

Cover sheet for response to an Ofcom consultation

BASIC DETAILS

Consultation title: **Speaking TV guides: would they help people with visual impairments and are they feasible?**

To (Ofcom contact): **Jacopo Genovese**

Name of respondent: **Guido Gybels**

Representing (self or organisation/s): **Self**

Address (if not received by email):

CONFIDENTIALITY

Please tick below what part of your response you consider is confidential, giving your reasons why

Nothing

☒

Name/contact details/job title

☐

Whole response

☐

Organisation

☐

Part of the response

☐

If there is no separate annex, which parts?

If you want part of your response, your name or your organisation not to be published, can Ofcom still publish a reference to the contents of your response (including, for any confidential parts, a general summary that does not disclose the specific information or enable you to be identified)?

DECLARATION

I confirm that the correspondence supplied with this cover sheet is a formal consultation response that Ofcom can publish. However, in supplying this response, I understand that Ofcom may need to publish all responses, including those which are marked as confidential, in order to meet legal obligations. If I have sent my response by email, Ofcom can disregard any standard e-mail text about not disclosing email contents and attachments.

Ofcom seeks to publish responses on receipt. If your response is non-confidential (in whole or in part), and you would prefer us to publish your response only once the consultation has ended, please tick here.

☐

Name **Guido Gybels**

Signed (if hard copy)



Response to Ofcom's consultation: "Speaking TV guides: would they help people with visual impairments and are they feasible?", dated 10 July 2014.

Guido Gybels, Independent ICT Expert – www.guidogybels.eu

Introduction

1. I welcome the opportunity to submit comments to Ofcom regarding this consultation on Speaking TV guides. I am an independent ICT Expert with significant background and experience in a broad range of technologies, including Digital Television and Text to Speech systems. Alongside an extensive career of almost 3 decades in mainstream Information and Communication Technology, including consumer equipment expertise, I also have a long track record in the field of accessibility and usability.
2. I am one of the contributors to International Standard IEC 62731:2013 on "Text-to-speech for television - General requirements" and am currently the editor of a draft standard on Digital Television accessibility being developed within IEC TC100.
3. I am also an Expert appointed to the CEN-CENELEC project delivering EU Mandate M/473 "Standardisation mandate to CEN, CENELEC and ETSI to include "Design for All" in relevant standardisation initiatives" which has some bearing on aspects of this consultation.
4. The information in this response is not merely opinion, but stems from my work practice and is thus based on factual evidence, including from constant interactions with service providers and equipment manufacturers, as well as industry-led standardisation bodies.
5. Because of commercial confidentiality and other contractual obligations I cannot name specific companies or products and I will not divulge privileged information.

Q1. Do respondents agree with Ofcom's initial assessment that apps for mobile devices have the potential to be useful for those people with visual impairments who feel confident using touch-screen technology and can afford a suitable mobile device? If not, why not?

6. Yes, I agree. Many blind and partially sighted people already use mainstream Android and iOS tablets through the accessibility features of these platforms. There is therefore an opportunity to harness this for spoken access to Electronic Programme Guides (EPG) and/or command and control over television solutions.
7. However, there remain a significant number of blind and other disabled users who do not have modern smartphones or tablets. As such, this approach will not address in full the access needs of all consumers with sight loss.
8. Even amongst those who do have smartphones and tablets with the necessary accessibility provisions to use them effectively, there remains demand for built-in spoken features in digital television solutions.
9. Receivers increasingly have provisions for external connectivity to apps on smartphones and tables, or to specialised alternatives to traditional remote controls. Equally, more and more receivers now extend the functionality of such connectivity interfaces from the home network to more remote operation, typically through a cloud based interface.

10. However, such connectivity provisions (in general an API available over the home IP network) frequently do not offer a complete suite of functions as required by external accessibility solutions.
11. Moreover, the theoretical potential for smartphones and tablets as accessibility accessories for television receivers is in practice constrained by the fact that the programming interfaces on television receivers are generally proprietary in nature, poorly documented and usually subject to confidentiality. The ability of a receiver to connect to a (own-brand) smartphone or tablet allows manufacturers to provide added value to their product and creates opportunities for upselling and vertical integration. Consequently, protocols are treated as (confidential) IPR and this makes it very difficult for third party application developers (who may specialise in accessibility) to create fully capable accessibility apps, of which a spoken interface to a television that incorporates a spoken EPG could be one example.
12. As such, the theoretical potential for opening up television receivers to blind people and other disabled users with specific abilities and preferences is only realised in a very partial manner.
13. Furthermore, even where apps for command and control of receivers exist, it is not always the case that these apps are able to collaborate in full with accessibility tools on the device, such as screen-readers.
14. These observations do not invalidate the general assessment that apps have great potential (and indeed there are some solutions out there already), but it does suggest more can be done to realise such opportunities. Ofcom could play a key role in facilitating the discussions and standardisation work needed to open up these opportunities and to create a better functioning market in this area.
15. It is not possible to create a single user interface which will successfully service all possible end users effectively and in full. As such, there is great scope for externally connected solutions to complement features of the built-in interface.

Q2. Do respondents agree with Ofcom's initial assessment that apps for mobile devices are less likely to meet the needs of the majority of visually-impaired people who are 65 or older, both because they are less likely either to own a suitable mobile phone and because touch screen apps present a number of actual and perceived barriers to use. If not, why not?

16. I agree in most part. It is clearly still the case that prevalence of smartphones and tablets amongst those aged 65 and over remains significantly lower than for other categories. This applies to both mainstream consumers as well as those with sight loss.
17. However, Ofcom's own data as well as that provided by other organisations, such as Deloitte's Global Mobile Consumer Survey, indicates the gap is decreasing as time goes by.
18. Nevertheless, despite the accessibility features now available on the main mobile Operating System platforms, substantial barriers remain. While some strongly motivated users with sight loss manage to work around barriers, other users find such devices hard or impossible to use.
19. In addition to specific accessibility features, the same general usability barriers that make technology in general challenging for much of the population also play a role (and these usability barriers may be amplified by accessibility issues).

20. Users that are generally confident with technology are more likely to be able to harness smartphones and tablets and this is not substantially different amongst blind and partially sighted people.
21. It should be noted that some platforms perform better than others in terms of accessibility while tools like screen-readers are not available for all mobile Operating Systems. In some cases, purchasing extra applications may be required in order to turn a smartphone into a properly accessible device for blind people.
22. As stated under the response to question 1, even amongst those who do have smartphones and tablets with the necessary accessibility provisions to use them fully, there remains demand for built-in spoken features in digital television solutions.

Q3. Do respondents consider that it would be reasonable for visually-impaired viewers to pay more than sighted viewers for the ability to use EPGs or substitutes for the same purposes as sighted viewers? If so, why?

23. I do not believe that it is realistic (economically nor technically) to expect built-in spoken interfaces to be available for all television receivers available on the market.
24. Similarly, external spoken solutions (including spoken EPGs) have some cost associated with them and even if this cost is subsidised or otherwise diverted, such solutions require smartphones or tablets that will need to be paid for one way or the other.
25. There is no evidence to suggest that spoken interfaces will become so popular in the mainstream that they will be incorporated as standard features across most receiver solutions.
26. It is an unavoidable reality that additional features in a product almost always require additional effort and additional cost. For high-end Connected TVs, a talking interface (including the EPG) is more attainable than for low or mid end devices where extra costs can quickly undermine the business case.
27. Connected TVs and many PVRs will have sufficient basic capabilities to allow them to support a spoken interface. A number of products currently on the market demonstrate that spoken interfaces can be offered in competitive, mainstream products. Nevertheless, these tend to be high end products where the additional cost of implementing the spoken interface can be justified and the product sold at an acceptable margin to generate a return on the investment.
28. While it is tempting to think that the model of using a secondary device (such as a smartphone or tablet) to provide the spoken interface could solve the question for low end devices with only basic capabilities, this is unlikely to be the case. For a receiver to be able to support external devices, it needs to provide a network and application stack, physical and logical interfaces and must be running software layers and APIs to enable the secondary device to communicate and control the receiver. Those capabilities require hardware and software not available in low end devices and which would substantially drive up costs of these low end devices, pushing them effectively out of their markets.
29. There remains one model based on secondary devices that does not involve additional capabilities or costs in low end products, namely where the spoken EPG (and possibly other functions) is delivered by a secondary device that impersonates an infrared signal, i.e. acts as a the infrared remote control for the

(low end) receiver. For instance, there exist tablets that incorporate an infrared sender which can be used as a remote control. Such a setup could offer a spoken interface to even a low end device without any additional cost on the receiver. However, the consumer would still need to buy an expensive tablet or other device (even if this of course could be used for other purposes). Moreover, infrared senders in smartphones or tablets are uncommon in the market and even decreasing.

30. Thus, for some time to come, spoken interfaces integrated into receivers will remain a feature for mid to high end equipment only, even if it can be expected that the range of products where such features are available will, over time, increase.
31. In the end, the issue that some disabled consumers may not in fact be able to afford the higher end equipment, or the specialised additional tools, that could offer them a more accessible experience is not one that can be solved by manufacturers and/or service providers alone.
32. As stated above, as time passes features presently currently found only in high end devices may, under influence of Moore's Law, evolving manufacturing processes and other factors, become available on cheaper devices. Other than in that specific way, the issue of affordability and who should pay is really a matter for social policy rather than something that receiver manufacturers and service providers can solve.

Q4. Do respondents agree with Ofcom's initial assessment that the speaking EPGs integrated into TVs and set top boxes may be easier for people with visual impairments to use than touch-screen apps? If not, why not?

33. I don't believe this question can be answered in a general manner as this ultimately depends very much on how well either solution is designed and implemented.
34. However, it is correct that *at present* integrated spoken interfaces are usually offering a better, more complete and more satisfying user experience. Quite obviously, where the spoken interface is integral to the receiver and designed by the manufacturer to be fully integrated and consistent with the rest of the interface, this can result in a very successful user experience.
35. It is also true that touch screens create specific accessibility barriers to blind users (as well as some other categories of disabled viewers), such as the lack of tactile feedback and the need for hand/eye coordination. Nevertheless, recent touch screen technologies have sought to work around the inherent problems that touch screens pose to people with sight loss through a range of features, including voice-over. This has meant that there are an increasing number of enthusiastic users of smartphones and tablets amongst blind and partially sighted people.
36. A well designed mobile app that has full access to the receiver for command and control and that can query the receiver to obtain detailed state and other information (including an EPG) could theoretically provide a level of sophistication and personalised control that would be hard to achieve using a default built-in interface.
37. If today it is still the case that built-in spoken interfaces tend to offer an overall better user experience, this is in part due to the observations I have made elsewhere in this response with regard to the constraints that developers, in

particular third party app developers, face when interacting with television receivers: the limited functionality exposed by protocols (which are not designed for accessibility purposes), the lack of documentation, the confidential and sometimes shielded aspects of hardware and software layers for communication and the fragmented nature of such solutions (both horizontally across the market as vertically between product generations).

38. A spoken EPG solution should be fairly tightly integrated with the television receiver for an optimal user experience.
39. There are some solutions conceivable (and indeed already in use) whereby a spoken EPG solution obtains EPG data independent from the television receiver (for instance from, often freely accessible, web resources), which can create discrepancies between the EPG presented to the user and the guide as stored by the receiver. Local programming in particular can be difficult to deal with in this scenario.
40. There are also solutions whereby the decoupling goes even further and the spoken interface obtains independent EPG data while operating the television receiver through an API or via infrared, but no state or other information is received back from the television receiver. This tends to result in a much lesser quality user experience and both sides of the solution (television receiver and spoken interface) can easily become out of sync or stop collaborating altogether (for example when the receiver has entered a modal dialogue while the spoken interface is unaware of this and/or unable to respond).
41. Finally, and as mentioned elsewhere, in the same way that there is no such thing as “the” average user, but instead different consumers have different abilities and preferences and use products in their own way, not all people with sight loss have the same profile of abilities and preferences. What ultimately is the best solution for an individual will differ from one person to another. Thus, there will remain a need for both types of solution, for the foreseeable future.

42. Q5. Do pay TV service providers such as Sky, Virgin, Talk Talk and BT TV see additional obstacles that would prevent them from committing to including text to speech capabilities in the next planned upgrades to the receivers they offer to subscribers? If so, what are these obstacles? Absent regulation, would these obstacles make it impossible on commercial grounds to commit to the necessary investment?

43. Adding features to existing technologies is almost never entirely problem- and/or cost-free. However, compared to other market players, pay TV service providers are in a much better position to influence the availability of spoken interfaces in their future products.
44. The starting point for all pay TV platforms’ roadmap should be that as their receivers evolve and are upgraded, proper consideration is given to making spoken EPGs (and ideally a fully spoken interface) available to consumers.
45. However, all too often in my discussions with such providers is it apparent that this possibility is not considered, even where it may be technically and economically viable.
46. Critically, most pay TV providers already operate on the basis of receiver technologies that have sufficient processing power, memory, storage and other capabilities necessary to support a spoken EPG and/or user interface. This

contrasts with the wider, horizontal market where many, especially lower end, products do not have sufficient capabilities.

47. For many pay TV subscribers, the ability to recover the development cost of a spoken solution in a future version of their receiver platform is genuine.
48. Because of the differences between platforms, it may be the case that for some pay TV providers offering an external spoken solution in the form of an accessible app on a smartphone or tablet is more feasible initially than a built-in spoken solution. Where that is the case, a pragmatic view should be taken that acknowledges the practical and economic realities that determine which solution (built-in or external) will be (first) made available for their platform.
49. Over time, it should be part of the planned platform development strategy to support both forms of solution for all major pay TV service providers.
50. Specialised, even niche, market opportunities with limited scope for reach are not by definition economically unfeasible. It is possible to produce solutions for smaller audiences at a reasonable cost and realising acceptable profits for industry. However, there is often a mismatch between the perceived market potential of specialised solutions or features in products versus realised value. Demographics and other aspects of market appeal can be overstated by user groups, leading to unrealistic expectations at the manufacturing side. It is in both parties' interest that business cases are not overstated and that all parties are aware at the start of the real numbers involved (in terms of user base, implementation cost, revenue potential).

Q6. If the cost of providing speech-enabled receivers to all those who subscribe to particular pay TV services would entail a substantial delay to the roll-out of such receivers to all subscribers, would it be feasible, quicker and more cost-effective to offer suitable equipment first to viewers with visual impairments?

51. I have no objection to this approach in principle, but I am unsure of how relevant it is in practice.
52. Where specialised receivers would be purpose-built for these users, perhaps using some form of subsidy scheme if one were available, the concept of targeting those users who need the solution (most) is sensible.
53. Similarly, where the cost and/or time involved in deploying a spoken interface to all consumers of a service or product would be substantial, it is perfectly acceptable to prioritise users where the need is highest.
54. Nevertheless, I am not convinced the above scenarios apply very widely, especially in the case of pay TV services.
55. The main cost and effort will be on development of the solution (either built-in or external). In contrast, once a solution exists and has been fully tested, deployment (other than where physical hardware needs to be shipped) is relatively straightforward and inexpensive (assuming such deployment is part of a wider upgrade package rather than just for a spoken solution).
56. It may be the case that the spoken solution (or the necessary interfaces to drive an external solution) requires upgrading or replacing the physical consumer equipment. Where that is the case, it is clearly reasonable to prioritise those consumers in greatest need.
57. A special case may be where an alternative product exists on the market that has the necessary spoken interface, but where this product is not normally available

for a given pay TV provider. It could be considered to make the alternative product available for users with sight loss and others who rely on the spoken interface, while not offering this to other audiences. One should, however, be cautious in that this may create significant future problems: as the mainstream platform of the provider evolves, they now have to maintain interoperability and functionality with a different solution as well, across all the features of their service. It may therefore be better in the longer term to provide spoken solutions as part of their standard receiver offering and/or apps.

Q7. Do respondents consider that it would be reasonable to expect visually-impaired viewers to pay extra for equipment that allows them to use EPGs or substitutes for the same purposes as sighted viewers? If so, why?

58. See my response to question 3.

Q8. Do licensors such as Freesat and Freeview see obstacles to using their leverage to require manufacturers to incorporate speaking EPGs in future versions of their branded products, such as Freetime and Freeview Connect?

59. Theoretically, it would be possible to define spoken functionality as a requirement for compliance with a trademark licence. However, in practice this seems unrealistic as it would substantially reduce the available equipment for these platforms and as such undermine fundamentally the free market business model, while substantially restricting choice (and likely increasing costs) for mainstream consumers.

60. At the very least platforms like Freesat and Freeview could stimulate manufacturers to offer more spoken solutions as part of their range, even if they should, ultimately, not mandate this.

61. Where manufacturers wish to offer spoken interfaces in their products for these platforms, the licensors could require them to develop solutions to be compliant with IEC 62731:2013, an industry developed standard for text to speech in digital television products. Manufacturers would still have the choice as to whether or not to offer a spoken interface and in which one of their products/ranges, but where they include a spoken solution it should be compliant with this industry standard. That would be beneficial to users too in terms of aptness and consistency of the talking features.

62. Connected platforms already rely on (a) programming interface layer(s) for much of their interactive and on-demand features. As such, licensors must design their protocols in such a way that they can not only be used for mainstream apps, but also for accessibility related applications, including spoken interfaces.

63. As stated in the response to question 1, provisions for external connectivity are still lacking in many respects as far as their potential for driving accessibility solutions is concerned. Many of these current shortcomings could be addressed in specifications for platforms such as Freeview Connect. They should define well documented, standardised APIs that are designed to be feature-complete with regard to accessibility applications (either running on the receiver or externally connected).

64. The Consumer Equipment market in general, and the digital television one in particular, form a difficult economic arena for industry. Many brands have reduced their involvement and offerings, or withdrawn entirely from this segment. Some companies have ceased trading altogether. The business models today

are very different from what they were even just 10 years ago. Margins, especially for low to medium end devices, are very thin. The market itself is highly commoditised.

65. Much of the UK's engineering base in television products has also been scaled down or disappeared and much equipment for the UK market is now manufactured outside the EU. Markets are no longer UK focused, but global and using global technical standards. Furthermore, real added value is increasingly shifting from equipment to services and content. These are all constraining factors that should be taken into account.

Q9. What are the main types of cost that pay TV service providers would face in incorporating speaking EPG features into the next generation of their set top boxes?

66. This is a commercially sensitive topic, so I am unable to provide a very detailed response to this question.
67. Furthermore, the challenges and opportunities for implementing spoken EPGs and other talking features into receivers are multifarious and can vary considerably from one manufacturer or service provider to another.
68. Due to the nature of Connected TV, CTV receivers are generally more suited for incorporating spoken EPGs and other TTS based features.
69. The best TTS engines incur a licence cost that is not entirely trivial (although feasible for higher end products). There are however acceptable low cost alternatives.
70. As with all new features, refactoring of existing software to accommodate the spoken features can require substantial developer time. The best way to mitigate this is to incorporate such changes within already planned broader software developments.
71. Pay TV service providers are in a good position to deliver a spoken solution within their existing roadmaps, recovering costs over some time through their subscription base.
72. As external connectivity (with mainstream apps for command and control) are already part of existing roadmaps (and indeed already present in many receivers currently deployed in this market), this offers potential for externally connected spoken solutions where built-in features may not be immediately attainable, or alongside built-in features.
73. As a general principle: having limited functionality is better than having no functionality at all, but the availability of initial, partial solutions should not mean that complete and properly integrated solutions should not be developed.

Q10. What is the scope for connected platforms to avoid the need for specific TTS provision within consumer equipment by using cloud-based resources (e.g. speech files on a central server delivered to the device as required)?

74. While it is technically possible to delegate the TTS process to an external entity (such as cloud based service), it is not clear to me what architectural or economic benefit this would have. In comparison with automatic speech recognition, the processing and memory requirement for TTS are modest.
75. As such, since delegating the TTS would still require substantial other logic in the local device or app (for controlling the process, handling network requests and rendering the results) and added functionality for managing delegation otherwise

not needed, I am not convinced this is necessarily a judicious model or that there is much, if anything, to be gained from it.

76. This is unlike Automatic Speech Recognition (ASR) where a cloud based architecture makes more sense, especially where the ASR results need considerable further processing for parsing a query, performing a search and other computationally intensive operations, before presenting results or feedback to the user.
77. In relation to the specific wording of the question, it should also be observed that TTS should not be understood in this context as simply playing out pre-recorded phrases. While play out of pre-recorded prompts requires only very basic capabilities in receivers (and indeed could be done by even very low end devices), the almost unconstrained language domain as represented by EPGs cannot realistically and cost-effectively be delivered on the basis of pre-recorded prompts. It requires on-the-fly speech synthesis through a TTS engine. The combination of pre-recorded prompts with TTS content does not necessarily provide a more acceptable user experience either.
78. While it could be argued that a cloud TTS service could use TTS to generate on the fly spoken output that is streamed back to a receiver and treated as a simple audio file play-out scenario by the receiver, the same problems as mentioned above around the additional logic needed in the receiver to manage all of this still applies. Such a receiver also needs extensive IP networking capabilities for using the cloud based service (and a connection to the home network), which would be incongruous with the notion of a simple, low-end device.
79. This model also introduces further considerations around who is going to operate the cloud service and what the on-going costs of such an operation would be.
80. Avoiding the need for built-in TTS is best achieved through a programming interface offering external connectivity, exposing all the necessary command and control functionality in the receiver as well as state, EPG and other information, through a standardised interface that can be used by apps or other external solutions.
81. As stated elsewhere, Connected TV platforms are generally already well capable of supporting spoken output in terms of processing power, memory and other capabilities.
82. It is unclear that licence costs related to TTS would necessarily be meaningfully lower when delivered through a cloud based TTS engine.
83. Using a cloud-based delivery model also means that if the connection to the cloud service is unavailable (or indeed the receiver is not connected to the home network, which represents a substantial proportion of all Connected TVs in the UK today), the spoken interface does not work.

*Guido Gybels,
7 September 2014.*