UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
UNECE Expert Meeting on Statistical Data Collection – Towards a New Normal?
26 to 28 October 2022, Rome, Italy

Session 2
19 September 2022

Automated assistance to respondents: the case of the Agricultural Census Measurement Error Survey

Claudia Fabi (Istat)
claudia.fabi@istat.it

Abstract

Between January and April 2022, five months after the end of the data collection process for the 7th General Census of Agriculture, Istat has projected and realized a "Measurement error survey" with the aim of evaluating the measurement error committed during the Census operation, regarding a subset of variables among those collected during the Census itself. In order to offer to the farmers a simple way to participate to the Measurement error survey, the toll free number was setted with an AI automated assistant, able to schedule appointments with the respondents. This work will show how the AI assistant worked during the four months of the data collection period, how many calls were received and when, response time and efficiency; analyze the capacity of the virtual assistant to fulfill respondents' expectations, how many appointments were fixed and their tipology, how many calls were abandoned by the respondent before the end; note the AI evolution within the four months of service, its ability to modify itself and to learn, from experience, how to better fulfill respondents' requests; compare the outcomes of a entirely "human" inbound service with those provided by the AI, for the same Agricultural Census.

UNECE 22 – Topic 1: "Automated Data Collection"

Automated assistance to respondents: the case of the Agricultural Census Measurement Error Survey

Claudia Fabi ISTAT, Rome, Italy – claudia.fabi@istat.it

Index

1.1. The context: the Agricultural Census Measurement Error Survey

Nearly five months after the end of the survey of the seventh General Agricultural Census, which took place between January and July 2021, Istat designed and implemented a Measurement Error Survey, with the aim of estimating the extent of the measurement error committed during data collection operations.
As frequently happens for Measurement Error Surveys, also in this case, the methodology has planned the repetition of a subset of interviews, with particular reference to agricultural and livestock companies that during the Census were defined as active. The Measurement Error Survey questionnaire was designed to be filled in only with the CATI technique, even for farms that during the Census had preferred to participate with another data collection technique (CAWI or CAPI).
This choice necessarily entailed the restriction of the field of observation only to the units for which there was at least one telephone number available, deriving from the census list or provided by the respondent during the Census, but, on the other hand, it allowed a greater speed of execution and a high degree of optimization of the support and management activities already performed in the previous months, during the Census.

In particular, about 135,000 units were extracted from the subset of active farms at the reference date of the Census, with a stratification that took into account the characteristics of the farms, their geographical location and the survey technique adopted during the Census. A questionnaire was therefore prepared, digitally developed by an external supplier, who also provided a network of nearly 100 CATI interviewers who were already experts on similar issues, for having been involved in the inbound and outbound CATI data collection process of the census survey.

The same supplier also managed, in support of the Measurement Error Survey, the toll-free number, which was used exclusively to offer to all respondents the possibility of making an appointment for the interview at their convenience, or to leave a telephone number to be called back, even if different from those contained in the census list. The toll-free number was not expected to also provide assistance on the questionnaire or information on the survey: this task was delegated to an email address managed by back office operators.

A BOT was therefore designed and implemented for the toll-free number service intended to fulfill the schedule of appointments activity, one of the first examples, in Istat, of the application of a similar technology to the context of a statistical survey. In the following, the main aspects and characteristics of the BOT activated between 26th January and 29th April 2022 will be explored. Furthermore, the most important results of the support activity offered through BOT will be illustrated, with the aim of proposing an initial evaluation and a comparative analysis between the services offered by an AI and a traditional service manned by human operators.

1.2. The BOT in the public assistance services

In public assistance services, provided both by private companies and public authorities, ChatBOTs as well as the most recent VoiceBOTs are increasingly widespread. Some of the latter are in - almost - all of our devices, starting with Siri (Apple) to continue with Cortana (Microsoft), Google Assistant, up to Alexa (Amazon), generating billions of interactions on an annual basis.

Beyond the most famous names, ChatBOTs and VoiceBOTs are now consolidated even within the main public support tools, such as the websites of companies operating in e-commerce markets, that allow their actual or potential customers to request information and services through a virtual assistant.

While, therefore, it is possible to find a basic literature regarding the BOTs used in e-commerce (literature however often offered by trade associations, sector analysts and stakeholders in the AI market), the study and analysis of the performance of a virtual assistance service to support statistical surveys, in particular official statistics, are not yet widespread.

But what is, actually, a VoiceBOT?
The most ancient telephone keypad guided data acquisition systems, which everyone has often experienced when dealing with branched and complex assistance services, can be considered the ancestors of today's VoiceBOTs. The term VoiceBOT combines the two terms "voice" and "robot", although VoiceBOTs have nothing of the appearance that we commonly attribute to an advanced robot. They are in fact powerful software, capable to understand human natural language and to respond in a relevant way to the questions that are asked for or, if necessary, to ask questions themselves to better understand clients' requests.

VoiceBOTs are based on Natural Language Processing (NLP) platforms with the addition, compared to ChatBOTs, of Automatic Speech Recognition (ASR) systems, i.e. AI techniques and algorithms for the automatic processing of information in natural language[1]. Systems like VoiceBOTs are able to learn from experience, refining and improving their understanding the more they deal with real requests from users. They are able to overcome interaction difficulties deriving from the use of jargons and dialectal terms, imperfections in the pronunciation of words or incompleteness of the requests that are addressed to them, becoming able to interact with the caller, through the appropriate questions, asking for further clarifications necessary to provide the requested service.

Therefore, the VoiceBOTs are designed to replace, at least partially, an assistance service made up of human operators, acting almost as a filter for the fulfill of less complex information requests and delegating to the traditional second level service the requests that need a more structured intervention.

1.3. The adaptive develop of the BOT for the Measurement Error Survey

In the case of the Measurement Error Survey, it was not necessary to set up a particularly complex reception system for the toll-free number, since the channel was only intended to provide the respondent with an additional tool to request to be interviewed on a more suitable day/time, eventually specifying a preferential telephone number.

Therefore, the structure of the BOT focused above all on two aspects:

- the ability to identify the caller, associating him with the corresponding record within the sample of farms selected for the survey;
- the ability to recognize and memorize dates and times proposed by the caller, and after that, interacting with the CATI software in order to schedule a priority contact (in the case of an immediate interview request) or an ordinary appointment, for outbound interviewers.

1.3.1. Identification of the calling farm

Each farm included in the census list, since the start of the Agriculture Census, has been associated with a 13-digit alphanumeric code, which has been the unique identification code for each unit throughout all the census survey. This unique code was also reported in the postal communications sent to respondents, so that the farms involved in the census were also aware of their personal code

[1] Eudata Blog: https://blog.eudata.com/chatbot/voicebot-cosa-sono-e-come-si-realizzano-nella-pratica/

eventually using it for self compilation of the questionnaire (CAWI) or provide it to the toll-free number operators, thus obtaining a more personalized assistance.

The same code was also adopted for the Measurement Error Survey, thus simplifying the subsequent processing procedures.

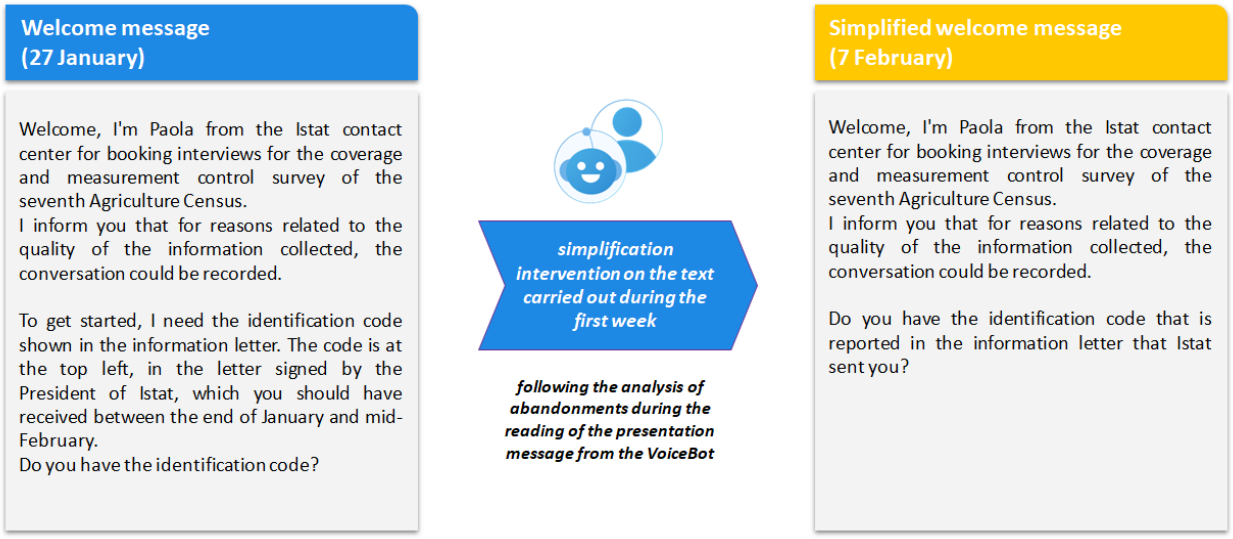
The structure of the code was setted as follows:

Letter identifying the initial DC technique	Istat municipality code	Letter identifying the farm's structure	Progressive number of the unit within the municipality
T	000000	U	00000
P	000000	M	00000

so a farm identifier code could have been, for example: T058060U00123.

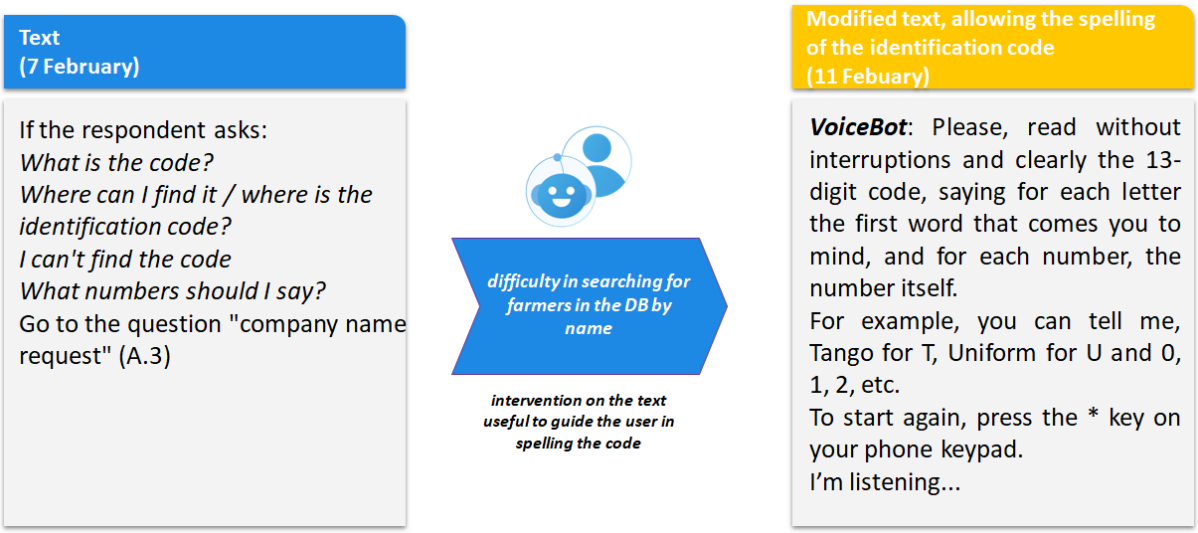
It is certainly not the simplest code for a BOT recognition, and, moreover, it was not initially designed for this purpose. The programming of the BOT therefore had to focus in particular on the most effective way to make the caller pronounce his own identification code, maximizing the probability of identifying it correctly even in the case of incomplete or abbreviated readings.

Since the start of the service, it was clear in fact that the way in which the BOT should have expressed itself needed a simplification, to make it more compliant with a natural language. Some changes to the script were adopted in the first days of the service. Below are the most important, relating to the approach to the caller and his identification.

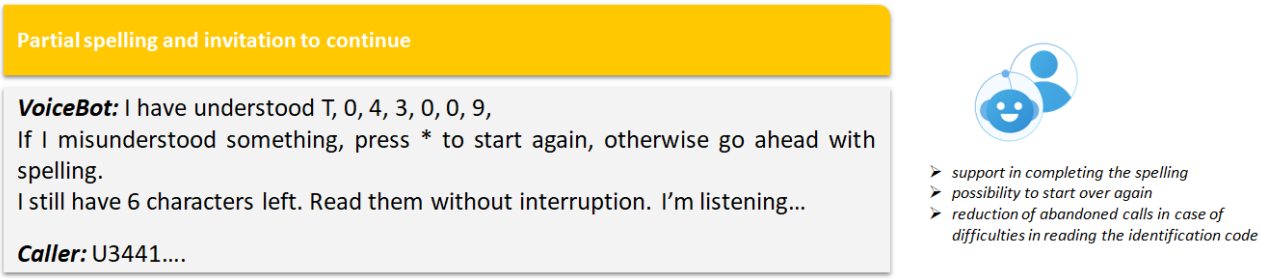


The most critical issue has been, predictably, the way in which the caller pronounced its own code, mainly due to the possible misunderstanding of the initial letters (the pronunciation of the consonants T and P can be misleading if listened to by phone). An attempt was therefore made to use the technique of spelling, allowing the BOT to also process codes dictated with words starting for the specific consonant and/or groups of numbers together.

In particular:



The BOT has been developed to be more interactive and able to ask to the caller whether or not it has really understood the right identification code, repeating it to the caller.



In this way, the BOT has progressively evolved itself offering support to the dictation of the identification code, possibly eliminating what has been received so far, if incorrect. This evolution reduced the use of the alternative dictation of the extended name of the farm or farmer, a more ambiguous identification method than a unique code, that has been left in the script as a last resource.

1.3.2. Appointment scheduling

The scheduling of appointments was much easier. In fact, the script provided a choice option, between two possibilities:

- to be called back within 10 minutes by an outbound interviewer, if on duty;
- to choose a day and a time within those possible, in which to be called[2].

In this case, natural language is easily identifiable by a BOT even when the caller uses approximations or other phrases to identify a date, for example by saying the words "tomorrow" or "the day after tomorrow" or "next Wednesday".
The BOT was immediately able to register the proposed appointments, requesting confirmation to the caller by repeating the date and time it has understood.
The caller, in this phase, could also indicate a telephone number to be called back, dictating it to the BOT. For this type of interaction no particular difficulties emerged, except for the interruption of the dictation by the caller, which made the developers reflect on the opportunity to also allow the

[2] Appointments or requests for immediate interviews could only be fixed on the days and times in which the outbound interviewers were active, therefore from Monday to Friday, from 9.00 to 21.00 and on Saturdays from 10.00 to 19.00 with the exclusion of national holidays.

typing of the telephone number via the numeric keypad of the callers' device. However, considering the greater speed of interaction via voice, it was decided to avoid any typing of keys.

1.3.3. Evolution and efficiency

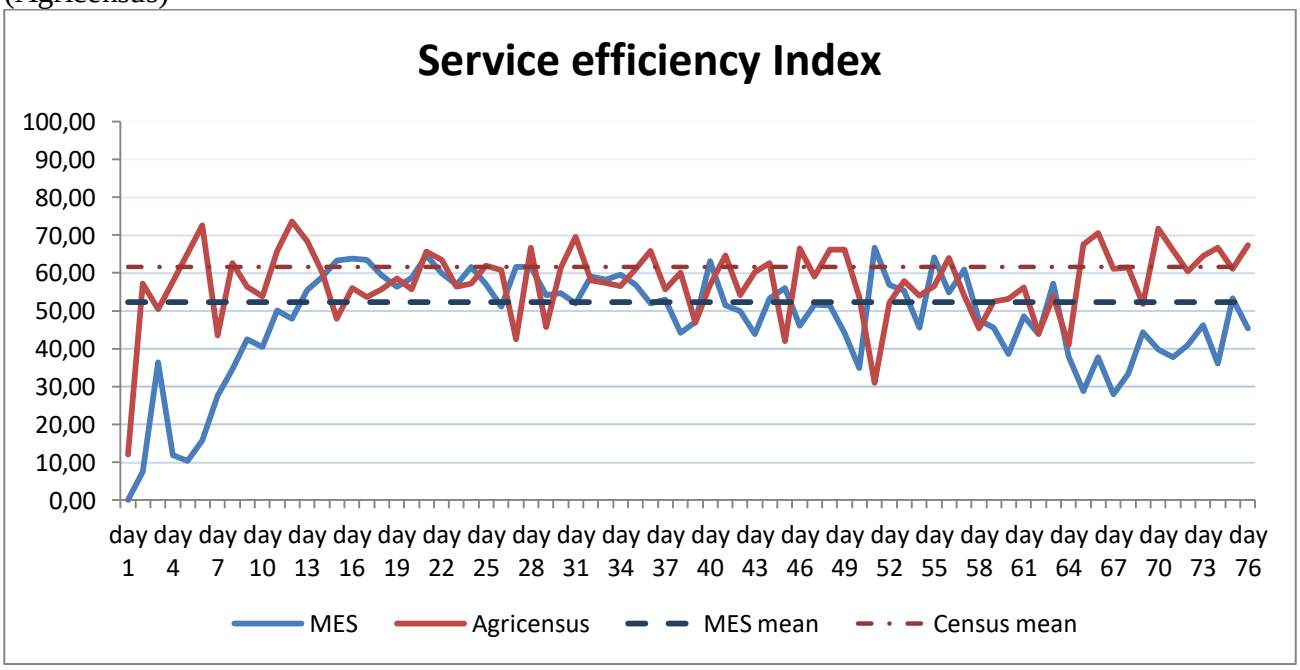
Although it is evident that a period of evolution and learning is inherent to the way in which a VoiceBOT works, it is advisable to try to evaluate the efficiency and speed with which the VoiceBOT stabilizes itself in its most effective form, trying to infer its efficiency from the traffic curves and the outcome of the calls handled.
The following graph 1.1 illustrates a possible efficiency indicator of the Toll Free Number service, processed both for the Measurement Error Survey (MES) and for the Agricultural Census (Agricensus). For the Census, the calculation was limited to the branch of IVR intended for booking interviews, in order to maintain the analogy with the operating methods of the BOT of the Measurement Error Survey.

Efficiency Index was calculated, for each day *t*, on the basis of the summary outcome of the calls received, using the following formula, reported in percentage values:

$$EffIndex_t = \frac{Satisfied\ requests\ for\ immediate\ interview + Satisfied\ requests\ for\ appointment}{Total\ handled\ calls}$$

in which the numerator represents the number of calls received, whose relevant request has been satisfied by the BOT (for MES) or by the operator (for Agricensus), while the denominator includes all the calls received, including those abandoned in the queue, incorrect or not managed.

Graph 1.1 – Trend of the efficiency index for VoiceBOT (MES) and traditional toll-free number (Agricensus)



While representing a very rough efficiency indicator, it is already evident that the efficiency curve exceeds the respective average value during the 13th day of service for the VoiceBOT with a progressive increase in efficiency throughout the first decade of operation. On the contrary, the service offered by human operators has a much steeper efficiency curve, settling from the first days

on values around the average of the observed period. It is reasonable to assume, therefore, that the BOT needed to evolve and improve over a period of about 10-15 days[3].
After the first evolutionary decade, the efficiency indicator no longer highlights differences between the two types of service. It should be noted that the comparison was truncated to the first 76 days of service of the toll-free number of the Census of Agriculture which lasted even longer, for a total of 171 days of activity.

A further indicator of efficiency is given by the First Call Resolution. In a traditional call center, the FCR is the percentage of requests processed directly online, i.e. without further interactions with the customer and without referrals to level II (if any).
On the basis of the available data, an attempt was made to elaborate an indicator of a similar type, however, considering only the calls actually resolved, for which the identification of the caller was available with certainty (thus allowing to distinguish those who called only once from those who have called several times) and, at the same time, that the call was relevant to the service offered.
The following table shows the FCR index referred to both the Measurement Survey and the Toll-Free Number of the Census of Agriculture: the indicator is slightly more efficient for the traditional Contact Center than the VoiceBOT. The small entity of the difference, also if it's statistically significant, is in itself surprising, considering that a VoiceBOT has the ability to relaunch and dialogue with the caller limited to its own script and capacity, compared to a human operator who can interact longer with the caller to understand and concretely satisfy his request.

	Requests resolved on the first call	**Total resolved requests**	**FCR Index[4]**
MES	4,494	5,569	80.7%
AgriCensus	47,737	58,285	81.9%

It should be borne in mind that in both cases, the resolution during the first call was conditioned not only by the performance of the service itself, but also by the efficiency of the outbound callback service, for scheduled appointments. In fact, a user who, despite having made an appointment, is not called back as agreed, can contact the booking service again, even if the latter has already concretely solved his need. However, since the outbound Contact Center was substantially the same for both surveys, it can be assumed that it did not significantly influence the above results.

1.4. The service offered by VoiceBOT: characteristics and peculiarities

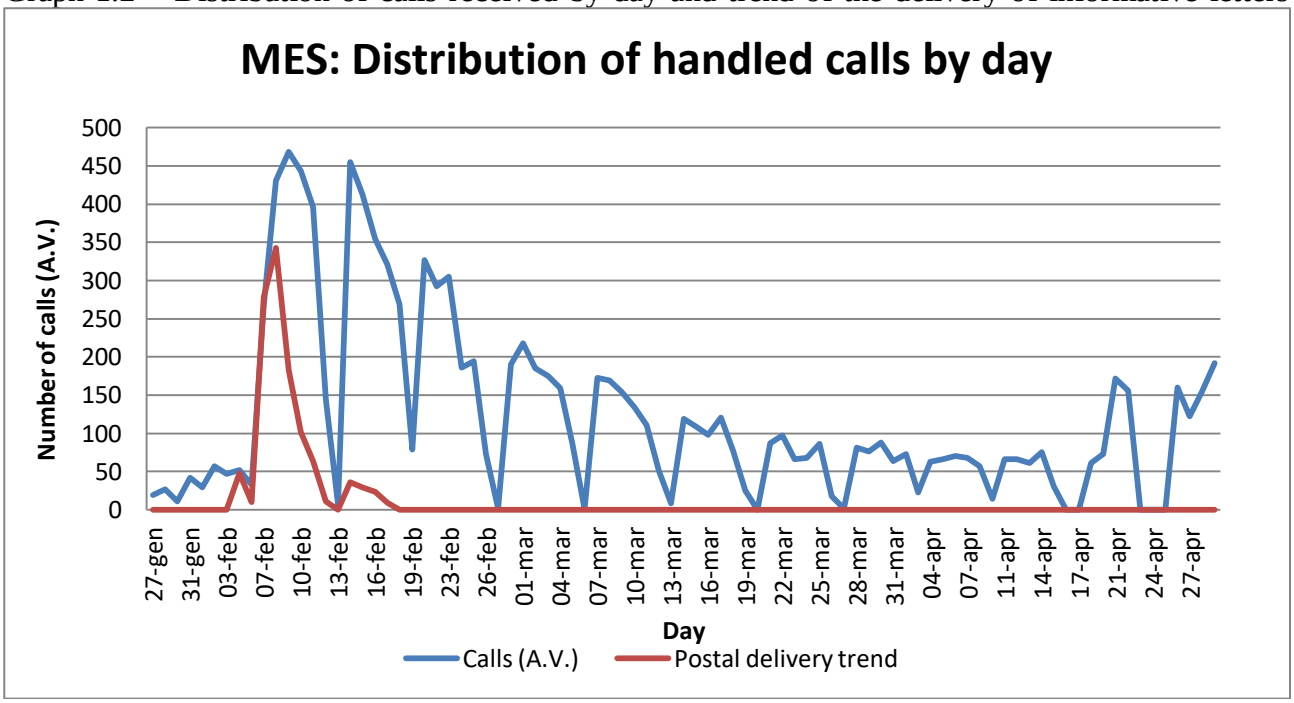
1.4.1. Call distribution and curves

During the 76 days of the VoiceBOT service, the distribution of received calls followed a trend very similar to that of any public assistance service. The peaks were recorded close to the delivery of informative letters to the farmers involved in the Measurement Error Survey, progressively decreasing in intensity until the period near the deadline for participation in the survey, in which it increased again. Graph 1.2 represents the calls received per day and has the characteristic sawtooth shape, in which the minimums correspond to the Sundays and public holidays in which the BOT

[3] It is important to underline that, especially in the first days, the BOT could not count on a significant influx of calls, so it is possible that the learning curve was slower, also due to the impossibility of having a sufficient number of cases on to gain experience. In this regard, see the distribution of calls received during the service period, Graph 1.2, paragraph 1.4.1.
[4] The difference between the two percentages is statistically significant (Chi-squared test) with a prob. level of 5%.

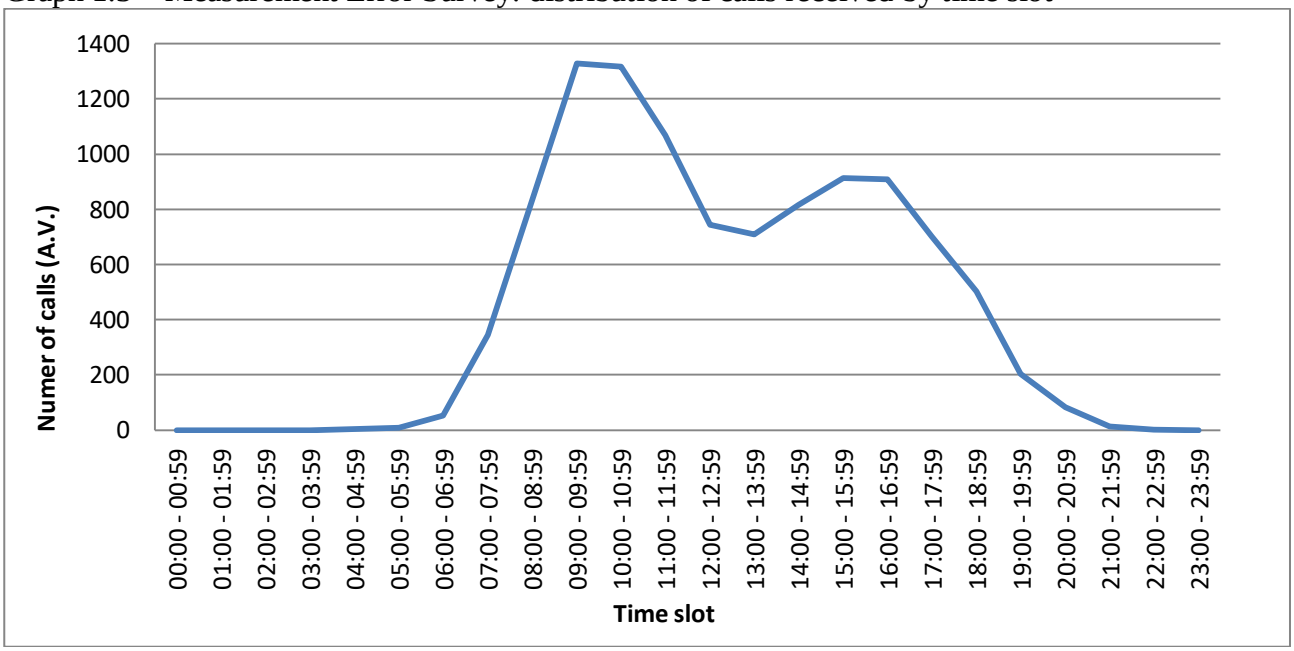
while remaining operational, offering the possibility of making an appointment for the interview but without being able to guarantee an immediate interview[5], received a rather small amount of calls.

Graph 1.2 – Distribution of calls received by day and trend of the delivery of informative letters



There are also no substantial differences in the distribution of calls by time slot, as can be seen from graph 1.3 which again traces a typical calls curve for a public assistance service.

Graph 1.3 – Measurement Error Survey: distribution of calls received by time slot

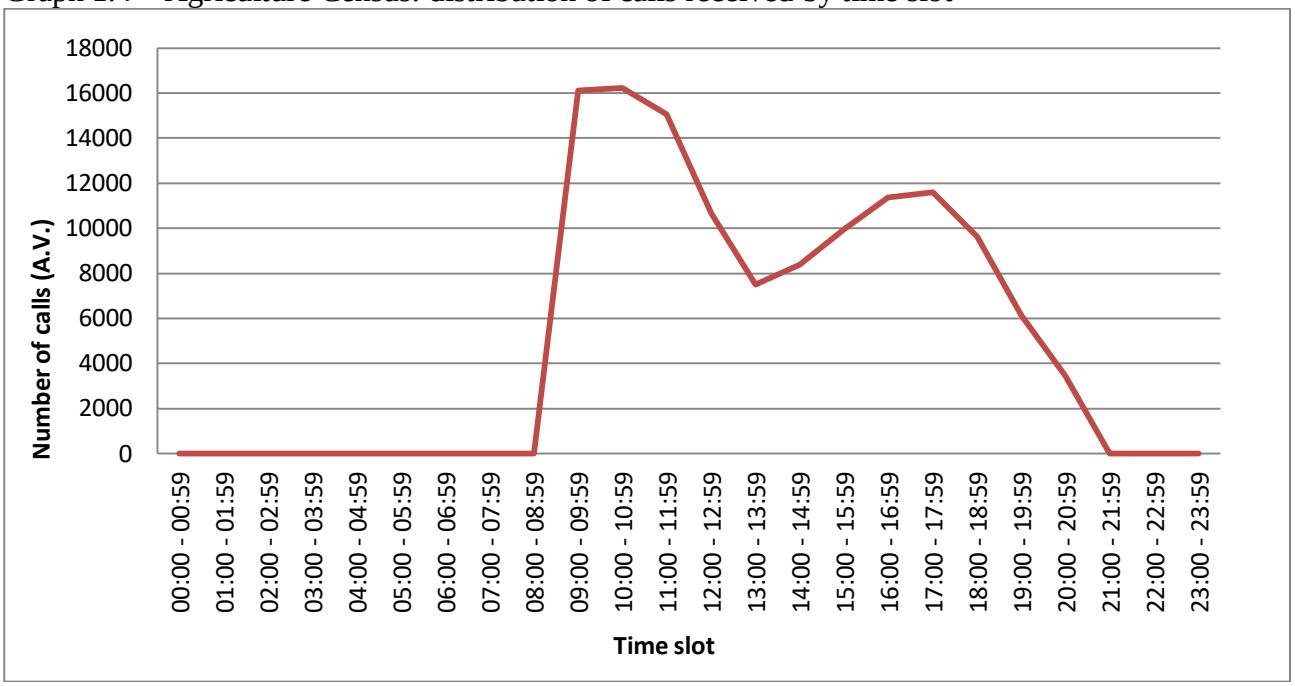


On the other hand, the slight difference with the same curve extrapolated from the distribution of calls received for the Agriculture Census is interesting (graph 1.4). It is possible to note how the

[5] The outbound survey was active only from Monday to Friday, from 9.00 to 21.00 and on Saturdays from 10.00 to 19.00 with the exclusion of holidays (see note 2).

limits imposed by the service time managed by the operators make the curve more abrupt in the two queues, presumably cutting out a percentage of calls that can be estimated by calculating the entity of the queues corresponding to the "out of service time" in the distribution of the Measurement Error Survey: 11.9% of the total calls received.

Graph 1.4 – Agriculture Census: distribution of calls received by time slot



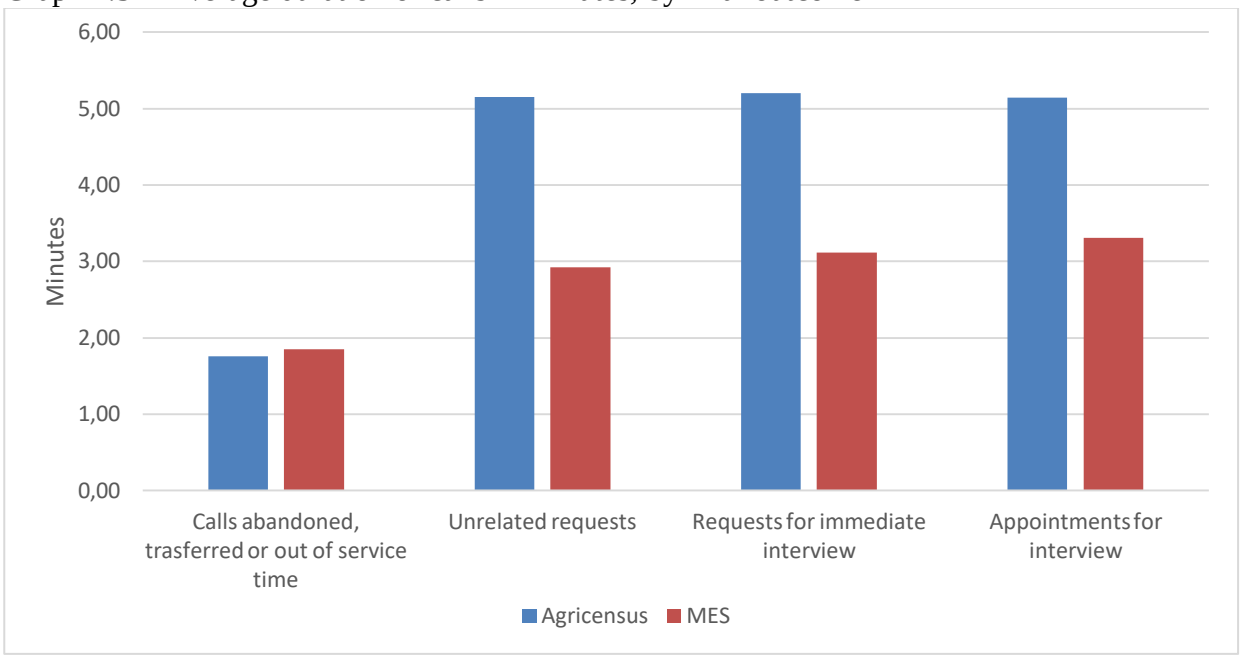
It should also be noted that the use of the BOT allows to evaluate the “natural” curve of an inbound service, without forcing the designer to make assumptions about the most appropriate service time for the target audience before the service is actually provided. In the case of the Agricultural Census, the start of the service managed by operators at 9:00 in the morning was probably too late, compared to the habits of the target audience.
In this sense, the use of a VoiceBOT allows to offer a service more compliant with the natural habits of its public. In general, there are no days or times in which the BOT has reasons to be out of service, while a service managed by human operators has necessary to deal with the organizational difficulties and the costs needed to cover even residual time slots (e.g. from 6:00 to 9:00 in the morning) with an adequate number of operators suitable to users' expectations. The hybrid combination of a BOT and a traditional service would probably be the most efficient way not to cut off a part of the users who, albeit marginal, prefer interaction times concentrated on queues.

1.4.2. Calls duration

Also observing the average duration of the calls received and managed by the BOT and, for comparison, by the service managed by the operators, it is possible to notice some differences.

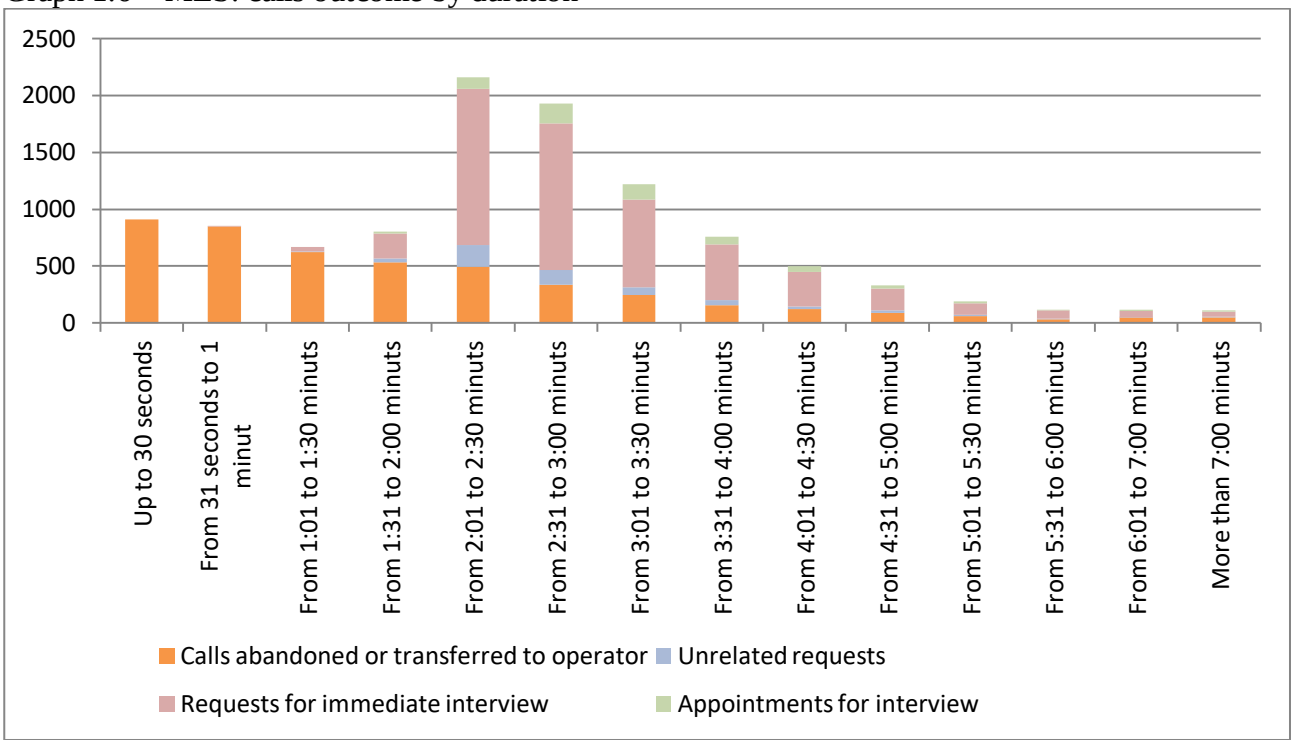
The interaction with the BOT is much shorter, for all types of calls, with slight differences only for abandoned calls, not managed because out of service time (Agricensus) or transferred to the operator (MES). Graph 1.5 shows the differences that, on average, are around 2 minutes even for resolved calls, that is, those ended with an interview appointment or with an immediate interview request.

Graph 1.5 – Average duration of calls in minutes, by final outcome

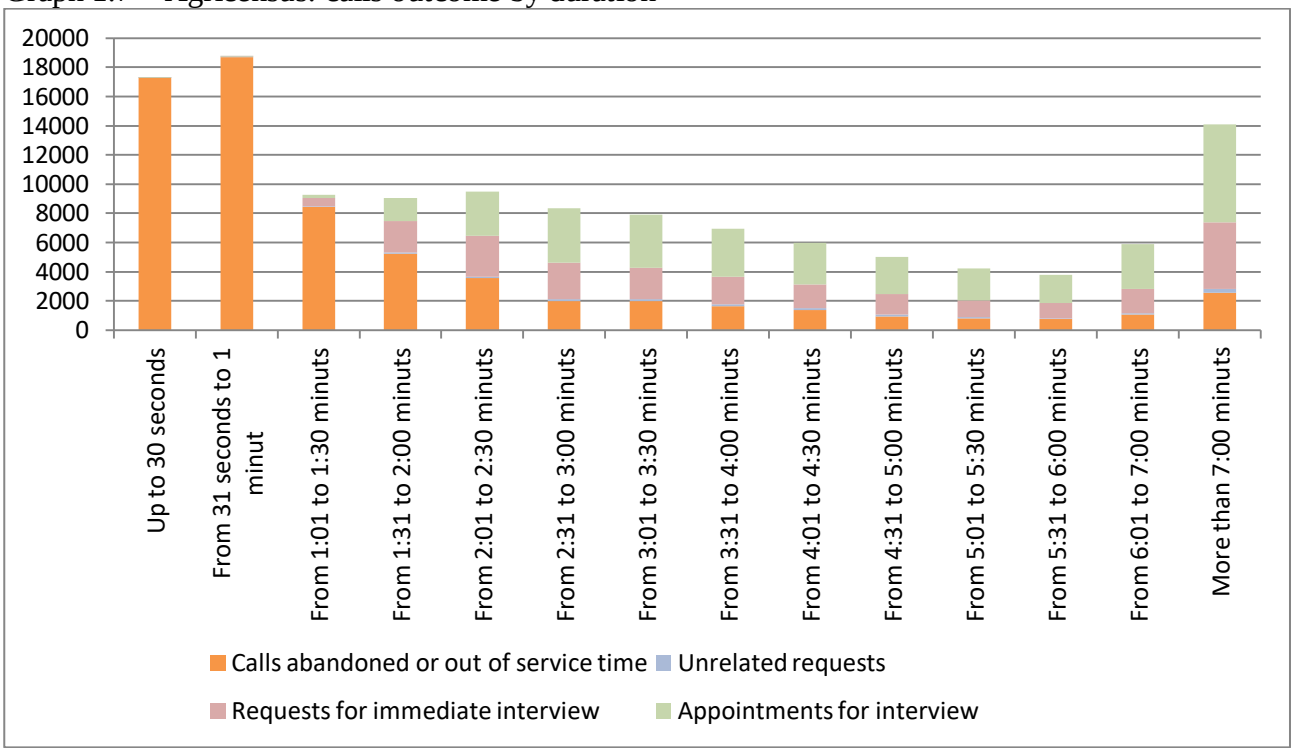


Analyzing in detail the efficiency of the service according to the duration of the call, it is clear that for the BOT an interaction between 2 and 3.5 minutes is sufficient to finalize the request, thus successfully passing through the recognition of the calling unit, for the appointment proposal and finally for the registration of the telephone number to which the farmer intends to be called back.
Graph 1.6 clearly shows where the distribution of resolved calls is concentrated, in terms of the duration of the conversation with the AI. The abandonment curve is concentrated in the first duration classes, while the distribution of resolved calls reaches its peak between 2 and 2.5 minutes. It is the same time that the BOT takes, on average, to understand that the request that is addressed to it is not relevant (unrelated request).

Graph 1.6 – MES: calls outcome by duration



Graph 1.7 – Agricensus: calls outcome by duration



The distribution for the toll-free number of the Census of Agriculture presents a different trend (graph 1.7). There is no longer a clear concentration of calls resolved within a narrow range of durations, but the natural unfolding of human conversation between two individuals ends up making the number of calls resolved along the entire duration axis more widespread, even over 7 minutes in total.

1.5. Summary, conclusions and perspectives

In the preceding paragraphs, an exploratory picture has been provided of what it was possible to detect on the functioning of a VoiceBOT for an assistance service to support the respondents of official statistical surveys.

The following table summarizes the main indicators proposed, computed both for the VoiceBOT and for the operator-assisted toll-free number service.

Indicators	**BOT**	**Traditional CC service**
Efficiency Index	52.3%	61.7%[6]
First Call Resolution	80.7%	81.9%
Average call duration (minutes)	2.58	3.37
Average resolved call duration (minutes)	3.14	5.16

Beyond the list of performance indicators, a further aspect to be evaluated is the savings obtained from the use of a VoiceBOT compared to a traditional service. Here, a very complex work should be made, for evaluating the economic costs sustained for both services typology, in terms of human and technological resources. Unfortunately, many of these data are not available. To provide at least an approximate estimate, it is possible to compute the total number of minutes taken by the

[6] Limited to the first 76 days of service. For the entire duration of the service, the indicator assumes a value of 57.2%.

VoiceBOT to manage the resolved calls, and consider this amount as an estimation of the labour of a traditional inbound operator service which has been saved.

Total number of minutes for resolved calls (MES): 17,476 minutes

This is obviously a very rough evaluation, which does not take into account the additional costs of personnel, their training, turnover, the provision of reporting and output, nor the investment necessary for the supply, preparation and programming of an evolved VoiceBOT.

Looking ahead, it is therefore necessary to start designing a set of indicators that are independent from marketing applications and more suitable for assistance services in support of official statistical surveys. The literature available in this regard is not very extensive, while, on the internet, sites and blogs sponsored by VoiceBOT or AI suppliers are swarming, offering indicators specially oriented to the satisfaction of a commercial clientele and to the ability to remunerate the investment made in a sales channel managed by AI[7].
It would also be interesting to evaluate the degree of user satisfaction with an assistance system managed by AI, by promoting small and rapid Customer Satisfaction surveys, among a sample of service users, similarly to what happens for services managed by traditional operators. This is important for the aim to evaluate not only the efficiency of a BOT service, but also its quality for the users and for the survey it is intended to support.

More consolidated and rich data are also needed: for example the possibility of separating the number of calls transferred to the operator from those abandoned by the caller before the end - which here have been presented only in aggregate form - to understand the share of users who, in any case, cannot or do not agree to complete the interaction with the VoiceBOT. This could be an estimate of the extent of the users which in any case, however perfected and efficient the system managed by AI may be, will always prefer the interview with a human operator.

In general, the VoiceBOT experience appears promising and productive, especially for its use in simple and repetitive services, such as obtaining contact information preparatory to subsequent telephone or in-person interviews, or scheduling appointments. It certainly has the advantage of not having waiting times or queues and of being able to accommodate even extremely high peaks of calls at any time, which often occur in the presence of massive sending of informative letters or reminders to the survey population. Considering how much VoiceBOTs have now made their entry into everyone's daily life, making them increasingly suitable and effective also in support of official statistical surveys will be the challenge that a not too distant future offers us all.

[7] For example: Rachana Chotia, How to Measure the Success of Your Voice AI Bot? - https://verloop.io/blog/measure-voice-ai/

Bibliography

Enuan - https://www.enuan.com/

Eudata Blog - https://blog.eudata.com/

Verloop.io - https://verloop.io/

Voicebot.ai - https://voicebot.ai/

Credits

A special thanks goes to Paola Antenucci and Simona Curto, of Teleperformance KS Italy, for their support about Voicebot performance in MES.