですけれども、私の場合、車かなと思っていまして、というのも、ドラレコは音声をずっと録っているので、もう慣れているということと、もちろん声でいろいろコマンドしたほうがよいというのは車であります。ただ、これまた車メーカーが会話をしていると、インフォテインメントなのか、コックピットの運転支援寄りなのか。運転支援のほうはもうちょっと真面目にやらなければいけないほうだねというところで、どっち側にこれを関わらせるのかということと、まさにエンジンで、先ほどの米中デカップリングのお話がありましたけれども、こちら辺で米国物なのか、中国物なのかをやるので、今後、車自体がすごくセンシティブになってしまうので、そこは気をつけたいねといった会話がありました。

現在思いますのは、GAFAが怖いという車であったり、家電であったり、また先ほどおっしゃった家であったりというのは、IDを取られてしまうと、Apple IDで車にログインされて、全部シームレスにされたときにどうするのか。オクケーグーグルでIDが、車のIDというのはなかなかなかったのですけれども、ずっとトラッキングはできていなかったのですけれども、その一助にこの音声アシスタンスがされてしまったらどうしようと。IDの取り合いという中では、一つのパワーになるのではないかと。

まさに今般出てきたアマゾンのAstro、自立的に動いているロボットです。コミュニケーションロボットは今ちょこちょこ出てきていますけれども、あれは音声のコマンドが入っていて、先ほどおっしゃったように、後ろで誰かが聞いているのではないとかいう話で、ロボットの顔をして、同じことをしてきている。

ただ、コミュニケーションは我々が思っているよりも広がる白地が何かあるかと思いますので、そうしたときに、いま一度、ここで学習した音声というものが非常にキーになってくるのではないかというようなことを思っていて、ぜひコミュニケーションロボット回りとかのお考えをお教えいただければ幸いです。

ありがとうございます。

○依田座長

ありがとうございました。

桑津理事、ここまでの伊永議員、川濱議員、塩野議員の御質問等について、御回答できる部分がありましたら、お願いいたします。

○桑津様

ありがとうございます。

まず、最初の点として、そもそも例えばボイスアシスタントは携帯と比べて脅威になっているのかという御質問があったと思います。

結論から言うと、なっていないです。当たり前ですけれども、アップルの製品を見れば明快のように、ボイスアシスタントの機能は別に携帯に入れても何の問題もない。むしろ使い勝手がよくなるということで、より包含するべきものです。独立して出さなければいけないという性格のものでは本来ではないと思います。

ただ一方で、携帯はパーソナルですので、家から出るときに、家の中に携帯は残っていません。ですので、そういう補完的な位置づけとして、家の中に端末を置いておきたいという議論の中で、家の中に入っているボイス、外に出たときは音声認識は使いませんという制御系に使うみたいな議論があるようで、プライムは携帯なのだと思います。人間が活動するところに置く。ただ、住宅という観点で補完し得る面で、ボイスアシスタントが入ってくる可能性が非常に高い。同時に、冷蔵庫や照明やテレビにそれが入ってくるという構造により近づいてきているのかなと思っています。

当初は、実は家の中で寝ている時間は有効だとかと言っていたのですけれども、これもApple Watchとかを使い始めたので、むしろそっちのほうが有効だという話になってしまっていて、正直、スマートスピーカー系で今、ヘルスケアはあまり言わなくなっているなと思っています。

2番目のパーソナルとホームとか、将来の競争軸とか、そもそも参入が限定されないのかという議論で、先ほどからのことと重複しているのですけれども、ボイスのスピーカーというのは、アングルでマーケットを切り出すことはできるのですけれども、スマートフォンのように包括的にパーソナルな領域で市場を全面的に牽引するという力の製品は出ていないのが実態です。あくまでスマホでできないところや、より利便性の高いアングルを切り出しているというように私どもは理解しています。ただ、それでは家の中のいろいろな家電等に入ると、思ったよりも使い勝手がよくなるよねと考えております。

それが実は最後のコミュニケーションかそうではないのかという議論につながってまして、当初、ボイスアシスタントはコミュニケーション系だと理解していました。つまり、それも人間対機械のコミュニケーションです。機能はあるのですけれども、人間同士で話し合うのはあまり使っていない。人間が機械に話しかけるためのコミュニケー

ションで、明かりをつけろだの、スイッチを消せだの、暑いからエアコンを入れろだの、そういう話があったわけです。あと、ニュースを聞くとかもあったのですが、今、データを収集している部分は人間が指示していないところ、例えば先ほど言った後ろ側で赤ん坊が泣いているとか、温度が高くて、そのときにエアコンのスイッチを入れろと言っているとか、3人以上でしゃべっているとか、横でスマートメーターで見たときに、同時にスイッチが入っていないのはこれだみたいに、どうもコミュニケーションはあくまで一角であって、その後ろ側にある神経系を狙ってくる方向になるかなと。だから、ボイスアシスタントはあくまでその入り口の皮一枚みたいな感じにして、人間との接点のインターフェースの部分がボイスアシスタントなのであって、スピーカー機能とネットワークとクラウドでいろいろなものを制御しましょう、判断しましょう、データを集めましょうみたいな方向感の入り口として、ボイスアシスタントは人間との接点の部分で使われている。

ただ、冰山ではないですけれども、見えない、もしくは聞こえない下の側のほうで、いろいろなことができますよ、データが取れます、センサーがつながっていますみたいな議論になっていると思います。

実際、アマゾンのアレクサ等の状況を見ていると、今はかなりつながってきているのが、スマートスイッチであったり、あるいはスマートリモコンであったりという感じになってきて、結構センサー系、人感センサーであったり温度であったりにつながってきているという状況で、暑いと判断したら、暑いからつけてと言うのですけれども、それを28度にしてとか、28度になったら自動に入れるみたいな機能に切り替えてきます。ですので、入り口はアレクサは音声で指示しているのですけれども、後半のほうになってくると、前提条件や無言でもやれるようになる。家の近くまで帰ってくると、携帯のGPSと連動して、エアコンのスイッチ、電気が入るよみたいな方向感が出ていまして、音声はあくまで機械と人間をつなぐ部分、インターフェースの部分のアンクルの一つなのだろうと思います。それは先ほどおっしゃった非常に自然なアンクルだと。つまり、あまり勉強しなくてもいいよねというところでポテンシャルがあって、誰でも使えるようになる方向感に向かっていくのかなと思います。

ただ、タイムフレームの御議論があったように、思ったよりも音声認識は、ある意味普及しない。いろいろな言い方があるのですけれども、あまりなじまない。この辺は多分カルチャーの違いなのかなと思っていて、今、例えば電話もそうなのですけれども、最近はワイヤレスのイヤホンで携帯をポケットに入れながら、しゃべりながら電話をしている人を時々見て、そんなに違和感がない。大きな声でしゃべっているときは、こいつは変なやつではないかという話になるわけなのですけれども、時々見るようになっていきますよね。

これは5～6年前だと、スマートイヤホンで独りでしゃべっている人間はどう見ても危ない人間で、みんな遠巻きにしていたわけでありまして、しゃべってコミュニケーシ

ョンを機械と取りますというところも2025年の段階になれば、それは当然のものだという世代が結構増えてくるかなと思っています。

音声で何でもできるというわけではなくて、音声は、まずはコミュニケーションなのですけれども、その後ろ側にあります制御とか神経系、内部に近いところにつながるインターフェースとして、結構そこは重要性が増してくる。ただ、将来的には人間がいない家の中でも、帰ってくるのであれば自動的にスイッチが入るというのは、音声ではなくて、普通につながっているアレクサとかそういったところの機能側で対応することになるのかなと思っています。ですので、ボイスとは結局何なのかという話になるわけですが、なじみやすく、かつ補完性が高いメディアの一つであるということは間違いないのですが、そこで完結するものではなくて、その水面下にかなりたくさんの機能があって、そちら側が今のボイスアシスタントを形成しているクラウドやプラットフォームの部分と改めて相性がいい。だからこそ単品の製品で普及するのではなくて、家を建て替える、車を買って替えるといったような母体になる製品がごろっと交換するタイミングにならないと、なかなか入ってこないというのが私どもの考えるところであります。

全部お答えし切れていないのですが、以上のように考えている次第です。

○依田座長

ありがとうございました。

では、最後に後半3人の議員から御質問いただきたいと思います。

まず、増島議員、お願いします。

○増島議員

ありがとうございました。

もうおおむね聞いていただいたのですけれども、1つだけ教えていただきたいのが、今おっしゃるような状態をつくるときに、結構データが膨大になるという話があって、分析をするという話なのですけれども、後ろ側のハードウェア側が、こっちも成長しているという話もありつつも、実現が可能なレベルなのか、つまりハードウェア側のリソースが技術的・経済的にそうした構想が成り立つ程度で提供できるのかという部分について、感覚的なものでもいいので少し教えていただけないでしょうか。画像ではないので、そこまでひどくはないというのは何となく分かるのですけれども、その点だけ教えてください。

○依田座長

ありがとうございます。

続きまして、山田議員、お願いいたします。

○山田議員

ありがとうございます。

2つ質問なのですけれども、一つはこういう新しいデバイスは、やはり収入源があるからこそやるのだと思うのですけれども、いろいろ本当に面白いプレゼンをいただいて、

研究される中で、彼らが収入源として明確なロードマップがあるはずなのですけれども、何を考えているのかということはどう思われますかという質問です。

アップルのiPhoneみたいにデバイスを販売することで、収入源として将来価格も高くなるかもしれないし、自動車、家にも入れるかもしれないしというデバイス販売を狙っているのか、それとも、フリーで提供して、広告費とか、アマゾンであれば販売促進ですよね。あるいは、先ほど人間との接点と何回かおっしゃっていたので、そこまではっきりした収入源というのがクリアになっていなくても、もっと抽象的に10年後にはもっといろいろなポテンシャルが、これをやっておけば、これを押さえておけばと出てくるだろうと。そういうロードマップで彼らがやっているのか、それをどう思われますか。独禁法の観点からのコメントなのですけれども、最初にプラットフォーマーのドミナントが広がっていくところを逃してしまったという前提で今いろいろなことが議論になっているわけなので、これら新しいデバイスについて今後どう広がっていくのだろう、という御質問です。

もう一個は簡単な話なのですけれども、中国製品がどのくらい日本市場に入ってくる可能性がありますかという点です。これは独禁法というよりも、安全保障とかいろいろなソーシャルセキュリティーの話です。

○依田座長

山田議員、よろしいでしょうか。

○山田議員

これで以上です。ありがとうございます。

○依田座長

では、続きまして、北野議員、お願いいたします。

○北野委員

北野です。

大変面白い情報、インフォーマティブなプレゼンテーションをありがとうございました。

動向とかを御存知ならばということなのですが、今の音声アシスタントの対話能力というのは、一問一答が基本ですよね。フォローアップしながら会話するという能力はなかなか難しいというか、結構いろいろなところのものを試してみたのですけれども、なかなかシンプルなものしかできないわけです。

3～4年前にグーグルがGoogle I/Oで発表したGoogle Duplexは、ドメイン限定ですけれども、結構な連続対話能力をデモとして発表したのです。そういうレベルのものが実装されるという動きは何か御存じかどうかということをお伺いしたいのです。

というのは、私は博士論文が対話型の実時間音声翻訳システムを作っていたので、自由対話の難しさというのは嫌というほど思い知っているというか、対話できる、できないかではなく、人間同士がそもそも自由対話がちゃんとできていないわけです。それを

機械に求めるのは不可能なことなのではすけれども、ただ、タスクオリエンテッドなところでというのは、私がドクターを取ったときですから30年ぐらい前もある程度できたし、今はかなりのレベルのものができらるうとは思ひ、Google Duplexの能力というのは非常にインプレッシブなものだつたと思ひのです。

それが出てくると、音声アシスタントの使い方とか、浸透の度合いがちょっと違ひ次元のものになると思ひのです。今は検索の代わりみたいなものなので、そのレベルで終わっているのではすけれども、そういう動きは何か察知しているものがござひますかというのが一つ。

もう一つは、キャッシュストリームをどうするかという先ほどの質問でふと思ひたのではすけれども、これがかなり日常生活に浸透する、例えばプロダクトプレイスメントみたいな、会話の中にふとリコメンデーションが入ったりとかというような形のキャッシュストリーム。あとは、アマゾンならばアマゾンのeコマースがあるから、そこに対するインターフェースなので、そこでわざわざ新しいリベニューストリームをつくらなくても、ECの間口を広くするというこゝでいけるのかなと思ひのですが、もしプロダクトプレイスメントすると、昔、『トゥルーマンショー』という映画があつたのを御存知ですか。あれは全部セットアップされているので、トゥルーマンの家族であつたりとかいろいろの人が、突然プロダクトプレイスメントを始めるのです。だから、あれに似たような状態になるというか、要するに、音声アシスタントがいきなりプロダクトプレイスメントをし始めて、何が起きているのかみたいな、あのようなことにはなるのではないかと、ふと今、思ひて、ちょっとおかしくなつてしまつたのではすけれども、何か今までとは違ひような対話のエコシステムになるのかなと思ひたのです。そこら辺に関して、何か御意見等をお伺ひできればと思ひました。

以上です。

○依田座長

ありがとうございます。

それでは、桑津理事、どうぞよろしくお願ひいたします。

○桑津様

桑津です。

まず最初に、後ろ側にどのぐらい膨大な音声処理ができるような体制もしくは仕組みが出来上がっているのかというお話だつたと理解しています。

これは私も画像認識のときとかにやつて、3,000人集めてすごいとか、グーグルのフォトで30万人集めてすごいとかという議論があつたと理解しておるのではすけれども、今はさらに大規模なデータをウォッチできる。ウォッチというか、ずっと記録して、データサンプルを取れるという状況だと聞いています。

言語対応を行うに当たつて、基本的にはスタート3,000、それから3万というのが、大体1か月間ぐらいデータを収集してやれますというのが割とルーチンにこなせますよと

ということですので、どのぐらいの人間の声の数を処理して、いろいろオペレーションな処理をできていますかということに関して言いますと、万ではなくて100万の単位が割と普通にできるという環境が整っているというか、ある種学習もできるような状況になっているということで、かなり大きいデータ量は処理しているのかなと思います。

2番目に収入源なのですけれども、御指摘のとおり、ハードウェアでは絶対にもうからないというのはみんな分かっています、理屈から言うと、ハードウェアにいろいろなアプリを載せて、そこに有料アプリが載るというスマホみたいな展開だったらよかったですけれども、実態はそういう方向には向かっていません。当初、収入化を目指していた自動車のところで条件を幾つか聞いたことがありますけれども、自動車のときは、ハードウェアをやらせてくれれば1個せいぜい1,000円ぐらいでハードウェアを入れて、先ほど言った自動車ならではの音声のシラバス、もしくは自動車ならではの最適化されたフィルターの処理みたいなものをセットで用意して、そのためのリージョンというか区画を用意する。そこを使ってくれということで、アマゾンのクラウドを使って、自動車専用最適化されたシラバスと、それから雑音処理、音声処理、あるいはシチュエーションの認識みたいなものをセットで用意して、その学習を整えるほうまで用意しますと。使ってくださいと。使ってくれていれば、1人当たり年間何ドルか、これは大した金額ではなかったと思いますけれども、4~5ドル下さいみたいな議論があったと聞いています。

これは実際にそこまで行かなかった。第1号でそこまで成果が上がっていないので、BMWはそこまで行っていないと聞いているのですが、ハードウェアを何個か売って、利益がどうこうというよりは、セグメント別のほうで巨大なセグメントに対して専用の環境を用意して、専用の音声認識のエンジンをカスタマイズして、専用のシラバスを用意して、運用をする。そこから費用を頂きますというモデルだと聞いております。

3番目の中国製品なのですけれども、実は初期の2014~2015年にLINEさんとかが入れたのは中国製品だったのではないかとされていて、本来入るはずだったのだろうなど。ただ、なぜか日本の場合には今のところ入ってきておりません。日本語のエンジンの最適化等ができないかと聞くと、アリババさんに聞くと全然難しくないからできるよみたいなことを言っていて、実際、アリババのスマホに入っている音声認識の自動翻訳のソフトを見ると、グーグルに全然負けていないなという感触は持っていますので、普通にできると思います。ただ、商売の優先順位上、10分の1もないマーケットで、かつ単価が高いのならばいいのですけれども、単価が高くなる要素はないので、今、現状は日本側でこれの展開をしようという計画は今はないです。

ただ、中国のマーケットに対して一定の決着がつけば、彼らは決済のときと同じで徐々にしみ出してきますので、2~3年後ぐらいに中国製のものがたくさん並んでいたとしても、それは特に違和感はありません。

全部テンセントクラウドで言語を取っていて、そっと聞かれているのではないかと

われると、その可能性はあるでしょうね。そういうことは特に違和感なく予想されるところだと思います。

4番目に、一問一答で連続対応できないではないかという話はおっしゃるとおりです。今、アマゾンのもので、連続対応モードというのがあるのですが、結局、はいといえを順番に答えるので、アレクサを言わなくてもいいみたいな、全然連続対応ではないではないかみたいな仕組みです。ただ、これもお伺いしていると、どちらかというところは学習で対応するような方向で、一遍に聞いたものを区分けして、順番に答えるとか、ある程度優先順位を入れ替えるみたいな連続対話対応は、学習させてくればできるようになりそうです。これもさっきの自動車と同じパターンで、エアコンのほうに関してそういうやり取りをできるようにしましょうとかということはやっているようですので、完全な連続対応が可能な学習モデルが明日出るという構造ではないのですが、それに近い物を用意する環境は整ってきていると思います。

最後の『トゥルーマンショー』に代表されるプロダクトプレズメントの議論なのですが、アマゾンは自分で商品売っているという面に関してはあります。私もアレクサを入れているのですが、お茶の賞味期限が切れましたとか、普通に言ってくるので、注文しませんかと聞いてきますので、はいといえば注文されてしまうと思います。ですので、プロダクトプレズメントではないのですが、そういう素朴なプロダクトインボークーションぐらいはできているなと思います。

ここに本当の広告宣伝が入るのかという議論なのですが、これは検討であるのです。

こういうので、日本企業はそこまで到底行かない段階になっていますので、検討の課題のテーマ自体が先に進んでいる。プロダクトプレズメントは普通にあると思います。言い方を変えると、そのバックボーンをサードパーティに販売するというのもあります。ただ、しゃべっているやつが突然商品をしゃべり出すというのは非常に気味が悪いですし、今は中国のスマートスピーカーはべらべらよくしゃべって、アマゾンの3倍ぐらいしゃべるのです。アマゾンがあれぐらいしゃべるようになると、うるさいからボリュームを下げろみたいな本末転倒になってしまうような気がしますので、その辺はまだ市場の環境というか認知度がどうなるか分からない。ただ、アプローチの一つとして、キャラクターというのが出てきて、その方面、アシスタントになるのか、キャラクターになるのか分かりません。昔のWindowsみたいに変なイルカみたいになると嫌だなと話しているのですが、そういう方向感で親和性を高めると聞いています。

以上です。

○依田座長

どうもありがとうございました。

最後のキャラクターの部分は大変興味深く拝聴いたしました。もう少し議論したいところではありますが、お時間が来ましたので、以上で本日の議論を終了いたします。

最後に、事務局のほうから連絡等をお願いいたします。

○成田審議官

ありがとうございました。

私のほうからは、次回のワーキンググループは10月13日の予定になっております。

私からは以上でございます。

○依田座長

ありがとうございました。

以上をもちまして、本日ワーキンググループを終了いたします。桑津理事、今日はどうもありがとうございました。

○桑津様

ありがとうございました。失礼いたします。