Speech Acts-Enhanced SIP Communication

Zohair Chentouf

Department of Computer Science,
King Saud University, KSA
zchentouf@ksu.edu.sa

Abstract

In workplaces, phone calls constitute a very common way of communication. However, the callee typically has no means to know what the caller wants to talk about. This can be counterproductive if the callee is busy performing a task that is more important and/ or urgent than the conversation that is about to be initiated by the caller. The present article addresses this issue and comes up with a new enhanced communication service. First, a communication model based on speech acts is proposed. Speech acts are used by the caller to indicate the call intention or subject to the callee in the most clear and unambiguous possible way. Then, an Internet telephony implementation of the speech acts-based model is designed to fit the Session Initiation Protocol (SIP). The solution is validated through a case study.

Keywords: *IP-PBX, SIP, speech acts, user preferences*

1. Introduction

Internet Telephony offers to workplaces voice communication products called IP-Private Branch Exchange (IP-PBX), which constitute an alternative to the traditional circuit-switched enterprise phone systems, i.e. PBX. This technology is attractive for workplaces mainly because it brings lower infrastructure costs over conventional telephony systems and it offers more flexible and richer multimedia communications that combine instant messaging, presence, voice calls, video calls, and conferencing. Another advantage of Internet telephony is the flexibility with which workers can communicate with their colleagues while they are at home or on the road like if they were in office.

Currently, most of the Internet telephony systems are based on the Session Initiation Protocol (SIP) [1] from the Internet Engineering Task and Force (IETF). The flexibility of SIP has allowed the IETF to design the presence service [2, 3] that is currently available with multimedia communication services in most of the SIP systems. The presence service allows the user to express his ability and willingness to communicate with other users on the network. For that, the user uses his communication device to set his presence status to be for example "busy", "out-for-lunch", "on-vacation", etc. In Internet Telephony, the presence service is usually combined with voice communication services. In this case, the presence status of a subscriber is stored and made available for the voice communication server in order for the latter to route the user's incoming calls depending on the presence status he has set. For example, if the presence status is "busy", the server will intercept the call and send back a "busy" signal to the caller's device.

1.1. Research Motivation

Phone calls constitute a very common way of communication in workplaces. However, the callee cannot know what the caller wants to talk about. This can be counterproductive if the

callee is busy performing a task that is more important and/ or urgent than the conversation that is about to be initiated by the caller. For example, suppose user A is the IT and software manager in a company. He is in office and is very busy preparing a product demonstration for important customers who are coming after half an hour. User B, a software tester, calls user A. User A's phone rings and displays the number of B. User A does not know what B wants. There are two possible scenarios. In the first one, A answers the call and B starts describing a technical problem he is facing. The problem is critical but less important than the demonstration. So, user A orders him to report the problem to the network administrator. In the second scenario, user B reports that the server which user A intends to use to demonstrate the product during the meeting with the customers is not available because of a hardware failure and that they need at least one hour to move data and re-install the software on another server. In such a case, user B's call is really important for user A. As this example suggests, it can be crucial for the callee to know the exact subject or intention of the call.

1.2. Research Plan

The present article addresses this issue and comes out with a new enhanced communication service. First, a communication model based on speech acts is proposed. Speech acts are used by the caller to indicate the call intention or subject to the callee in the most clear and unambiguous possible way. Then, an Internet telephony implementation of the speech acts communication model is designed to fit the Session Initiation Protocol (SIP) and the SIP-based presence service. To validate the solution, a case study is conducted.

Section 2 presents the proposed communication model based on speech acts. A SIP implementation of the communication model is described in Section 3. Section 4 contains a detailed case study. Section 5 overviews the related work. Section 6 concludes the article.

2. Speech Acts-based Communication Model

2.1. Speech Act Theory

The Speech Act Theory has been introduced by Austin [4] in the beginning of the 1960's then refined by Searle during mid 1970's. According to these authors, utterances of a given language statements represent speech acts. For example, "I promise", "I apologize", etc., are utterances that convey an action. Searle [5] classified all speech acts into five categories:

- Assertives: information about some state of affairs. Example: "It is snowing".
- Directives: the speaker requests the hearer to perform an action. Example: "Please, bring me a tea".
- Commissives: the speaker commits to perform some action. Example: "I promise to meet them today".
- Expressives: the speaker expresses his attitude about some state of affairs. Example: "I like ice cream".
- Declaratives: the speaker reports some change of state of affairs by the speech act. Example: "I hereby pronounce you husband and wife".

Other authors proposed more specific speech act models. Examples are Conversion for Action [6], Language-Action Perspective (LAP) [7], and Dynamic Essential Modeling of Organizations (DEMO) [8].

2.2. Communication Model

Applying the Speech Act Theory in modeling real communication situations requires adapting them to the specific problem's domain. This implies choosing suitable speech acts so they express real actions. In order to establish specific speech acts for our research purpose, an analysis of emails of a software development company has been performed. The methodology here has been based on the assumption that if user A needs to communicate with user B, user A chooses a phone call instead of an email based only on whether he wants an immediate reaction from user B, not on the content of the communication. In other words, phone calls and emails have the same content. The sole difference is that phone calls are used to cause an immediate effect.

Table 1. Proposed Speech Acts

Searle's Speech Act Categories	Proposed Model's Speech Acts
Assertives	Feedback-to-you
	Inform-you
	Improvement
	Remind-you
	Personal
Directives	Feedback-to-me
	Coordination
	Assign-you
	Authorize-me
	Help-me
	Ask-you
	Personal
Commissives	Help-you
	Personal
Expressives	Concern
	Personal
Declaratives	Announcement

Based on the analysis of emails, the following domain specific speech acts have been derived. In Table 1, they are classified under their corresponding Searle's speech act categories.

- Feedback-to-you: "as reply to your request"; "your testing server is ready".
- Feedback-to-me: "what is the progress status on that task?"
- Inform-you: "server down"; "I installed a new version for you".
- Ask-you: "what is the IP of that server?"
- Help-me: "I am stuck with this bug"; "I need a new build".
- Help-you: "I can update your database"; "do not restart the switch because I already did it".
- Assign-you: "new task for you"; "server down, restart it".
- Remind-you: "meeting at 2 pm"; "you did not send your report".

- Concern: “I am afraid the overnight test will fail”.
- Coordination: “let us meet now”.
- Authorize-me: “need to leave early today”.
- Improvement: “I have a new design idea”; “I have a new OS version”.
- Announcement: “I consider version 2.5.32 as GA candidate”.
- Personal: “I invite you to dinner tonight”.

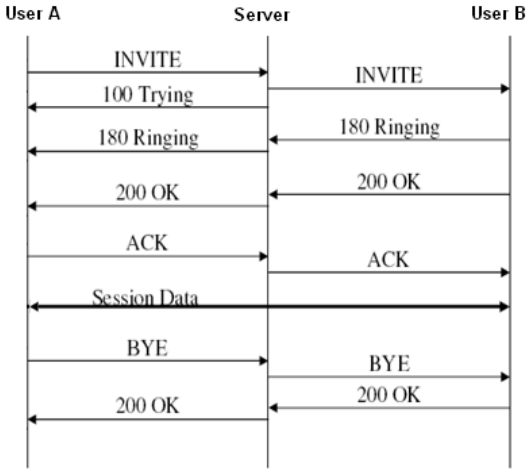

Figure 1. SIP Call Message Sequence

3. SIP Implementation

SIP calls are initiated by INVITE messages. The server receives the INVITE, processes it, then routes it to the next hop. In Figure 1, the call is routed to user B. A part of an INVITE message content is presented in Figure 2.

3.1. Implementation Requirements

The implementation of the proposed communication model requires the following:

- The communication terminals should be suitably equipped so they allow callers to choose a speech act among those listed in Table 1 just before dialing a number.
- The communication protocol should convey the speech act sent by the caller.
- The communication terminals should also be able to display the speech act received from the caller while ringing.

The second requirement is straightforwardly implementable using SIP. Indeed, the latter has been designed to include a “Subject” header field that provides a summary of the call purpose [1]. We propose to use this header to convey speech acts. This does not interfere with any other SIP feature that would use the “Subject” header since, at the best of our knowledge, current SIP phones do not allow the caller to fill in this header.

```
INVITE sip:bob@biloxi.com SIP/2.0
Via: SIP/2.0/UDP pc33.atlanta.com;branch=z9hG4bKkjshdyff
To: Bob <sip:bob@biloxi.com>
From: Alice <sip:alice@atlanta.com>;tag=88sja8x
Max-Forwards: 70
Call-ID: 987asjd97y7atg
CSeq: 986759 INVITE
```

Figure 2. Part of INVITE Message Content [1]

```
1: status = in-a-meeting -> redirect to voicemail
2: status = on-road -> redirect to 0565555288
3: status = on-vacation -> redirect to 325
```

Figure 3. User Call Routing Preference Rules

3.2. SIP Implementation

In IP-PBX, SIP call processing is integrated with the presence service [2–3]. The latter allows users to set their status as "online", "on-the-phone", "busy", "out-for-lunch", "in-a-meeting", etc. Users can specify their call routing preferences based on the presence service. Examples are depicted by Fig. 3. Consequently, when the SIP server receives a call, say from user A to user B, it loads the preference rules of user B and uses user B's current presence status to infer a routing action. If no routing action is derived, the SIP server simply sends the call to user B. In this case, user B's terminal will display the speech act sent by user A. As example, suppose user B's preference rules are those of Fig. 3 and the presence he has set just before leaving his office is "on-road". Any incoming call to his office will be intercepted and redirected by the SIP server to 0565555288 (line 2, Fig. 3), which is his cell number.

The second requirement (Section 3.1) encompasses two sub-requirements:

- Users use speech acts and presence status to specify their call routing preference rules (instead of presence status only).
- The SIP server triggers preference rules using the callee's presence status and the caller's "Subject" header content (instead of presence status only).

For example, user B may have specified preference rule 2 (Fig. 3) as:

status = on-road AND subject = coordination -> redirect to 0565555288

In such a case, the user A's call will be routed to user A's cell only if user A has chosen "coordination' as speech act before dialling user B's number. Otherwise, the call will be sent to user B's office phone, which will ring a certain number of times then it will be routed to the voicemail server, as it is the most common service configuration in IP-PBX (i.e. to forward unanswered calls to the voicemail).

Table 2. Author's Emails Analyzed

Speech Act	Occurrence
Feedback-to-you	61
Feedback-to-me	5
Inform-you	33
Ask-you	8
Help-me	11
Help-you	2
Assign-you	9
Remind-you	7
Concern	3
Coordination	13
Authorize-me	2
Improvement	5
Announcement	7
Personal	11

4. Case Study

Despite the fact that the here specified communication model has been derived by analyzing emails of a specific software development company, we conjecture that it encompasses the most common and important work interaction subjects that are used in any workplace. The aim of this section is to validate the communication model and the SIP call processing here proposed. To validate the communication model, an analysis of the author's business emails received during one month has been done. Those emails convey communication that pertains to an academic environment. Table 2 summarizes the results. It shows the completeness of the model, since no missing speech act has been detected.

The validation of the call processing has been performed through an implementation on an open source SIP server developed by NIST [9]. The usual processing of a SIP call consists in the following ordered steps:

- Parse the received SIP packet (UDP in our case) in order to extract the values of some headers, mainly "Request-URI", "From" (caller's address), "To" (callee's address), and "Call-Id" (see Fig. 2).
- Load the preferences of "To" to infer the address to which the call has to be routed.
- Route the call.

Handling the speech acts-enhanced service simply adds extracting the "Subject" header and comparing its value with the condition parts of the callee's preference rules. SIP call load tests have been conducted using two Windows XP computers with 1 GB RAM and 2.8 MHz CPU. One computer has roomed a simple SIP call generator developed in Java by the author. The second contained the NIST SIP server [9].

A first test has aimed at determining the maximum load capacity that the SIP server can handle with a success rate of 99.999%. The latter value is a well known performance requirement in the telecommunications world. The maximum capacity of the server has been found to be 22000 calls/h. Then, the SIP server has been updated with the speech acts-enhanced SIP routing algorithm, and the same load test has been performed again. The maximum capacity has been of 22000 calls/h. Since the maximum capacity has been the same

in the two tests, we conclude that the enhancement of the SIP routing service with speech acts generates no SIP call processing overhead.

5. Related Work

At the best of our knowledge, there has been no previous work on communicating the caller's intention to the callee or using speech acts in voice communication. That is why, in this section, we will present some research works that dealt with email communication. They are related to our research in two aspects: they have the motivation to enhance communication and they use speech acts.

Given a team of users that finished a common task or project, the objective of [10] is to analyze the emails they exchanged in order to determine the exact role performed by every team member. The analysis is based on speech acts. The main aim of [11] is to compare the performance of two classifiers (Random Forests and Support Vector Machines) in identifying the intent of emails by means of speech acts. In [12], emails are categorized based on the sender's intent, which is derived using a Support Vector Machines classifier and speech acts. The problem addressed in [13] is the one of the email information overload that users have to handle. Authors use the Semantic Web and a speech acts-based ontology to classify emails and support users in their actions of sending and receiving emails. In [14], email classification uses speech acts and dependency networks. The latter are elaborated based on sequential correlations between emails. The authors claim that such sequential correlations do exist and that they can help predicting the intent of email messages. In [15], authors present a system that analyzes emails based on speech acts. The system's output is a summary that focuses on the tasks that the email's sender wants the receiver user to perform. In [16], speech acts are used to annotate workplace emails. Then, a natural language processing tool is used in order to derive features like for example which verb forms are most common in a given speech act. The aim of the research is to pave the way to better understanding the pragmatic dimension of the workplace email communication.

6. Conclusion

In this paper, we proposed a new communication model to enhance IP-PBX services. The model has been based on speech acts. Despite having been deduced using an analysis of a software company's emails, the proposed speech acts have proved to be completely sufficient to express communication intentions and subjects in an academic organization. This leads us to conjecture that the proposed model is suitable for any other type of workplaces. The communication model has also been implemented using SIP to fit the IP-PBX usual services. This required using the SIP "Subject" header to convey the speech act specified by the caller. It also necessitated updating the call processing method at the SIP server level. The main update consisted in integrating speech acts with the presence-based call routing service. SIP call simulations showed that this update generates no significant overhead.

As future work, we intend to study the provisioning of speech acts-enhanced call routing user preferences to the SIP server.

References

[1] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley and E. Schooler, "SIP: Session Initiation", IETF RFC 3261 **(2002)**.
[2] M. Day, J. Rosenberg and H. Sugano, "A Model for Presence and Instant Messaging", IETF RFC 2778 **(2000)**.
[3] J. Rosenberg, "A Presence Event Package for the Session Initiation Protocol (SIP)", IETF RFC 3856 **(2004)**.

[4] J. L. Austin, "How to do things with words", Harvard University Press, Cambridge, **(1980)**.
[5] J. R. Searle, "in Mind and Knowledge", Edited K. Gunderson, Language, Minneapolis, **(1975)**.
[6] T. Winograd and F. Flores, "Understanding Computers and Cognition: A New Foundation for Design", Ablex Norwood, N.J., **(1986)**.
[7] F. Flores and J. Ludlow, "in Decision Support Systems:Issues and Challenges", Edited G. Fick and R.H. Sprague, Pergamon Press, New York, **(1980)**.
[8] J. Dietz, "Understanding and Modeling Business Processes with DEMO", Proceedings of the 18th International Conference on Conceptual Modeling Paris, **(1999)** November, Paris, France.
[9] NIST SIP, available at: http://www-x.antd.nist.gov/proj/iptel/nist-sip-downloads.html, **(2012)** June.
[10] A. Leuski, "Email is a stage: discovering people roles from email archives", Proceedings of the Annual ACM Conference on Research and Developmentin Information Retrieval, **(2004)**, Sheffield, UK.
[11] J. Goldstein and R. E. Sabin, "Using Speech Acts to Categorize Email and Identify Email Genres", Proceedings of the 39th Annual Hawaii International Conference on System Sciences, **(2006)**, Hawaii.
[12] W. Cohen, V. Carvalho and T. Mitchell, "Learning to classify emails into speech acts", Proceedings of the Conference on Empirical Methods in Natural Language Processing, **(2008)**, Barcelona, Spain.
[13] S. Scerri, B. Davis and S. Handschuh, "Improving Email Conversation Efficiency through Semantically Encanced Email", Proceedings of the 18th International Workshop on Database and Expert Systems Applications, **(2007)** September, Regensburg, Germany.
[14] V. Carvalho and W. Cohen, "Improving Email Speech Act Analysis via Ngram Selection", Proceedings of the ACTS Workshop, **(2006),** New York.
[15] S. H. Corston-Oliver, E. Ringger, M. Gamon and R. Campbell, "Task-focused summarization of email", Proceedings of the ACL-06 Workshop: Text Summarization Branches, **(2006)** July.
[16] R. De Felice, "The pragmatics of the workplace: insights from speech act-annotated email corpora", Proceedings of the fifth Inter-Varietal Applied Corpus Studies (IVACS) group International Conference, **(2010)**, Edinburgh, UK.

Author

Zohair Chentouf has been a faculty member at the Department of Computer Science at the King Saud University since 2009. He received his Ph.D. degree in 2005 from the University of Sherbrooke, Canada. With 6 years working in the telecommunications and VoIP industry in Canada, Zohair's research interests are mainly in computer communication software engineering, including VoIP reliability engineering, service engineering, and communication architecting.