# To Call or to Recall? That's the Research Question

JUAN PABLO CARRASCAL, Universitat Pompeu Fabra
RODRIGO DE OLIVEIRA and MAURO CHERUBINI, Telefonica Research

We present findings of a study with 62 subjects who had 796 of their outgoing mobile phone calls recorded and transcribed for their later annotation—by highlighting important information shared during calls. We found that patterns in these calls (numbers, names, interrogative adverbs), as well as some contextual parameters, are better indicators of annotation needs than the callers' profile or call quality. Callers highlight information in both parties' turns (caller and callee) more often than highlighting solely information provided by the callee, which is mostly due to annotating questions with contextual information for the highlights in the callee's turns. We discuss how this behavior changes according to call purpose. Finally, we found that annotation needs change over time: whereas some annotations might not be considered relevant after weeks, others originally considered irrelevant might become important archival notes. We present implications of these findings for the design of mobile phone annotation tools.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms: Human Factors, Experimentation, Design, Measurement

Additional Key Words and Phrases: Phone call, annotation, context, mobile information, temporal influence

**ACM Reference Format:**
Juan Pablo Carrascal, Rodrigo de Oliveira, and Mauro Cherubini. 2015. To call or to recall? That's the research question. ACM Trans. Comput.-Hum. Interact. 22, 1, Article 4 (March 2015), 30 pages.
DOI: http://dx.doi.org./10.1145/2656211



## 1. INTRODUCTION

Mobile phones are one of the most pervasive personal electronic devices ever made. With a number of mobile subscriptions that approaches the 7 billion [World Bank 2013], these devices are used for various purposes, such as keeping in touch with friends, doing business, or even for emergency situations. Moreover, making mobile phone calls has been part of people's day-to-day life in the past couple of decades. A lot of information is exchanged through phone calls. Whereas a consistent part of this information could be ephemeral as supporting our social needs, another part might be worth remembering as being functional for our lives. For example, we might make a phone call to confirm

This work has been partially funded by the EEE (TIN2011-28308-C03-03) project. Telefonica Research participates in the Torres Quevedo subprogram (MICINN), co-financed by the European Social Fund, for researchers recruitment.
R. de Oliveira is currently affiliated with Google Inc., San Bruno, CA. M. Cherubini is currently affiliated with Google Inc., Zurich, Switzerland.
Authors' addresses: J. P. Carrascal, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Carrer de Tanger, 122-140 Barcelona, 08018, Spain; email: jp.carrascal@upf.edu; R. de Oliveira (current address), Google– YouTube, 901 Cherry Ave, San Bruno, CA 94066; email: oliveirar@google.com; M. Cherubini (current address), Google, Brandschenkestrasse 110, 8002 Zurich, Switzerland; email: mauro.cherubini@acm.org.

DOI: http://dx.doi.org./10.1145/2656211

whether our partner is feeling better, to inquire about which groceries to buy on our way back home, to discuss a specific topic with a work colleague, or to double-check the exact time and place to meet a friend during the weekend. Whereas the first example deals with information that is consumed immediately and does not necessarily require notes to be taken, the other ones suggest the need to archive information for future use.

To date, however, there is little support of applications for annotating this information appropriately. Previous work has focused mostly on investigating how people take notes during work-related meetings [Bothin and Clough 2012; Whittaker et al. 2008; Geyer et al. 2005; Wilcox et al. 1997] and in other general settings [Kleek et al. 2011; Bernstein et al. 2008; Lin et al. 2004]. Indeed, mobile phone calls could be seen as two-person meetings, but they are quite different from formal meeting contexts: calls are usually less planned, less structured, shorter in time, and performed in mobile contexts. Aside from these differences, the cumbersome situation of having hands busy and needing to take notes while on the phone is often experienced by many—a restriction typically not experienced during work meetings. Perhaps the scarce number of note-taking solutions for mobile phone calls is due to the lack of studies that have tackled this issue. In consequence, we have little evidence of what is important to remember during mobile conversations, as well as which factors play a role in this process. Furthermore, it is unknown how phone call annotation needs change over time and which actions should be taken to support such dynamic needs.

In a previous work, we disclosed initial findings related to some of these topics, which pointed to the importance of studying more adapted forms of phone call annotation [Carrascal et al. 2012]. In this article, we expand our previous work by tackling all topics mentioned previously and using a larger dataset with increased statistical power (CL: 95%, MOE: $\pm 3.4\%$). More specifically, we discuss results obtained with a 2-month user study conducted with 62 subjects who had their outgoing phone calls transcribed for annotation by means of highlighting. The analysis of the participants' calls ($N = 796$) and highlighting behavior confirms that there are specific call patterns that are frequently annotated, and that some contextual variables might influence the note-taking behavior.

We observed that callers surprisingly highlight only information on their turns as often as highlighting only information on the callee's turns. We also observed that people's annotation needs and strategies do change across time, and these dynamics should be taken into account when designing mobile phone call annotation tools.

The main contributions of this work include our findings on the following:

(1) People's needs to annotate mobile phone conversations, as supported by a survey conducted with 62 subjects
(2) Factors that seem to influence the note-taking activity, as supported by the analysis of 2-month user study data in which participants (i) highlighted parts of the transcribed mobile phone calls that they considered to be important and (ii) provided contextual details for each call by answering a postcall questionnaire
(3) The frequency and circumstances that each party in a phone call—caller and callee—provide information considered to be relevant for annotation
(4) Influence of time on annotation needs, which was investigated by a follow-up study phase enabling participants to (re)annotate—if needed—a subset of their calls
(5) Derived recommendations for the design of note-taking support tools for daily mobile phone conversations.

We expect our work to provide new insights to these topics and to stimulate the design of memory prosthetic support for mobile phone calls.

## 2. RELATED WORK

The act of taking notes is an activity that people frequently perform to record information from other sources for later use. Geyer et al. [2005] state that "personal notes primarily serve as a memory aid for individuals to remember important facts, actions, ideas, and decisions but are hardly useful for persons other than the author." Annotations may be as simple as highlighting passages or producing a new written product. In the first case, certain fragments of information from the original piece are marked as they are considered important. In the second, new written pieces are composed on the basis of the original information.

These free-text annotations might provide additional information beyond the original piece that can enrich the information, such as for providing context or indicating what to do with the information in the future. To provide similar functionality, text highlighting can be enriched by certain behaviors, such as highlighting contextual information beyond the specific information to be remembered (e.g., instead of highlighting only the address of a meeting point, the annotator can also highlight the question related to the address). In both note-taking strategies, there is the need to preserve certain pieces of information for later access and use.

It is often the case that one has to take notes in cooperative situations, such as in meetings, at lectures, or during phone calls. Related research in recent years has focused primarily on studying *needs* for annotation in work-related meetings. In this setting, notes have a structured form and usually include action items. Often these notes are recorded to create a shared group memory and to make the meeting more efficient [Geyer et al. 2005]. Alternatively, notes can be used as memory cues for participants to recall events of a meeting rather than being full recordings of the activity [Whittaker et al. 2008]. Moreover, these notes can serve as markers to add structure to meeting recordings [Whittaker et al. 1994; Geyer et al. 2005; Wilcox et al. 1997; Bothin and Clough 2012]. In both cases, attention and active participation is required, and taking notes at the same time may become an additional cognitive load [Piolat et al. 2005] that reduces the person's ability to participate [Whittaker et al. 2008].

These needs have been lately supported by various artifacts, including electronic annotation tools that leverage desktop computers [Kleek et al. 2011; Bernstein et al. 2008] or mobile devices [Hinckley et al. 2007; Kam et al. 2005; Cragun 2009; Hayes et al. 2004], as well as the common paper and pen approach, still frequently used in the form of Post-It notes, miscellaneous text files, or the corner of other printed documents [Bernstein et al. 2008; Whittaker et al. 2008].

The work presented herein focuses on annotations of daily mobile phone calls—that is, parts of a phone conversation that participants of a phone call would consider worth preserving for later use. Phone calls are different from work-related gatherings in a number of ways. For example, typical mobile phone calls tend to be relatively shorter, are frequently not planned beforehand, and lack the structure of a meeting, relaying instead a series of salutations and informal dialogs. In consequence, the pieces of information that the participant of a phone conversation may need to remember—to annotate—and his or her motivations to take notes differ from those of a participant of a work meeting. In addition, it has been observed that during phone calls, participants often have their hands busy, either by performing another activity (e.g., driving) or by holding the phone, documents, or other objects [O'Hara et al. 2002]. Despite these differences, annotation of mobile phone calls has received little coverage in current research literature.

These observations support the need to study note-taking habits explicitly in contexts of informal mobile phone calls. As an initial step of this work, we aimed at understanding current needs and artifacts for annotating mobile phone calls as stated by our first research question:

*RQ1: What needs and artifacts do people have to annotate information exchanged in mobile phone calls?*

Researchers have also been dedicated to understanding *influencing factors* when annotating information, particularly in work meetings. As Lin et al. [2004] identified previously, the first step in the lifecycle of a note is the need to annotate something. Studies conducted by Bothin et al. [2010] and Bothin and Clough [2012] revealed that demographics and contextual information might be significant influencing factors of note-taking in work meetings. According to their findings, women tend to take more notes and do so more often than men. They also observed that older participants take more notes than do younger ones. Additionally, the role of the participant in the meeting was found to have a direct influence on the amount of notes taken: meeting leaders talk more and hence have less opportunity to take notes, whereas project managers have to take more notes because of their responsibility to produce meeting minutes.

Note-taking influencing factors have also been studied in the reverse order—in other words, by looking at annotation occurrence to infer other pieces of information. For example, Bothin and Clough [2012] explored the relationship between topics in meetings and the annotating behavior of their participants, and suggested that the presence of annotations could have some predictive power to estimate when something important is about to be discussed.

However, there is little evidence that the aforementioned findings can be generalized to annotations of mobile phone calls. The second goal of our research is therefore stated as follows:

*RQ2. Which factors mostly influence the need for creating annotations during mobile phone calls?*

The information on which we take notes is usually the information that we consider to be the most important to be preserved for later use. Understanding how often each party—caller and callee—can be the source of such important information is key to simplifying automatic approaches for mobile phone call annotation. To investigate this, we leveraged concepts from Conversation Analysis, a research area that has recently focused on studying communication in mobile phone calls [Arminen and Leinonen 2006; Hutchby and Barnett 2005; Laurier 2001; Weilenmann 2003] beside its frequent studies on landline phone calls [Schegloff 2002; Whalen et al. 1988; Lee 2009]. We used the concept of turn [Sacks et al. 1974] when analyzing our dataset to investigate whether callers annotate information shared by the callee more often than they annotate information shared by themselves. We aimed to understand this both on a general level as well as in relation to contextual variables—such as relationship with the callee, location, and companion at the time of the call and the objective of the call. Hence, our third research question:

*RQ3. How often and in which circumstances does each mobile phone party—caller and callee—provide information that the caller considers to be worth remembering?*

For simplicity, we split this broad research question into the following ones:

*RQ3.1. Do callers annotate—by means of highlighting mobile phone call transcripts—information shared by the callee significantly more often than information shared by themselves?*

*RQ3.2. Which factors mostly influence callers to highlight information in the turns of the two parties involved in a call (i.e., caller's turns, callee's turns, or both parties' turns)?*

Finally, scholars have looked at temporal effects on people's annotation needs in contexts of meetings and lectures. As people's memory recall abilities diminish with time, several tools for aiding their memory have been applied, thus composing their prosthetic—rather than the natural organic—memory [Kalnikaité and Whittaker 2007]. In the lab study conducted by Kalnikaité and Whittaker [2007], participants' prosthetic memory supported by paper and pen notes were shown to have high recall value and retrieval efficiency in the short term. However, accuracy decreases rapidly to the point of becoming useless after 1 month when compared to organic memory. This trade-off between recall accuracy and retrieval efficiency has been observed by Whittaker et al. [1994].

As time passes, actions taken over notes vary. After initial information consumption is completed, notes are either discarded or archived [Lin et al. 2004]. Strategies used to perform these actions vary [Kleek et al. 2011]. Whittaker et al. [2008] found that most annotators access their personal records of meetings afterward, and 75% access them frequently. The same work found that meeting participants try to keep notes accurate, and half of them even rewrite their notes. In terms of archiving, 75% of the participants keep their notes for a year on average.

These findings are mostly related to the context of work meetings and hence might not be generalized for annotation of informal mobile phone calls. We therefore state our fourth research question:

*RQ4. Does the annotation behavior—by means of highlighting mobile phone call transcripts—change over time?*

In case we find the answer to this question to be affirmative, additional questions should also be investigated to better understand such change:

*RQ4.1. Does the topic and amount of information in highlights change across time?*

*RQ4.2. How does this information change—if any—unfold across time?*

*RQ4.3. Does the importance of the highlighted information change across time?*

*RQ4.4. What are the reasons presented by participants for the change in their highlighting behavior?*

*RQ4.5. How can annotation behavior change—if any—be explained in terms of the previously investigated factors?*

The rest of the article is organized as follows. In Section 3, we explain the design and methodology of the study. In Section 4, we address each research question separately, presenting relevant results and discussing them in detail. Finally, in Section 5, we present a set of implications for the design of a mobile phone application to highlight important information shared during mobile phone conversations.

## 3. METHODOLOGY

We deployed a user study consisting of two phases. The first phase ($P1$), which spanned over 64 days, allowed us to collect a large sample of outgoing[1] mobile phone calls; the highlights of important information inside them, if any; and contextual parameters at the time of the calls. The second phase ($P2$) was based on questionnaires that produced a second round of highlighting on a subset of the same calls, allowing us to understand the effects of time on the highlighting behavior of the participants.

[1]We considered only outgoing calls given that these are intentional with a clear user need, thus reducing the number of random and/or undesired calls in our sample.

### 3.1. Participants

A total of 62 subjects (20 female), with a mean age of 31.5 years ($s = 7.52$, $min = 20$, $max = 51$) participated actively in the user study by answering the prestudy questionnaire and contributing at least one mobile phone call. All of them were residents of Spain and reported being fluent in the Spanish language—a requirement of the study. The sample was geographically well distributed (38 unique cities) and included only subjects who had received basic education at least (primary school: 3.2%), followed by 3.2% who finished secondary school, 79% who concluded technical school or obtained a bachelor degree, and 14.5% who had either a master's or doctorate degree. The reported annual income suggests that all social classes were represented in the sample (27.4%, 19.4%, 25.8%, 19.4%, and 8.1% earned up to €10K, €20K, €30K, €40K, and more than €40K a year, respectively).

### 3.2. Procedure

Participants were recruited among people who voluntarily registered after following advertisements in popular Web portals in Spain. We opted for asking participants to install a specific VoIP application in their smartphones to enable recording and transcription of their calls for later analysis. The application was available for Android and iPhone platforms only. Candidates who owned a mobile phone with either of these platforms were invited via email to be part of the study and were asked to answer an online prestudy questionnaire. Besides collecting general demographics, the questionnaire also asked participants about their calling habits and general note-taking habits during phone calls. We then conducted a user study divided in two phases, as explained in the following section.

*3.2.1. Study Phase One ($P1$).* The goal of this phase was to gather a large sample of mobile phone calls, including their audio, full transcription, contextual information, and textual fragments of each call transcription that the respective caller would consider important or worth remembering in the future. This data was used to address $RQ1$ and $RQ2$.

*Phone call setup.* During $P1$, we offered participants free calls to mobile or fixed phone lines inside the Spanish territory. To be able to make free calls, they had to install a VoIP application on their mobile phones and configure it to connect through our servers. Once the application was installed, making mobile phone calls through the application was transparent for users: whenever a call was dialed using the native phone keypad, it was automatically redirected through the VoIP system.

*Privacy and information security.* We explained to participants how to deactivate the application to prevent us from having access to more private phone conversations. To further comply with privacy protection laws, whenever participants placed a call through our system, a short message was played to both the caller and the callee, informing them that the call was going to be recorded and transcribed.

*Call transcription.* Since it was out of the scope of our study to work on or improve the state of the art in speech-to-text, we hired an external service provider to transcribe the phone conversation recordings into text. Transcriptions were first generated by the provider automatically—by means of a hidden Markov models–based method [Juang and Rabiner 1991]—and later manually inspected and corrected by a human expert before being presented to participants. The resulting transcription for each call was an annotated file that contained the text of the whole call and the adequate labels for both caller's and callee's turns. Besides helping to identify each party's turns, we used the labels as a separator for parsing the transcriptions. This way, by using a simple regular expression, we could divide the call into individual turns for quantitative

analyses. Example 1 shows a fragment of a transcription as provided by the transcription service[2]:

*Example 1*:

```
         (...)
CALLER:  Hey... where are you?
CALLEE:  Still here in Fnac.
CALLER:  Ok, we are down here, by the door. Near the tree, in Plaza de Callao.
CALLEE:  The tree? Is there a tree there? Well, I'm here still undecided, but
         I'm going down, but we probably have to pay a book that Irene bought,
         so...
CALLER:  Ok, then come to Callao when you are out, to the Christmas House.
CALLEE:  Ok, good, we'll go as soon as possible. See you later.
CALLER:  See you.
```

*Annotations in the form of highlights.* Annotations are usually studied in the form of spontaneous, free-text *note-taking*, which involves a high cognitive load and effort for synthesizing the original information and composing new pieces of derived text [Piolat et al. 2005]. Another form of annotation is that of *highlighting* text fragments, which is consistent with the offline/explicit indexing strategy explained by Geyer et al. [2005] for marking meaningful points in meeting records. Both cases have the same goal: to preserve pieces of important information for later use.

On this basis, we opted for studying annotations in the form of highlights rather than free text generated by the note-taker for a couple of reasons. First, generating free-text form annotations for each phone call would demand higher effort from the participants, which could lead to negligent behavior when deciding whether a call should be annotated or not. Second, collecting useful fragments from calls would allow us to later investigate whether annotations can be generated automatically at an acceptable error rate.

*Call questionnaire.* We developed a Web application that allowed participants to interact with call data. It displayed a list of their calls, and for every call, its transcription, date, duration, status, and callee number, as well as a related call questionnaire. Participants were free to complete this questionnaire whenever they wanted for the duration of $P1$. They were also allowed to delete[3] a call within a 24-hour period if they considered it to have sensitive content.

In the call questionnaire, we gave participants the following task for all of their transcribed calls: *Highlight the parts that you consider to be important and that you would like to remember for later.*[4] We therefore asked participants to regard information as *important* when it was worthy of being preserved to be used at a later time. In cases when they did not find any important text to highlight, we asked them to explicitly declare so. Participants received a monetary incentive for answering call questionnaires regardless of highlighting anything in the call transcript or not. We expect that this approach motivated participants to provide us with more detailed information about their calls without biasing their highlighting behavior.

Finally, participants were presented a series of contextual questions related to each phone call: (1) relationship with callee, (2) who was with the caller at the time of the call, (3) location of the caller at the time of the call, (4) objective of the call, (5) level of

[2]Call transcription examples were translated from Spanish (their original language) to English by the authors.

[3]Participants deleted a total of 65 calls recorded during the study, which represents 7.5% of the final working dataset ($N = 796$ calls).

[4]This is the exact phrasing that we used, as translated from the original Spanish version: "*resalta las partes que consideres importantes y que quisieras recordar para después.*"

importance of the call, (6) level of importance of the notes, and (7) general questions about sound and transcription quality.

Given that participants reported objectives of calls in free text, we needed to develop a coding scheme to further analyze this variable. To minimize investigator bias, we asked a colleague researcher who was not aware of this study to inspect the reported objectives and to suggest a coding scheme. He then met with one of the authors, and they agreed on a final coding scheme that included the following categories: "discuss a topic," "ask and/or receive specific information," "set an appointment," "social," and "other."

They both then proceeded to classify the answers separately. Interrater reliability was tested using Cohen's kappa, and it was found to be highly acceptable ($K = .86$, $p < .001$).

*Analysis of highlighted turns*. We leveraged some concepts from the Conversation Analysis area to study the call transcriptions to obtain insight on how frequently important information is highlighted among the contributions of the caller and the callee.

A phone call conversation is composed by several alternating interactions contributed by the two involved parties. Each of these interactions is called a *turn* [Sacks et al. 1974]. They can be considered as the basic building blocks of any conversation. Consecutive turns constitute units called *sequences*. For instance, example 2 comprises a sequence of four turns extracted from one of the conversations of our dataset (text in bold indicates the highlights made by the caller):

*Example 2*:

4 (...)
5 CALLER: **Have you bought something for mom already?**
6 CALLEE: No, I was going to do it, but I don't know... after the next week or the other one.
7 CALLER: Did you have something in mind already?
8 CALLEE: As I told you **I thought about buying her a coat** because she once said she would like to have one.
9 (...)

Therefore, we first selected the calls that contained at least one highlight ($N = 299$) and parsed their corresponding transcriptions to extract the turns. Afterward, for each call, we counted the number of highlighted turns—that is, the turns containing at least one highlighted piece of information. It is important to note that we asked participants to highlight the pieces of text from the transcription that they considered to be important or worth remembering. This is our definition of *important information*. With this in mind, we consider that if a turn contains a piece of highlighted information, the turn itself contains important information (i.e., it is a highlighted turn).

Having counted highlighted turns for both parties in every call, we determined whether (1) the caller highlighted information only in his own turns, (2) the caller highlighted information only in the callee's turns, or (3) the caller highlighted information in both parties' turns. In other words, the caller highlighted some information that she or he said during the call, *as well as* some other information—not necessarily the same information—that the callee said during the call. Every call was therefore classified into one of these three groups.

To further explain these groups, let us reconsider example 2, which has turns 5 and 8 highlighted. Given that the former is a caller's turn (turn 5) and the latter is a callee's turn, we conclude that the caller highlighted information in *both parties' turns* for that specific call. In example 3, however, all of the highlighted turns[5] are caller's turns, and

[5]Although example 3 shows only a fragment of the call, no other turns, besides those highlighted in bold, were highlighted in the entire call.

hence we conclude that the caller highlighted information *only in his or her turns* for the given call.

*Example 3*:

```
7           (...)
8   CALLER: Listen, I am calling to tell you that I am going to Burgos tomorrow
            in the morning.
9   CALLEE: Really? Why?
10  CALLER: Well because Diego will leave no sooner than six and I'll leave
            tomorrow at twelve so I have bought the one o'clock ticket.
11  CALLEE: Good, Ok. So will you be home for lunch?
12  CALLER: I don't know for lunch, because one plus three is four... You eat,
            if you can leave something for me, then great.
13          (...)
```

After manually inspecting the highlighted turns, we observed that many of them contained questions and that they mainly provide context to the actual information sought by the person who asked. As stated by Sadock and Zwicky [1985], interrogative sentences "signal the desire of the speaker to gain information from the addressee." Given this main role of questions, they were highlighted by participants most likely to provide contextual information for framing a forthcoming answer. For instance, in example 2, the highlighted question in turn 5 (*Have you bought something for mom already?*), provides context to the information sought by the caller and hence highlighted in turn 8 (*I thought about buying her a coat*). Although both turn 5 and turn 8 were highlighted, we consider that the callee (turn 8) provided relevant information, whereas the caller's turn (turn 5) was highlighted mostly to provide context.

Therefore, in a similar way as we studied the frequency of highlights in call parties' turns, we also studied how frequently *questions* were highlighted among the caller's, callee's, and both parties' turns.[6] This helped us to understand who the party is (either the caller or the callee); who provides more questions—the main information seeker; and thus, in those cases, who provides more important contextual information. To round up this analysis, we also analyzed the highlighted turns excluding questions to determine how frequently highlighted information, excluding contextual information contained in questions, appears in each parties' turns.

Finally, we searched for correlations between the presence of highlighted information—in each and in both parties' turns—and the contextual variables obtained during the study. All of the aforementioned measures were used for addressing $RQ3$.

*3.2.2. Study Phase Two (P2).* From $P1$, we obtained an initial dataset to understand the highlighting needs of participants in a time close to their calls. The aim of $P2$ was to extend this data with a second round of highlights that participants could take further in time from calls, which would allow us to study how annotation needs change over time. This data was used to address $RQ4$.

To collect such data, we invited participants to fill out a questionnaire—similar to the call questionnaire in $P1$—about the calls they made during the first phase of the study; $P2$ was not mandatory, and participants in $P1$ received a monetary incentive to also participate in $P2$. Probing participants on every call they made in $P1$ would reduce reliability of our results due to survey response fatigue, given that each subject made 13 calls on average during $P1$. We hence decided to restrict participants' questionnaires to a subset of their calls. In a pilot conducted before $P2$, test subjects—not related to the study sample—answered the questionnaire for up to six calls without any fatigue-related complaint. Therefore, we opted for selecting a maximum of six calls per

[6]To detect questions, we first manually reviewed the dataset to verify the correct presence of question marks (in Spanish ¿ and ?) in turns representing questions and then parsed the transcriptions.

participant: three calls that she or he previously highlighted in $P1$ (if any), and three calls that she or he did not highlight in $P1$ (if any). The selection of the six calls had to be representative about each user's participation in the entire $P1$ study. To achieve this goal, we ordered each participant's *highlighted* calls according to their annotation date and time, and then selected his or her first call, mid-point call, and last call. The same process was applied for *nonhighlighted* calls toward selecting a maximum of six calls per participant. By following this procedure, we obtained a more representative set of calls per participant covering their entire experience in $P1$ as evenly as possible. Furthermore, this procedure balanced the amount of data points across time, thus maximizing statistical power for the study of $RQ4$.

The $P2$ questionnaire look and feel was similar to the one used in $P1$. For each of the selected calls, we again offered participants the opportunity to highlight the pieces of information from the transcription that they considered relevant. It is worth remarking that we were not measuring recall, as participants could highlight whatever information in $P2$ they considered to be important at that particular time. Up to this point, we did not present initial highlights made in $P1$, if any. If participants considered that call transcriptions in $P2$ had no important information, they were required to inform that. Next, participants were asked about the level of importance of the call, reasons for highlighting it (if applicable), and level of importance of the notes.

Once they finished this step, participants were presented with a dynamic section, which could branch into one of the following:

(1) *Call was highlighted both in* $P1$ *and* $P2$: Both sets of highlights were presented. Participants were asked if the highlights made during $P2$ were less important, equally important, or more important than those made in $P1$, and to explain why.
(2) *Call was highlighted in* $P1$ *but not in* $P2$: Original highlights were presented and participants were asked about the reason for not highlighting anything in $P2$.
(3) *Call was not highlighted in* $P1$ *but was highlighted in* $P2$: Recent highlights made in $P2$ were presented, and participants were asked about the reason for not having highlighted anything in $P1$.
(4) *Call was not highlighted in either* $P1$ *or* $P2$: No further questions were asked.

Participants' reasons for highlighting different information in $P1$ and $P2$ were coded using the same procedure described for coding the objectives of calls. The categories used were "temporal effect," "information refinement," "archiving," "same content," and "no clear answer." Interrater reliability for this coding scheme was tested and evaluated as highly acceptable ($K = .85$, $p < .001$).

Finally, we aimed to quantify thematic differences between notes taken in both phases. To achieve this, two coders inspected all highlights in $P1$ and $P2$ and then agreed on classifying differences in four categories:

—*No change of information*: Both notes dealt with the same topic and conveyed the same amount of information.
—*Information increase*: Both notes dealt with the same topic. Notes from $P2$ conveyed more information than notes from $P1$, such as *Where are you, I am in the garage* versus *I am in the garage because I had to get some oil and it took some time*.
—*Information decrease*: Both notes dealt with the same topic. Notes from $P2$ conveyed less information than notes from $P1$, such as *I'm going to take it back / the T-shirt / Don Jaime street / I will wait for you at <supermarket name>* versus *at <supermarket name>*.
—*Different topic*: The topic of the notes changed from $P1$ to $P2$, thus the information they contain could not be compared, such as *Let's meet tomorrow at five / at Sevilla on February the 19th* versus *I work from home on Friday / call you later*.

The two coders applied this classification scheme, with a highly acceptable inter-rater reliability ($K = .81$, $p < .001$).

### 3.3. Data Preprocessing

Before analyzing data collected in both phases, we first reviewed all of the participants' comments and objectives of calls reported in the $P1$ call questionnaires. From this initial analysis, we observed that most of the participants' first calls were justified as being test calls. These calls were removed from our data. Furthermore, calls that could not be properly transcribed were also filtered out. The sample reported herein (796 calls and 62 participants) is therefore our final working dataset after applying these filtering heuristics.

## 4. RESULTS AND DISCUSSION

During the study deployment, participants made 796 calls with a total duration time of 141,741 seconds ($\bar{x} = 178.07$; $s = 364.55$). Quality of the calls was considered acceptable ($\tilde{x} = 3$: acceptable; $Q1 = 3$: acceptable; $Q3 = 4$: good). Transcriptions of the calls and highlights yielded a total of 1,241,956 characters ($\bar{x} = 1{,}560.25$; $s = 3{,}094$) and 49,382 characters ($\bar{x} = 62.04$; $s = 198.01$), respectively. The average number of calls per participant was 12.84 ($s = 11.57, min = 1, max = 49$), and they highlighted an average of 4.92 of their calls ($s = 5.78, min = 0, max = 30$). Hence, 37.6% of all phone calls were annotated by highlighting, which is consistent with the participants' self-reported annotation habits captured by the prestudy questionnaire (34% and 45% indicated taking notes frequently using paper/pencil and their mobile phone, respectively). Likewise, participants called family members more often than friends, and called friends more often than work colleagues, which reveals the same order reported in the prestudy questionnaire. These findings support consistency between the participants' behavior in the study and how they perceive their behavior in real life. Next we present and discuss results related to each of the research questions described in Section 2.

### 4.1. Addressing RQ1: Phone Recall Needs and Artifacts

According to what participants reported in the prestudy questionnaire, recalling information from call conversations is a frequent need and not necessarily an easy task. Almost half of the sample agreed that this need occurs sometimes (47%), and more than one third indicated that it happens frequently (37%), whereas only 16% of the participants reported that this need rarely occurs. No one reported the absence of this need. When evaluating the easiness to recall information obtained in phone calls, 39% said that it is either easy or very easy, 35% reported that it is neither easy nor difficult, and the remaining 26% agreed that the recall task is at least difficult. These results suggest the importance of supporting recall of phone conversations.

In addition, we observed that only 4% of the characters in the transcribed calls were highlighted by the participants. This indicates that having full transcripts of conversations is not an optimal solution for phone recall, as they would most likely overload users, switching their problem from recall to information retrieval. As Lansdale [1988] points out, this is one of the major problems in personal information management that could be targeted by mobile applications. An optimal approach should include the recognition of annotation patterns together with contextual information of the call so that the user does not have to go through the entire transcript to retrieve the important pieces of information. Then the stored contextual information (i.e., metadata) can be used to optimize retrieval. This is in agreement with Whittaker et al. [2008] on their suggestions for work meeting–capturing tools.

With respect to the main artifacts used for recall, mobile phones and regular paper and pencil were reported as the most important ones—45% and 34%, respectively, use them frequently for this task. More specifically, participants reported taking offline notes of phone conversations using text-based notepads and audio-based memo applications (24% record audio notes for phone calls at least once a week).

It is worth noting that our sample is composed of smartphone users only, hence the popularity of mobile phones as the primary annotation source. Nevertheless, our results indicate that (1) the majority of this population segment often have needs to recall information exchanged in daily phone calls, and (2) smartphones are becoming relevant tools for annotating these conversations. In fact, smartphones seem to be the most convenient devices for that purpose, as most commonly reported annotation needs can be handled by a specific mobile application: 33.9% reported usually taking notes of phone calls to remember calendar-related information (dates and appointments), 33.9% said call notes are useful to remember contact information (phone numbers, names, and emails), and 22.3% wanted to remember to-do items (e.g., shopping lists).

### 4.2. Addressing RQ2: Variables That Influence Call Annotation

We looked into several variables that could potentially reveal the likelihood for one to highlight information in any given phone call. We collected and analyzed four types of variables:

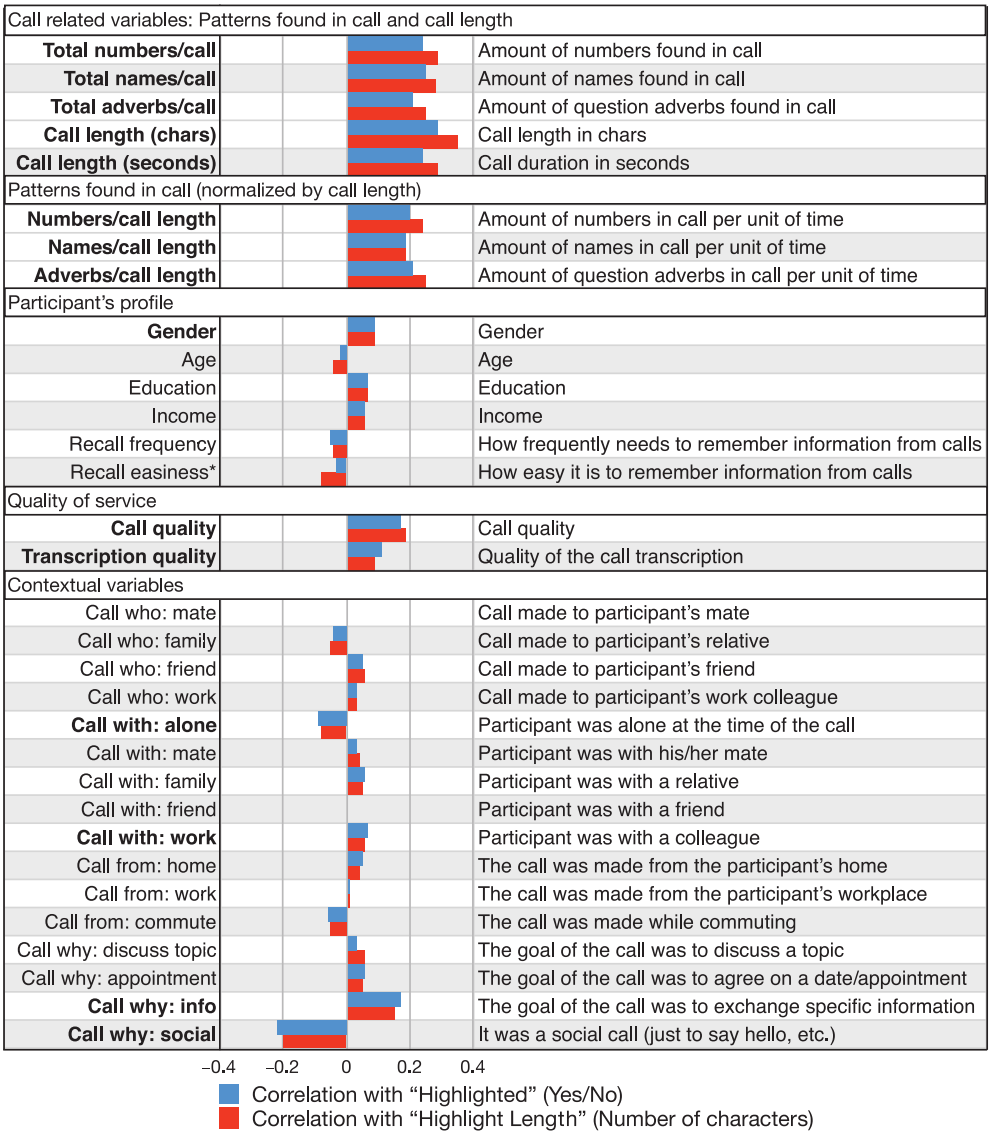
—*Patterns and call related variables*: In the preliminary questionnaire, participants reported the kinds of information that they find themselves trying to remember after a phone call. About 81% mentioned pieces of information that necessarily include numbers (e.g., phone numbers, dates, prices, addresses), and 47% mentioned information related to names (e.g., addresses, contacts). On this basis, we implemented two parsers to count numbers and names in call transcriptions, aiming to evaluate whether this information could be a relevant driver for highlighting information in the call. Example of text fragments that were considered to have a number include "three hundred euro," "half kilo," and "fourth street," among others. We also implemented a third parser to count interrogative adverbs (i.e., why, where, how, when). Our reasoning is that the occurrence of an interrogative adverb implies the presence of a question, which is, by definition, an explicit request for information. Such a request would therefore be followed by important information that might not necessarily appear in the form of numbers or names. Additional call-related variables include call length both in characters and seconds.

—*Participant's profile*: These variables were collected in the preliminary questionnaire, and they helped us understand whether highlighting behavior can change based on demographic features.

—*Quality of service*: We asked participants about the quality of calls and transcriptions to investigate how they might influence highlighting.

—*Contextual variables*: We supposed that the context of a call might influence the need for highlighting information. Furthermore, given that we asked participants to highlight *important* pieces of information, we also looked for an association between the stated importance of the highlights and the stated importance of the call.

Figure 1 shows the tested variables, their correlation with the highlighting parameters (*Highlighted,* i.e., whether the call was highlighted or not, and *Highlight length* in characters), and their meaning. We report both variables that displayed a significant correlation and those that did not. This helps us provide a more complete discussion of which would be more useful for predicting annotation of mobile phone calls as well as those that might not. Next we discuss these results. We refer to the names of studied variables in italics.




*Recall easiness was found to be significantly correlated with highlight length only.

Fig. 1. Correlations/associations between highlight-related variables (*Highlighted*: Yes/No; *Highlight Length* in number of characters) and other variables related to the call task. Correlations between ordinal and nonnormal interval variables were assessed using Spearman's rho ($\rho$). Associations between dichotomous variables were assessed using the $\chi^2$-derived phi coefficient ($\phi$). Variables in bold have significant coefficients at $p < .05$.

*Patterns in mobile phone call annotations are better indicators of note-taking than most of the variables observed in the study*. According to Figure 1, these pattern variables (*Total numbers/call, Total names/call, Total adverbs/call*) have some of the highest correlation coefficients with the length of the highlights ($\rho = .29$, $\rho = .28$, $\rho = .25$, respectively).

We further normalized values of these pattern variables by call length, and correlations were also among the highest ones observed in our study ($\rho = .24$, $\rho = .19$, $\rho = .25$, respectively). In other words, the more numbers, names, and interrogative adverbs mentioned in a call, or the higher the amount of them per unit of time, the higher the probability to highlight information—take notes—and also the longer the highlights might be. In the specific case of interrogative adverbs, the ones with higher correlation with highlight length were "when" ($\rho = 0.24$), "how much/how many" ($\rho = 0.18$), "who" ($\rho = 0.15$), and "where" ($\rho = 0.10$). This suggests that the questions that would most frequently lead to highlight information are respectively those related to planning events (e.g., meetings, special events, holidays), discussing numeric values (e.g., prices), people, and places, respectively. Last, *Call length* is the only variable that surpasses correlation results for patterns (in characters: $\rho = .35$; in seconds: $\rho = .29$).

Additionally, we found some weak yet significant correlations between contextual variables and the occurrence of patterns in calls. However, they are worth being reported, as they might indirectly influence highlighting behavior. We found that calls made to the caller's mate tended to have less proper names ($\rho = -0.1$), whereas calls made to friends seemed to have more proper names ($\rho = 0.13$). This suggests that couples tend to discuss topics less related to things that need to be referred to by a proper name (e.g., places, people, or brands) and maybe focusing more in their common daily activities. On the other hand, friends seem to mention more information related to third parties, news, recently discovered places, and so forth (e.g.,[7] *CALLER: (...) Who's coming? Aitor, Raul...? CALLEE: Aitor, Gorka, Raul and I don't know who else. (...)*). In terms of the objective of the call, calls made with the goal to set up a meeting/appointment were positively correlated with the number of names mentioned during the call ($\rho = 0.17$). This is most likely due to the mention of addresses or meeting points during the call, as these items frequently include streets or specific place names (e.g., *CALLEE: (...) Should we meet at eight in ... some bar of the San Juan Street? Is it good for you?*). The opposite happened in social calls ($\rho = -0.13$). Although this kind of conversation might also exchange names casually, social calls are more frequently made without a specific goal, instead with the need to "say hello" or to see how the callee is doing. In consequence, they less frequently contain proper names.

*Profile and demographic information of the caller do not seem to be related to phone call annotation.* According to Figure 1, most of the callers' demographic variables (i.e., age, education, and income) did not reveal significant correlations with the highlighting variables. *Gender* (coded as 1: male; 2: female) seems to be an exception, implying that women might highlight more information than men. Bothin et al. [2010] obtained a similar finding for the context of work meetings. However, they reported a medium effect size for women to take notes more frequently and for longer than men, whereas we observed an effect below the standard weak correlation threshold ($\rho = .09 < .10$) [Cohen 1988]. Similarly, our participants' self-reported *Recall easiness* also revealed a below-weak correlation with length of highlights, implying that the easier one considers the recall task, the shorter his or her highlights ($\rho = -.08 > -.10$). Future work should clarify whether gender and recall easiness are indeed significantly correlated with phone call annotation.

*Quality of service (QoS) parameters are weakly correlated to phone annotation.* All QoS variables—as reported by participants—were positively correlated to making phone call highlights. Moreover, *Call quality* was positively correlated to both the highlighting activity ($\rho = .17$) and length of highlights (in characters: $\rho = .37$; *in seconds* : $\rho = .26$). One possible explanation is that the poorer the quality of calls, the less

[7]Actual examples from the study dataset.

users speak and spend time in a phone conversation, thus reducing the probability of highlighting information.

*Contextual variables play distinct roles in the note-taking activity.* Information related to the call place, caller's companion, and callee information did not reveal any significant relationship with call highlighting, or rather only below-weak correlations (see variables *Call from*, *Call with*, and *Call who* in Figure 1). However, call objective (*Call why*) seems to be more connected with the users' highlihgting needs. For example, whenever participants made social calls (*Call why:social*, i.e., call objectives reported as *just to chat* or *to say hello/goodbye*), less highlights were made ($\rho = -.22$). On the other side, calls to give and/or ask for information (*Call why:info*) received more highlights ($\rho = .17$).

*Importance of calls and importance of their notes are related to call context.* By means of studying associations between call/notes importance and contextual variables, we observed some significant—although mostly weak—correlations that are worth mentioning. We found that calls made to the caller's mate had the tendency to have lower importance, as evidenced by a negative correlation ($\rho = -0.12$), whereas the opposite happened for calls made to friends ($\rho = 0.1$). It is possible that phone conversations between partners tend to have routine information that is more predictable and then frequently regarded as less important, whereas phone conversations with friends include information considered to be new or unexpected, thus considered to be more important. Call importance tended to be lower for social calls ($\rho = -0.35$), whereas calls made to discuss a topic showed a tendency to be more important ($\rho = 0.13$). As mentioned earlier, social calls are mostly made for very informal or casual purposes, and thus these calls tend to be regarded as unimportant. On the other hand, calls made with the goal of discussing a topic have a clear goal, which is probably the reason they are more frequently considered important. With regard to the importance assigned to notes, we observed a positive association with calls made to a friend or to a service provider, like restaurants, shops, and so forth ($\rho = 0.13$ and $\rho = 0.14$, respectively). This suggests that highlighted information from these calls are frequently important for conducting further actions, such as attending appointments based on highlights of location details, buying artifacts based on highlights of shopping items, or performing specific tasks based on highlights of instructions. Still regarding note importance, we found a similar positive association with calls made with the objective of setting up a meeting/appointment and for discussing a topic ($\rho = 0.15$ and $\rho = 0.13$, respectively). The first case might be due to the importance of the information in the notes for attending the commitment, and the second suggests that the discussion of topics frequently produces information that is important for later use. Finally, social calls tended to have less important notes ($\rho = -0.23$). This once again confirms that social calls—and notes derived from it—are generally not very useful for information exchange. Their purpose is probably to keep in touch with people we know to nourish our social circles.

*Important notes are taken from important calls*. As a last result in this section, our experimental data corroborate what one would expect: importance of calls as evaluated by the callers strongly correlates to the importance that they attributed to highlighted information in those calls ($\rho = .51, p < .001$). This finding might indicate that the higher one thinks is the importance of a call, the more likely important annotations will be created for the conversation.

These results support and expand our major findings from Carrascal et al. [2012] and further clarify some early doubts. For example, in our previous work, we inquired whether the associations between the *Call why:info* variable and the highlighting-related variables (*Note Taken* and *Note Length*) were significant. With the larger dataset used in the work presented herein, we were able to corroborate that these associations were indeed significant. Similarly, the early apparent significant association

between *Gender* and *Notes Taken* was corroborated in this work, which is also in accordance with results from Bothin et al. [2010].

In summary, our findings indicate that users tend to highlight—or annotate—mostly information containing numbers and names, such as phone numbers, addresses, dates, shopping lists, or contacts. These patterns can be easily identified, signaling the importance to annotate calls. Profile and demographic information of the caller are most likely not relevant indicators of annotation, whereas QoS parameters could potentially inform it—poor quality calls may lead to shorter calls with fewer notes. Finally, contextual information such as the call objective and call length seem to be good indicators for call annotation: whereas calls that intended to give or receive specific pieces of information did require annotations to be taken, social calls did not. Future work should investigate whether these variables related to mobile phone call annotation could also influence note-taking in work-related meetings [Bothin et al. 2010; Bothin and Clough 2012].

### 4.3. Addressing RQ3: How Often and in Which Circumstances Does Each Mobile Phone Party—Caller and Callee—Provide Information That the Caller Considers to Be Worth Remembering?

With this research question, we want to understand, in terms of turns, how frequently and in which circumstances the important information exchanged during a mobile phone call appears in each of the parties' turns. The dataset used to address this question—all calls that contained information highlighted by participants in the initial study phase ($N = 299$)—contained a total of 11,958 turns ($\tilde{x} = 25$ turns per call, $min = 2$, $max = 463$). The amount of caller's and callee's turns per call were balanced across the dataset (caller: $\tilde{x} = 12$ turns per call, $min = 1$, $max = 232$; callee: $\tilde{x} = 12$, $min = 1$, $max = 231$). From all annotated turns, 833 (7%) contained at least some information highlighted by the participants—that is, they were *highlighted turns*. From these highlighted turns, 441 (3.7%) were caller's turns and 392 (3.3%) were callee's turns. Next we will go deeper into understanding where the highlighted information tends to appear more frequently.

*4.3.1. RQ3.1. Do Callers Annotate—by Means of Highlighting Mobile Phone Call Transcripts—Information Shared by the Callee Significantly More Often Than Information Shared by Themselves?* As explained in the methodology section, we defined three groups of calls depending on where the highlighted information was contained: in the caller's turns, in the callee's turns, or in both. Figure 2 shows the frequencies of these groups (confidence intervals in Figures 2 and 3 are exact binomial confidence intervals, calculated according to Clopper and Pearson [1934]).

An analysis using McNemar's chi-squared test using a significance level of 0.05 indicates that the proportion of the first group (information highlighted in caller's turns only) is significantly higher ($\chi^2 = 5.20$) than the second (information highlighted in callee's turns only), and the proportion of the third group (information highlighted in both parties' turns) is also significantly higher ($\chi^2 = 14.80$) than the second. Therefore, there is a higher proportion of calls with important information provided by the caller (or by both parties) than by the callee alone.

This result may appear counterintuitive, as one might think that given that the caller is taking the initiative when dialing, probably more information should be contributed by the callee, as a response to the caller's inquiries. As stated in Section 3, we looked for a possible explanation by performing a manual inspection of call transcriptions. This revealed that in many cases, the callers' highlighted turns were questions. Furthermore, the information included in those questions is contextual, frequently highlighted for adding meaning to the information exchange. For instance, in one of the calls

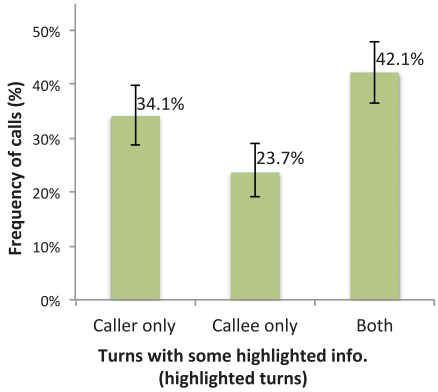



Fig. 2. Binomial proportions of calls according to what party—caller, callee, or both—had some information highlighted in his or her turns. Error bars indicate a 95% confidence interval.

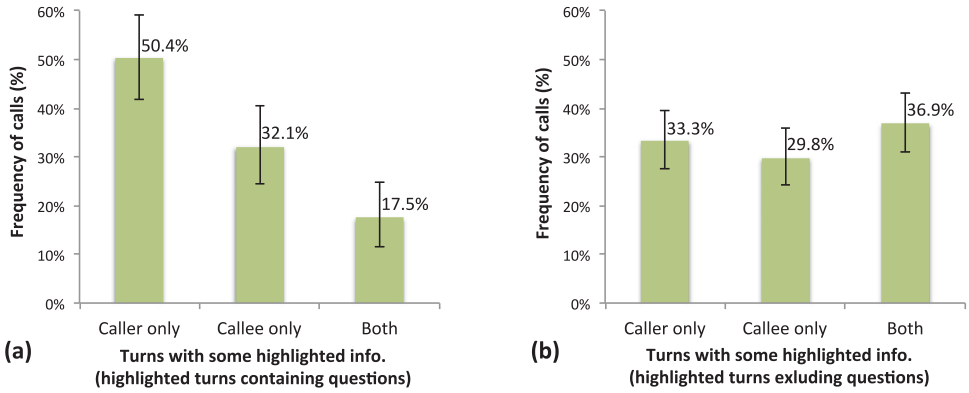



Fig. 3. (a) Binomial proportions of calls according to what party—caller, callee, or both—had information highlighted in his or her turns. (b) All highlighted turns excluding those with questions. Error bars indicate a 95% confidence interval.

including a discussion about buying a gift, the participant highlighted both the question *What color?* and the answer *Dark grey, because I wasn't sure if I should buy the green or the blue one [...]*. The actual information requested is in the second turn, whereas the first was highlighted to add meaning to it.

We found that about one quarter of the highlighted turns (23%) were questions. Figure 3(a) shows the proportions of highlighted turns containing questions, according to their location on each parties' turns. A McNemar's test confirmed that there are more calls with highlighted questions in the caller's turns than calls with highlighted questions in the callee's turns ($\chi^2 = 5.1$) or in both parties' ($\chi^2 = 20.82$). According to these results, from the caller's perspective, it is he or she who makes most of the important questions in a given call while also providing context for information exchanged during the call.

We further looked at the *number* of highlights—rather than the presence of highlights per call—in each party's turns that included questions versus those that did not. From the 441 caller's highlighted turns, 107 (24.3%) were questions—that is, context establishing turns—and the remaining 334 (75.7%) could be considered as integral to what is worth remembering. On the other hand, from the 392 callee's turns that were highlighted, 82 (20.9%) were questions and 310 (79.1%) were not. In this

analysis, however, we did not observe a significant difference between the rate of highlighted questions in the caller's turns versus those in the callee's turns (as we observed previously when analyzing presence of highlights *per call*). From these findings, we argue that it is more likely for calls to have questions highlighted only in the caller's turns than only in the callee's turns, which suggests that context to important information tend to be provided by the caller. However, whenever callers highlight questions only in the callee's turn, often more questions per call are highlighted.

Figure 3(b) shows that removing the highlighted questions from the highlighted turns made the proportions of highlighted turns between the three groups similar. In this case, no significant difference was found using McNemar's tests ($\chi^2 = 0.4$, $p = 0.53$; $\chi^2 = 0.3575$, $p = 0.55$; $\chi^2 = 1.7$, $p = 0.19$). In this case, roughly one third of the calls contained highlighted information in the caller's turns exclusively, one third contained highlighted information in the callee's turns exclusively, and one third contained highlighted information in both parties' turns. These results suggest that when looking for important pieces of information exchanged during a mobile phone call (excluding questions that usually just provide context), one might find them in any of the parties turns, or even in both of them.

Although the exchange of information excluding highlighted questions might give a clear view of where the important information more frequently appears, the data suggests that questions should be involved in the annotation process—especially those made by the caller. It is worth noticing that this is backed up by the results from Section 4.2, where we found that the presence of interrogative adverbs was one of the variables that displayed highest correlation with the note-taking activity. Although questions may play a number of roles in a conversation [Freed 1994], most of the time they are used for gaining information [Sadock and Zwicky 1985]. However, given our findings, we believe that an automatic annotation system should not consider questions as being merely markers for important information that is yet to appear during the conversation. Instead, questions themselves contain relevant contextual data that should be linked with information from their corresponding answers to construct meaningful notes.

Regarding *RQ3.1*, we conclude that callers do *not* highlight information on the callees' turns more often than on their own turns. In fact, callers more often highlight information in their turns exclusively or in both parties' turns. Analysis based on manual inspection of calls suggests that this is mostly due to annotating a caller's own questions to provide context for the information highlighted in the callee's turns, thus composing one meaningful note. Moreover, these findings can indicate collaborative construction of important information by both caller and callee. Next, we look further into this latter case.

*4.3.2. Which Factors Mostly Influence Callers to Highlight Information in the Turns of the Two Parties Involved in a Call (i.e., Caller's Turns, Callee's Turns, or Both Parties' Turns)?* We investigated if call context is related to the occurrence of important information in each parties' turns. More specifically, we looked for associations between contextual variables (e.g., calling a family member) and the presence of highlights in the caller's turn exclusively, in the callee's turn exclusively, or in both parties' turns. Associations between binary variables were calculated using the $\chi^2$-derived phi coefficient ($\phi$), and a significance level of 0.05 was used.

*Calls to one's significant other are collaborative: highlights are usually made in both parties' turns.* For calls made to the caller's partner (variable *Call who: mate*), we found a significant positive association with the presence of highlighted information in both parties' turns ($\phi = 0.12$). We also found a negative association between calling the partner and highlighting information in his or her turns—that is, callee's turns

($\phi = -0.12$). These results suggest that when having a phone conversation with their significant others, callers found important information in both what they said and in what their partners said, as opposed to finding it in one party's turns exclusively. Manual inspection of call transcripts suggests that calls between members of a romantic relationship tend to be collaborative, sharing common interest items such as activities to attend together and discussing what groceries to buy. A recurrent example of this is the case of conversations where partners synchronize their schedules (highlighted turns in bold):

*Example 4*:

CALLEE: So that's it. What time are you going to the gym?
CALLER: **I think today I'll go at 8:15, so we can meet before.**
CALLEE: **Fine. Yes, I hope so. In fact, we have work here, but I think I'll leave no later than 6. So at 7:00 I should be at home.**
CALLER: Fine it's just that...
CALLEE: Yes, tell me, tell me.
CALLER: **Of the activities they had before, I didn't like anything. Then I'll go to Body Combat, which is at 8:15.**

*Social calls are also collaborative: highlights are usually made in both parties' turns.* We also found a positive association between making a social call and highlighting information in both parties' turns ($\phi = 0.16$). After inspecting transcripts of these conversations, we observed how they are frequently related to daily and informal events, or to common acquaintances and news. Moreover, these conversations frequently flow without time constraints, and thus new conversation topics arise, bringing up spontaneous bits of information that can be considered relevant. For instance, note the conversation between a woman and her mother shown in Example 5.

*Example 5*:

CALLEE: They have a baby girl, Alba, and now they are waiting a kid, in about a month, they told us.
CALLER: And how they are going to call him?
CALLEE: What? I couldn't hear you
CALLER: I said how they are going to call him?
CALLEE: **Mario, they told us.**
CALLER: Vicky's child is going to be called Mario, too.
(...)
CALLER: **And next thursday is La Candelera*, isn't it?**
CALLEE: Yes, this thursday, La Candelera. Your father has a party, he says he is going to buy some seeds.

*A religious celebration.

*In informational calls, the caller has important things to say.* Conversely, we found a positive association between making phone calls to exchange specific information (variable *Call why: info*) and highlighting the caller's turns ($\phi = 0.16$). For this type of call, the need to share information with the callee is probably the main trigger for placing the call, and thus the important information is more frequently in the caller's turns. This happens frequently in situations when the person calls to inform the callee of past or forthcoming events, such as *Hey, about the scarf I told you somebody forgot in my house, well I think it's yours. [...]*, or *I called to tell you that I am going to Burgos tomorrow in the morning*. A negative association between making these information-based calls and highlighting information in both parties' turns ($\phi = -0.18$) further suggests that these types of calls tend to be task oriented. Therefore, giving or—less frequently—receiving information is enough without much collaboration between parties to generate content worth highlighting in both turns.

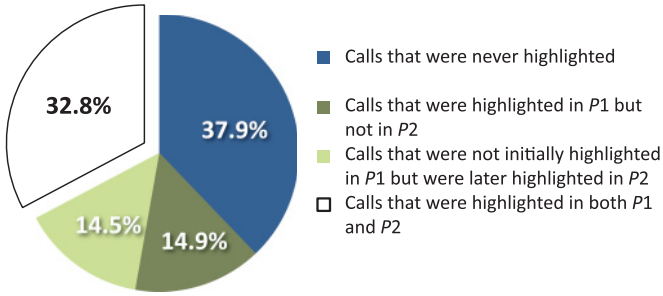



Fig. 4. Overview of all calls used to study the temporal effect on highlights ($N = 235$). Calls in shades of green (67.2%) suffered a clear change in highlighting behavior. Calls in white need to be further investigated.

Other contextual variables, such as companion during a call (variable *Call with*) and location at the time of the call (variable *Call where*), did not show significant associations with highlighting a certain party's turns (i.e., caller's turn, callee's turn, or both turns). In the next section, we study time-based changes in annotation behavior to address *RQ*4.

### 4.4. Addressing RQ4: Temporal Effect on Highlighting Needs

As described in Section 3, the effects of time on highlighting behavior were studied using a subset of calls ($N = 235$) that participants revisited during $P2$. From this set, 37.9% were neither highlighted in the first phase of the study ($P1$) nor in the second phase ($P2$), meaning that participants did not change their highlighting behavior for these calls. On the other side:

—14.9% of the calls were highlighted in $P1$ but not in $P2$. This decision was justified by the participants in most cases as due to a temporal effect in the original highlighting needs (e.g., participant 54: *this event already happened*).
—14.5% of the calls were not initially highlighted in $P1$ but were later highlighted in $P2$. These calls were later considered to have useful information, thus indicating that users' first intention about discarding information exchanged in mobile phone calls can also be affected by time.

These results alone corroborate our fourth research question given that almost one third of the calls had their highlights drastically changed as an effect of time (29.4%; Figure 4). Nevertheless, it is worth inspecting calls with highlights in both $P1$ and $P2$ (32.8%, $N = 77$) to investigate differences as a result of the influence of time.

*4.4.1. RQ4.1. Does the Topic and Amount of Information in Highlights Change Across Time?* As explained in Section 3, differences between highlighted information in $P1$ and $P2$ were categorized, and we then used these categories to calculate the proportion of highlights in $P2$ that differed, in terms of their topic, from highlights in $P1$. Our findings reveal that 6.38% of the calls had highlights with different topics in $P1$ and $P2$, 10.64% of calls had highlights with the same topic but more information in $P2$, and 10.21% of calls also had highlights with the same topic but more information in $P1$. On the other hand, only 5.53% of the calls kept the same amount of information in highlights made during both phases of the study (Figure 5).

These results indicate that 56.6% of the calls (133 calls) used for studying temporal effects were approached differently by participants in $P2$, either by highlighting different pieces of information, by highlighting information in calls that were initially considered not to have valuable content, or by not highlighting the calls anymore. Conversely, only 43.4% (102 calls) of the calls received the same highlighting approach in

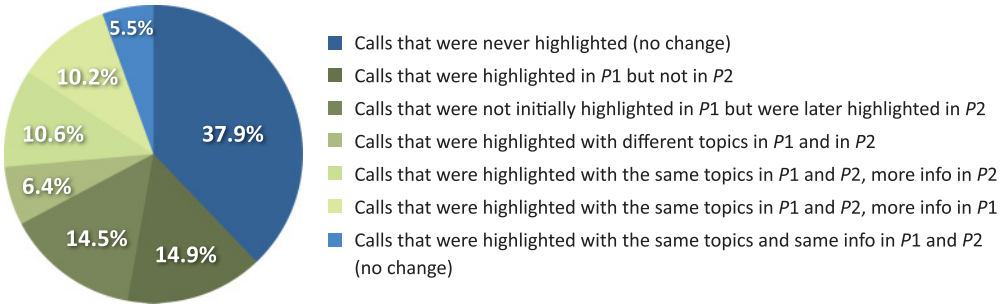



Fig. 5. Overview of changes on highlighting behavior between $P1$ and $P2$. Calls in shades of green (56.6%) were approached differently by the participants, whereas those in shades of blue (43.4) suffered no change.

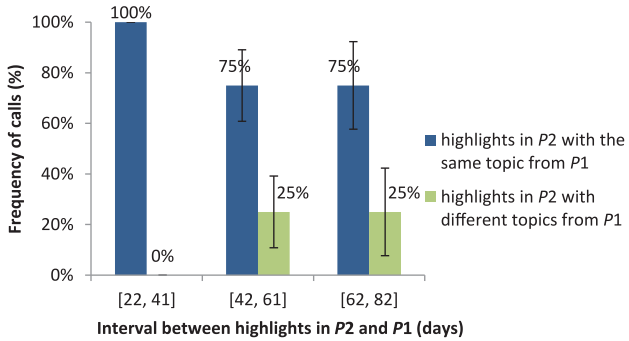



Fig. 6. Change of topics for highlights made in $P2$ compared to highlights in $P1$. Error bars indicate a 95% confidence interval.

both phases, because they were never highlighted or because the highlighted information was the same.

More than half of the mobile phone calls involved in the study suffered a change in their highlights after time passed. The forthcoming research questions are aimed to further investigate and understand this phenomenon. To answer them, we considered the time difference between highlights in $P1$ and highlights in $P2$. This difference spanned from a minimum of 22 days to a maximum of 82 days. We grouped calls into temporal bins according to these time differences. We tried different numbers of bins (two bins of 31 days each, three bins of 20 days each, etc.), and our results were consistently similar for all combinations. Therefore, we opted to use three bins since they provided enough temporal detail to understand highlighting dynamics and increased statistical power for data analyses across bins (three bins have more data points per bin compared to four or more bins). Hence, $Bin1$ contained calls with an interval of 22 to 41 days between their highlights in $P1$ and $P2$ ($N = 54$), $Bin2$ contained calls with an interval of 42 to 61 days ($N = 105$), and $Bin3$ contained calls with an interval of 62 to 82 days ($N = 76$). We used this grouping strategy to address the next research questions.

*4.4.2. RQ4.2. How Does This Information Change Unfold Across Time?* In Section 3, we explained how highlights in $P1$ and $P2$ were manually compared to identify whether highlights preserved the same topic. Figure 6 shows the proportion of calls for which highlighted information changed topics in $P2$ compared to $P1$ for each of the three temporal bins. Whereas none of the highlights in $Bin1$ changed topic between $P1$ and $P2$, 25% of the highlights in the following temporal bins ($Bin2$ and $Bin3$) did change

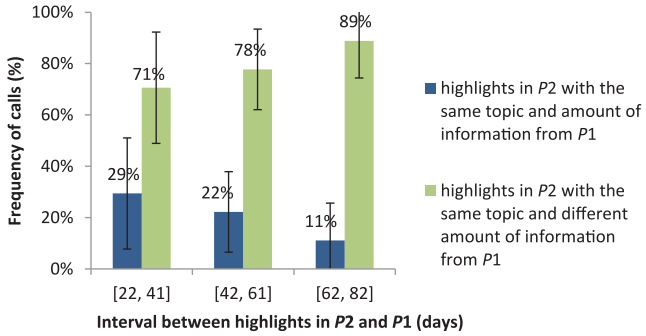



Fig. 7. Change of amount of information highlighted in $P2$ compared to $P1$ for call highlights with the same topic. Error bars indicate a 95% confidence interval.

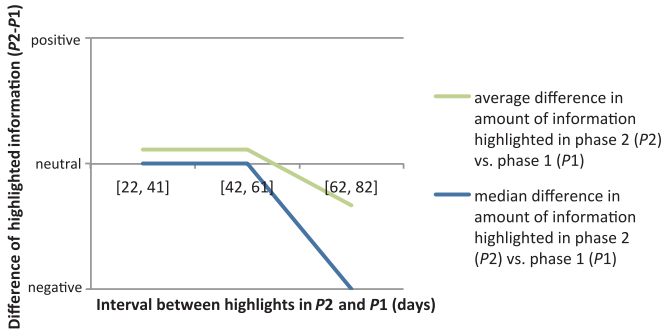



Fig. 8. Difference between the amount of information highlighted in $P2$ compared to the amount of information highlighted in $P1$. Participants made fewer highlights in $P2$ compared to $P1$ when these phases were separated by 62 to 82 days ($Bin3$).

topic. This suggests that time affected participants' highlighting behavior in a way that after 42 or more days, they found some pieces of important information that were thematically different to the ones they originally highlighted.

We further looked into those calls that preserved the same topics in their highlights in both study phases ($N = 62$) to verify if the change in *amount* of highlighted information could also be explained by a temporal effect. Figure 7 shows the proportions of calls with the same topic and amount of information in $P1$ and $P2$ versus calls with the same topic and different amount of information in $P1$ and $P2$ ($p < .05$). It shows that the former decreases and the latter increases across bins—that is, across time. This means that participants had the tendency to change the amount of highlighted information more often for calls that were highlighted later in time.

To better understand how this change happens for calls that kept the same topic in their highlights during both study phases, we looked at their difference in amount of information between highlights from $P2$ and highlights from $P1$ (Figure 8). During the first two time bins (i.e., when $P2$ and $P1$ are separated by 22 to 61 days), the same amount of information is highlighted in $P2$ compared to $P1$. However, in $Bin3$, apparently less information is highlighted in $P2$ for this set of calls.

So far we have looked into how changes of topic and amount of information for calls highlighted in both study phases evolved across time. Next we investigate if the perceived importance of highlights also changes as time passes, and later the reasons for all of these changes. Whereas *RQ4*, *RQ4.1*, and *RQ4.2* were based on the analysis of

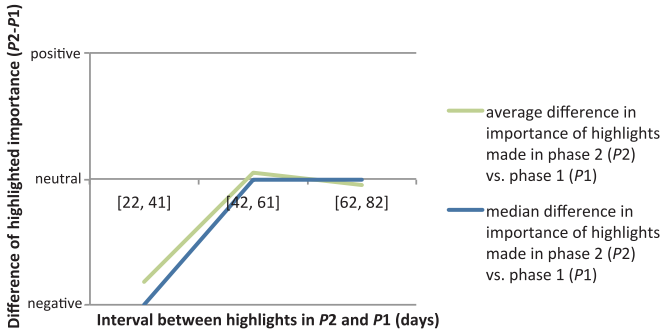



Fig. 9. Difference between the importance of highlights made in $P2$ compared to the importance of highlights made in $P1$. The importance of $P2$ highlights is lower than the importance of $P1$ highlights in $Bin1$ (22 to 41 days), and it increases later on.

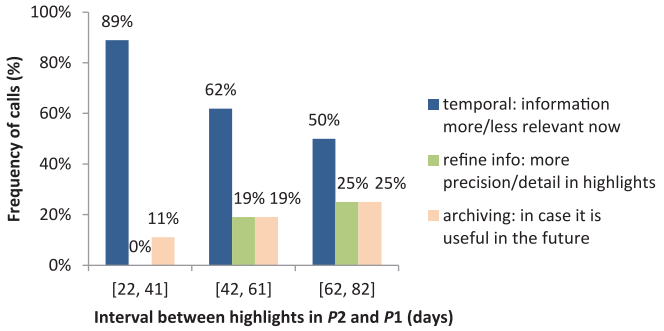



Fig. 10. Reasons for changing highlighting strategies in $P2$ compared to $P1$.

the highlighted information, the next two research questions leverage the participant's self-reported perceptions of these changes.

*4.4.3. RQ4.3. Does the Importance of the Highlighted Information Change across Time?* Figure 9 shows how the importance of highlights made in $P1$ changed for highlights made in $P2$, as reported by the participants. For $Bin1$, the increase in importance is negative, meaning that highlights made about a month after the first highlighting phase were perceived to be less important. For longer intervals between highlights, they were perceived to have an importance similar to that of the previous bin ($Bin2$ and $Bin3$: 42 to 82 days). The change of importance observed in $Bin1$ was significantly different than the change of importance observed in $Bin2$ ($Z = -2.698$, $p < .01$) and in $Bin3$ ($Z = -2.823$, $p < .01$). Conversely, the change of importance observed in $Bin2$ was not significantly different than the one observed in $Bin3$ ($Z = -.281$, $p = .78$).

*4.4.4. RQ4.4. What Are the Reasons Presented by Participants for the Change in Their Highlighting Behavior?* We studied participants' answers to our question regarding *why* their annotation behavior changed, when it was the case. As explained in Section 3, participants' explanations for the differences—if any—between recent highlights in $P2$ and older highlights in $P1$ were manually categorized into a few reasons. We found that time between highlights seemed to influence these explanations. Figure 10 shows how the reasons evolved across time. While reasons related to *temporal effect* tend to decrease over time, participants cited increasingly more reasons related to *archiving* highlighted information and *refining* highlights to explain the changes of highlighting.

*4.4.5. RQ4.5 How Can Annotation Behavior Change Be Explained in Terms of All Previously Investigated Factors?* From the calls studied during both study phases ($N = 235$), 56.6% revealed a change in participants' annotation behavior—that is, they highlighted calls that were not highlighted before, they did not annotate previously highlighted calls, or the information they highlighted was different between study phases. Next, we describe and discuss the possible causes for this change over time.

About 1 month after the first annotation (*Bin*1: from 22 to 41 days), all highlighted calls were related to the same topic (see Figure 6). This result might indicate that participants were still highlighting information based on the *original purpose of the call*. By the time they answered call questionnaire in $P2$, probably highlighted information was already consumed, thus explaining why highlights made in $P2$ were considered to be less important than those taken in $P1$ (Figure 9). In addition, for 89% of these calls, differences between highlights in $P1$ and $P2$ were justified to be due to a temporal effect. In other words, while still highlighting based on the original purpose of the call, participants perceived the new highlights to be less important because the old ones were already consumed, and due to a temporal effect, they were not important anymore (e.g., participant 54: *this event already happened*). This seems to agree with findings from Kalnikaité and Whittaker [2007] for the context of work meetings, in which pen and paper notes were found to be missing important information after around 30 days.

Up to 2 months after the first annotation (*Bin*2: from 42 to 61 days), participants started to change their annotation behavior by *annotating new topics* beyond the original purpose of the call (see Figure 6). By considering the newfound topics worthwhile, the overall valuation of highlights increased with respect to the previous bin: whereas in *Bin*1 the valuation of highlights taken in $P2$ was depreciated with respect to the valuation given to highlights made in $P1$, in *Bin*2 both valuations were similar, revealing an appreciation for highlights made in $P2$ in the midterm (see Figure 9). A possible explanation for this is found in Figure 10, which shows that participants' annotation strategies changed from addressing immediate needs to recording information that might be useful in the long term (e.g., participant 23: *Because it is important to remember the name to be able to search information about it whenever it is possible*) or refining previously highlighted information (e.g., participant 18: *I've highlighted more details*). Although our results are related to daily mobile phone calls rather than work meetings, they might shed light on the reasons previous work has found that meeting participants try to keep their notes accurate after meetings and why half of them even rewrite them [Kalnikaité and Whittaker 2007; Whittaker et al. 2008].

Finally, from 62 to 82 days after the first annotation (*Bin*3), participants' behavior was quite similar to that observed in the previous interval *Bin*2: importance of new highlights in $P2$ was once again the same as highlights in $P1$ (see Figure 9), annotations were again taken beyond the original purpose of the call (see Figure 6), and more often participants considered archiving information for future use rather than addressing an immediate need (see Figure 10). Nevertheless, the amount of information highlighted in $P2$ decreased with respect to information highlighted in $P1$, a behavior not observed in any of the previous intervals *Bin*1 or *Bin*2 (see Figure 8). Therefore, this information is captured in a different way across time (see Figure 7), particularly with fewer details due to a temporal effect in the long term.

From the results discussed herein, we conclude that there is an appreciable effect of time on people's annotation needs of their daily mobile phone calls. Particularly, we found it to be quite interesting that users can depreciate information in the short term and look at them with renewed appreciation in the long term, with the inverse relationship also possible—valuing information only in the short term. These findings suggest the need to identify and archive information that could potentially be useful at

some point in time (either in the short- or long term). Information overload might be reduced by allowing users to hide annotations that are not relevant at a specific point in time.

## 5. IMPLICATIONS FOR DESIGN

Our results suggest the need to create tools to support annotation of mobile phone calls. In this section, we describe some design recommendations that might assist in that goal, aiming to satisfy the requirements of mobile phone users. The section was organized according to the tasks that such an application should perform: recognizing potentially useful notes, annotating them in a way that satisfies users' needs as reported in our study, supporting users in taking advantage of their notes, and finally facilitating the consumption of annotated information.

### 5.1. On What to Annotate

Similarly to recommendations in the context of work meetings [Whittaker et al. 2008], we found that saving complete transcriptions of calls is not an adequate solution: in our study, only 4% of characters in the transcribed calls were annotated. A more efficient approach is to either process the call audio or parse its transcribed text toward automatically identifying potential fragments that the user would be interested in annotating. Deciding which parts of a call should be annotated without consulting the caller's opinion is not a trivial task. Nevertheless, our study revealed a number of *patterns* in mobile phone calls that are usually annotated, such as phone numbers, dates, addresses, prices, shop/to-do lists, contact names, and activities. In fact, we automatically identified these recurrent patterns in participants' calls using simple text parsers—as described in Section 4—and verified that their presence in calls are indeed significantly correlated to whether calls are annotated or not. Other techniques shall be used to automatically annotate patterns without requiring full transcription (e.g., dynamic time warping as applied by Cherubini et al. [2009]).

Our study reveals that given the collaborative nature of mobile conversations, information worth being annotated is not always going to appear in a few phrases of single turns. Instead, it is going to be spread over a number of turns contributed by both participants. For example, whereas questions help to find *when* important information might be about to appear (completion of a question–answer adjacency pair [Schegloff 2007]), annotations should not be focused on answers only. Instead, it should start by analyzing questions to understand call purpose [Freed 1994], then extracting and analyzing meaningful relationships with the corresponding answers. A similar approach extended to other types of relations between turns can lead to the construction of useful notes.

Although our results show that the caller finds information worth being annotated in both parties' turns, some deviations from this behavior were also found. For instance, in informative calls, it was more frequent for callers to highlight information in his or her own turns. If noteworthy information in this type of calls could be obtained by only recording and analyzing the caller's channel, bandwidth, and audio, processing might be reduced with the consequent improvement of QoS by providing faster and/or better results.[8]

Therefore, the annotation application should also take into account *call-derived data*, such as relationship with the interlocutor and call time, as well as leverage embedded sensors to gather relevant *contextual information* for the note-taking activity, such as

[8]Additionally, unintentional privacy breaches could be limited by complying with one-party consent laws, such as those found at http://www.gpo.gov/fdsys/pkg/USCODE-2011-title18/pdf/USCODE-2011-title18-partI-chap119-sec2511.pdf.

the objective of the call (e.g., to determine if a call fulfills only social purposes or if it was made to get or receive specific pieces of information). By identifying call context, call QoS parameters—via analysis of the microphone signal, and patterns in the calls, the need to annotate a phone call might be detected and potential annotations inferred.

### 5.2. On How to Annotate

The majority of participants in our study reported using their mobile phones to annotate information exchanged in daily phone calls. Although our sample was composed of only smartphone users, these devices are becoming predominant worldwide[9] and should be considered one of the most relevant annotation media for daily phone conversations. They are also very convenient given that users always have them nearby during a phone call, whereas other annotation artifacts are not necessarily available at the same time, such as laptops or paper and pen used in the context of work meetings [Bernstein et al. 2008; Whittaker et al. 2008].

Although being more convenient, mobile phones also impose a significant restriction for annotating calls. When making a phone call, one's hands are usually busy, thus preventing annotations to be appropriately taken on-the-fly. In fact, in only 8.8% of all calls made during the study, participants had both hands free for taking notes. Our findings suggest that annotations of phone calls should be better addressed using an offline approach—that is, performed after the call. In addition, the process seems to have potential to be automated given evidence of general nonpersonalized patterns in phone conversations that are usually considered to be relevant (i.e., numbers and names in the form of addresses, phone numbers, codes, prices, contact names, shopping lists). Nevertheless, we found mostly moderate associations between the presence of patterns in phone conversations and annotation of the calls, which indicates that fully automated methods might not achieve high accuracy rates when based mostly on these factors. Further research in audio processing and sentiment analysis might reveal other important sources of information that shall enable fully automated annotation tools.

Given all of the aforementioned recommendations, we conclude that a *semiautomatic annotation approach* to phone call annotation is required. As mentioned previously, important information patterns should be automatically detected, extracted, and stored immediately after the call. This would reduce the information overload that browsing through complete call recordings or transcriptions would mean for the users. On the other hand, we observed that only a fraction of the automatically identified patterns were actually annotated by participants in our study (e.g., 46.6% of numbers in calls that had numbers annotated, 42.5% of names in calls that had names annotated). That said, users should also have the option to manually inspect these preannotated patterns and approve those of particular interest to them, thus reducing overload in future recall tasks. The semiautomatic approach should also enable user-derived notes.

### 5.3. On What to Do with Notes

Phone call annotations can be related to a number of activities and used in different ways. For example, one might annotate details of a doctor's appointment discussed over the phone and—right after the call—transpose the notes to an electronic calendar tool. This suggests that once notes are automatically detected and presented to users, the annotation tool should allow them to take *actions* on the given notes by associating them with the appropriate application, such as creating reminders or appointments

[9]According to the International Data Corporation (IDC), in the first quarter of 2014, a total of 281.5 million smartphones were shipped worldwide (press release: http://www.idc.com/getdoc.jsp?containerId=prUS24823414).

in the phone's calendar tool and saving phone numbers and email addresses in the contacts list.

Other possible actions that may be taken on notes are related to sharing. Given the usual two-person setting of most phone calls, we foresee the opportunity to explore collaboration between caller and callee for the note-taking activity. Collaboration can also be implemented to connect third parties mentioned during calls. According to data from our study, 40% of all annotated calls mentioned either the caller's or callee's relatives, friends or colleagues. The possibility to share (by email, tweets, SMS, etc.) information extracted from mobile phone calls in an effective way might lead to the design of innovative collaborative tools.

### 5.4. On How to Consume Notes

Our findings reveal strong influence of time in people's phone call annotation needs. Although users initially annotated information related to the purpose of the call and disregarded the remaining information exchanged with the callee, later on they considered the nonimportant pieces of information to be worth remembering. Moreover, what was once said to be important, later on it was not annotated at all. This means that users' first impression about annotating any given call is commonly related to their short-term needs, which does not exclude the remaining information from being useful in the long term. This aspect is notably missed by audio-buffer–based solutions such as those depicted by Hindus and Schmandt [1992] and Hayes et al. [2004]. Although their proposed solutions provide fast access to the most recent segments from a conversation, they do not offer mechanisms for deeper browsing of additional pieces of information that may become more relevant in the long term.

To support these time-based needs, we suggest that mobile phone annotation tools should (1) automatically *record every note candidate* to avoid discarding notes that shall become prominently important in the future and (2) offer to the user the possibility to provide feedback on which of the automatically detected notes are more relevant, so as to highlight them using a *multilayer annotation visualization interface*. One possible implementation of this visualization technique is suggested in the following: notes that users manually select as important should be put in the first layer for privileged retrieval; remaining notes should be stored in the second layer, effectively preserving every annotation for long-term annotation needs (e.g., archiving); and finally the whole call transcription—or recorded audio—could be stored in the third layer (in case such information is available). The whole transcription might be specially useful for providing important contextual information not initially contained in the automatically generated notes, such as inside questions asked during the conversation. When users browse or search notes, first-layer information should be ranked higher and thus presented before second-layer notes. If users cannot find the information they are looking for in the first two layers, the third layer could be used. Temporal effect on importance of notes can be further addressed by letting users manually downgrade notes from the first layer whenever they become less important. Similarly, users should also be able to upgrade second-layer notes to reflect their dynamic annotation needs.

Another interesting result on the temporal effect of notes is the change of annotation strategies over time. We observed that notes taken about 1 month (22 to 41 days) after the first annotation still focus on the original purpose of the call, whereas beyond this point people might annotate other pieces of information not previously considered to be important. That said, the proposed multilayer interface should leverage this finding and provide awareness on annotations "time-to-live" by decreasing the emphasis on the first-layer notes and/or increasing the emphasis on the second-layer notes as time passes. A sudden switch of layers should be avoided, as the majority of users tend to keep annotating information related to the original purpose of the call.

Alternatively—and more appropriately—the interface could keep the relevance ordering of layers while using special cues for the temporal effect. An example of implementation is to attach a thermometer-like indicator to the first-layer notes to inform how "fresh" they are with respect to notes in the second layer. Other examples include greying first-layer notes and highlighting access to second-layer notes.

We consider that to facilitate the eventual retrieval of notes, the preservation of additional contextual information is of great importance. As pointed out by Lansdale [1988], the interpretation of the context where information arises greatly influences our ability to remember it. Metadata related to the calls, such as party name, time, and location at the time of the call, which are easily gathered by modern smartphones, can assist in the retrieval of important information.

As a summary, we propose that a semiautomatic phone call annotation method, implemented by means of a mobile phone app developed after our design implications, should perform these steps: (1) storing the full call transcript, (2) automatically detecting candidate notes and presenting them to the user, and (3) letting users fine-tune these notes if they want to. The app could present notes by means of a multilayer annotation visualization interface, as described in the previous subsections. Our ongoing work leverages findings presented herein toward investigating how mobile technology can best support the annotation of daily mobile phone calls.

## ACKNOWLEDGMENTS

We thank Diego Urdiales, Isabel Ordas, Miguel Angel Santiago, Sergio Romero, Edgar Olivares, Martin Pielot, José San Pedro, Nuria Oliver, Josep Blat, and all participants of the study.

## REFERENCES

Ilkka Arminen and Minna Leinonen. 2006. Mobile phone call openings: Tailoring answers to personalized summonses. *Discourse Studies* 8, 3, 339–368.

Michael Bernstein, Max Van Kleek, David Karger, and Monica Schraefel. 2008. Information scraps: How and why information eludes our personal information management tools. *ACM Transactions on Information Systems* 26, 4, 1–46. Available at http://portal.acm.org/citation.cfm?id=1402263&dl=GUIDE&coll=GUIDE&CFID=97795080&CFTOKEN=56870801.

Antje Bothin and Paul Clough. 2012. Participants' personal note-taking in meetings and its value for automatic meeting summarisation. *Information Technology and Management* 13, 1, 39–57. DOI:http://dx.doi.org/10.1007/s10799-011-0112-7

Antje Bothin, Paul Clough, and Gunilla Bradley. 2010. Quantitative analysis of individual differences in note-taking and talking behaviour in meetings. In *Proceedings of the IADIS Multi Conference on Computer Science and Information Systems*. Available at http://ir.shef.ac.uk/cloughie/papers/IADIS-2010.pdf.

Juan Pablo Carrascal, Rodrigo de Oliveira, and Mauro Cherubini. 2012. A note paper on note-taking: Understanding annotations of mobile phone calls. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI'12)*. ACM, New York, NY, 21–24. DOI:http://dx.doi.org/10.1145/2371574.2371578

Mauro Cherubini, Xavier Anguera, Nuria Oliver, and Rodrigo de Oliveira. 2009. Text versus speech: A comparison of tagging input modalities for camera phones. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI'09)*. ACM, New York, NY, Article No. 1. DOI:http://dx.doi.org/10.1145/1613858.1613860

C. J. Clopper and Egon S. Pearson. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26, 4, 404–413.

Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Earlbaum Associates, Hillsdale, NJ.

Brian J. Cragun. 2009. Telephony annotation services. US Patent. US 7602892 (October 13, 2009). http://www.prior-ip.com/patent/71213098/.

Alice F. Freed. 1994. The form and function of questions in informal dyadic conversation. *Journal of Pragmatics* 21, 6, 621–644.

Werner Geyer, Heather Richter, and Gregory D. Abowd. 2005. Towards a smarter meeting record–capture and access of meetings revisited. *Multimedia Tools and Applications* 27, 3, 393–410. DOI:http://dx.doi.org/10.1007/s11042-005-3815-0

Gillian R. Hayes, Shwetak N. Patel, Khai N. Truong, Giovanni Iachello, Julie A. Kientz, Rob Farmer, and Gregory D. Abowd. 2004. The personal audio loop: Designing a ubiquitous audio-based memory aid. In *Mobile Human-Computer Interaction—MobileHCI 2004*. Lecture Notes in Computer Science, Vol. 3160. Springer, 168–179.

Ken Hinckley, Shengdong Zhao, Raman Sarin, Patrick Baudisch, Edward Cutrell, Michael Shilman, and Desney Tan. 2007. InkSeine: In situ search for active note taking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, New York, NY, 251–260. DOI:http://dx.doi.org/10.1145/1240624.1240666

Debby Hindus and Chris Schmandt. 1992. Ubiquitous audio: Capturing spontaneous collaboration. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work (CSCW'92)*. ACM, New York, NY, 210–217. DOI:http://dx.doi.org/10.1145/143457.143481

Ian Hutchby and Simone Barnett. 2005. Aspects of the sequential organization of mobile phone conversation. *Discourse Studies* 7, 2, 147–171.

Biing Hwang Juang and Laurence R. Rabiner. 1991. Hidden Markov models for speech recognition. *Technometrics* 33, 3, 251–272.

Vaiva Kalnikaité and Steve Whittaker. 2007. Software or wetware? Discovering when and why people use digital prosthetic memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. ACM, New York, NY, 71–80. DOI:http://dx.doi.org/10.1145/1240624.1240635

Matthew Kam, Jingtao Wang, Alastair Iles, Eric Tse, Jane Chiu, Daniel Glaser, Orna Tarshish, and John Canny. 2005. Livenotes: A system for cooperative and augmented note-taking in lectures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*. ACM, New York, NY, 531–540. DOI:http://dx.doi.org/10.1145/1054972.1055046

Max Van Kleek, Wolfe Styke, and David Karger. 2011. Finders/Keepers: A longitudinal study of people managing information scraps in a micro-note tool. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*. ACM, New York, NY, 2907–2916. Available at http://portal.acm.org/citation.cfm?id=1979374.

Mark W. Lansdale. 1988. The psychology of personal information management. *Applied Ergonomics* 19, 1, 55–66.

Eric Laurier. 2001. Why people say where they are during mobile-phone calls. *Environment and Planning D: Society and Space* 19, 4, 485–504.

Seung-Hee Lee. 2009. Extended requesting: Interaction and collaboration in the production and specification of requests. *Journal of Pragmatics* 41, 6, 1248–1271.

Min Lin, Wayne G. Lutters, and Tina S. Kim. 2004. Understanding the micronote lifecycle: Improving mobile support for informal note taking. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, 687–694. Available at http://doi.acm.org/10.1145/985692.985779

Kenton O'Hara, Mark Perry, Abigail Sellen, and Barry Brown. 2002. Exploring the relationship between mobile phone and document activity during business travel. In *Wireless World*. Springer-Verlag, New York, NY, 180–194. Available at http://dl.acm.org/citation.cfm?id=510801.510817

Annie Piolat, Thierry Olive, and Ronald T. Kellogg. 2005. Cognitive effort during note taking. *Applied Cognitive Psychology* 19, 3, 291–312. DOI:http://dx.doi.org/10.1002/acp.1086

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 4, 696–735.

Jerrold M. Sadock and Arnold M. Zwicky. 1985. Speech act distinctions in syntax. *Language Typology and Syntactic Description* 1, 155–196.

Emanuel A. Schegloff. 2002. Beginnings in the telephone. In *Perpetual Contact: Mobile Communication, Private Talk, Public Performance.* Cambridge University Press, New York, NY, 284.

Emanuel A. Schegloff. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis*, Vol. 1. Cambridge University Press, New York, NY.

Alexandra Weilenmann. 2003. "I can't talk now, I'm in a fitting room": Formulating availability and location in mobile-phone conversations. *Environment and Planning A* 35, 1589–1605.

Jack Whalen, Don H. Zimmerman, and Marilyn R. Whalen. 1988. When words fail: A single case analysis. *Social Problems* 35, 335.

Steve Whittaker, Patrick Hyland, and Myrtle Wiley. 1994. FILOCHAT: Handwritten notes provide access to recorded conversations. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'94)*. ACM, New York, NY, 271–277. DOI:http://dx.doi.org/10.1145/191666.191763

Steve Whittaker, Simon Tucker, Kumutha Swampillai, and Rachel Laban. 2008. Design and evaluation of systems to support interaction capture and retrieval. *Personal and Ubiquitous Computing* 12, 3, 197–221. DOI:http://dx.doi.org/10.1007/s00779-007-0146-3

Lynn D. Wilcox, Bill N. Schilit, and Nitin Sawhney. 1997. Dynomite: A dynamically organized ink and audio notebook. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'97)*. ACM, New York, NY, 186–193. DOI:http://dx.doi.org/10.1145/258549.258700

World Bank. 2013. ITU Releases Latest Global Technology Development Figures. Retrieved December 22, 2014, from http://www.itu.int/net/pressoffice/press_releases/2013/05.aspx#.VJgyWI4CCA.

Received June 2013; revised October 2014; accepted November 2014