

Diss. ETH No. 11600

Towards Computer Supported Audio Conferencing

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of
Doctor of Technical Sciences

presented by

CHRISTOPH ANDREAS BURKHARDT

Dipl. El.-Ing. ETH

born July 2, 1964

citizen of Matzendorf and Härkingen, SO

accepted on the recommendation of

Prof. Dr. B. Plattner, examiner

PD. Dr. H. P. Lubich, co-examiner

1996

Abstract

This work focuses on the use of audio in distributed synchronous communication situations. Whereas in tasks like joint editing, discussing previously distributed documents, or conducting a research meeting, audio capabilities can provide valuable support, there are much more tasks, such as making arrangements and/or decisions, which are perfectly suited for being supported by audio conferencing systems. In comparison to other forms of meetings, audio conferencing can be very efficient and can help to produce high quality output for such electronically conducted meetings. Furthermore, the “social presence” of meetings carried out through audio conferences is an important success factor in academic and business applications alike.

Audio conferencing is only efficient if the right tools are used and if the flow of task-inherent work processes is not disturbed by unsuited technology. Therefore, we suggest to structure the original task into two asynchronous preparation phases, one without and one with a link to the audio conference. This is followed by the actual audio conference and the assessment phase, which ends one iteration of processing the original task.

We designed our audio conferencing system such that it provides a flexible audio conferencing service for such situations. It is based on a client-server model with a central audio conference server providing the actual audio conferencing service. Our conference paradigm supports three roles: chairperson, member, and operator. Due to the use of an out-band signaling channel, each conference participant in either role has instantaneous access to the switching equipment based on a state-of-the-art and intuitively usable graphical user interface. Each member can control a number of personal parameters, while the chairperson can control and set parameters for the whole conference. Private side conversations are supported, and in addition, the chairperson can dynamically form subgroups, as well as reassemble them into one main conference.

The object-oriented audio conference software is written in C/C++ and runs on Sun workstations under SunOS 5.3. For the development of the GUI, DevGuide, the window developers guide from Sun Microsystems, was used. The audio mixing part of our prototype is based on IBM compatible PCs running DOS. The PCs are equipped with ISDN interfaces, an analogue telephone interface, which has been built in our laboratory, a digital signal processor board (DSP), performing the actual switching, as well as Ethernet interface cards allowing network access over TCP/IP.

As a consequence of using the Public Switched Telephone Network for audio trans-

mission and the Internet (TCP/IP based communication) for sending conference control data to and from the audio conference server, we take advantage of two well established, widely accepted and available services. The Public Switched Telephone Network provides audio transmission with at least constant telephone quality and the Internet provides a packet oriented transmission service which is perfectly suited to transfer bursty data such as control packets of our audio conferencing system. As a consequence of that, our audio conferencing service with good audio quality can be fully accessed and controlled over the Internet.

In contrast to our architecture, existing audio conferencing systems either lack a user-friendly interface, or do not provide reasonable audio quality at a reasonable price. Our system proved to be realizable with existing low-cost workstation technology, but additionally provides a good quality audio conference signal and a user-friendly interface. Based on observations of the shortcomings of existing audio conferencing systems, a variety of new, useful end user and control functions is provided by our system. At the time of writing, no other system is known to us that is able to fully support our design requirements.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit den Anwendungsmöglichkeiten von Sprachkonferenzsystemen in Situationen, in denen verteilte, synchrone Kommunikation eingesetzt wird. Sprachkonferenzen können nicht nur in Aufgaben wie dem gemeinsamen Editieren von Dokumenten, dem Diskutieren von bereits verteilten Dokumenten oder dem Durchführen von Forschungsgruppensitzungen einen wertvollen Dienst leisten, sondern sind ausgezeichnet geeignet, um Aufgaben wie das Aushandeln eines gemeinsamen Termins oder Entscheidungsfindungsprozesse gewinnbringend zu unterstützen. Im Gegensatz zu anderen Sitzungsformen können Sprachkonferenzen sehr effizient geführt werden und erreichen demzufolge ein qualitativ hohes Sitzungsergebnis. Im weiteren vermitteln Sprachkonferenzen das Gefühl von "gegenseitiger Verbundenheit", das ein entscheidender Erfolgsfaktor für Sprachkonferenzsysteme ist, sowohl im akademischen als auch im geschäftlichen Anwendungsumfeld.

Sprachkonferenzen sind nur dann effizient, wenn geeignete Unterstützungswerkzeuge benutzt werden und wenn der aufgabeneigene Prozessablauf nicht durch unpassende Technologien behindert wird. Deshalb schlagen wir eine Strukturierung der eigentlichen Aufgabe in vier Teilaufgaben vor. Die ersten beiden Teilaufgaben erfordern asynchrone Bearbeitung, sowohl mit als auch ohne konkreten Bezug zur eigentlichen Sprachkonferenz, die dann im dritten Teilschritt durchgeführt wird. Den Abschluss bildet die Nachbearbeitungsphase, welche die Voraussetzungen für eine neue Iteration schafft.

Unsere Sprachkonferenzanwendung wurde unter dem Gesichtspunkt entworfen, einen flexiblen Sprachkonferenzdienst für die Unterstützung der oben erwähnten Aufgaben zur Verfügung zu stellen. Sie basiert auf einem Client Server-Modell mit einem zentralen Sprachkonferenzserver. Das zugrundeliegende Konferenzparadigma unterstützt drei Rollen: "Chairperson", "Member", und "Operator". Da wir für die Steuerung einen eigenen Signalisierungskanal verwenden, hat jeder Konferenzteilnehmer in einer der erwähnten Rollen die Möglichkeit, die Sprachkonferenzanwendung mittels einer fortschrittlichen, intuitiv zu benutzenden graphischen Benutzerschnittstelle zu bedienen. Während jeder Konferenzteilnehmer persönliche Parameter der Konferenz setzen kann, hat die Konferenzleiterin die Möglichkeit, konferenzweite Einstellungen vorzunehmen. Private "Nebengespräche" (chat) werden genauso unterstützt wie das Bilden von Unterkonferenzen und das Zusammenführen von Unterkonferenzen zu einer gemeinsamen Konferenz.

Die objekt-orientierte Sprachkonferenz-Software ist in C/C++ geschrieben und läuft

auf Sun Arbeitsplatzrechnern unter SunOS 5.3. Die graphische Benutzerschnittstelle wurde mit dem DevGuide, dem "Window Developers Guide" von Sun Microsystems, erstellt. Die Sprachmischung für die Sprachkonferenz basiert auf IBM kompatiblen PCs, die mit dem Betriebssystem DOS 6.2 betrieben werden. Die PCs sind zusätzlich ausgerüstet mit einem analogen Telephon-Interface, das an unserem Institut entwickelt worden ist, mit zwei ISDN-Interfaces, mit einer digitalen Signalprozessor-Karte (DSP), auf der die eigentliche Sprachmischung durchgeführt wird, sowie Ethernet-Interface-Karten, die die Netzwerkverbindung über TCP/IP ermöglichen.

Da wir für die Übertragung der Sprache das öffentliche vermittelte Telefonnetz und für die Übertragung der Konferenzsteuerdaten das Internet verwenden, basiert unser Sprachkonferenzdienst auf zwei bestens eingeführten und weit verbreiteten Diensten. Dabei garantiert das öffentliche vermittelte Telefonnetz die Übertragung von Sprache mit mindestens Telephonqualität und das Internet stellt auf der Basis von TCP/IP einen paketorientierten Transportdienst zur Verfügung, der ausgezeichnet geeignet ist, um die in unregelmässigen Zeitabständen anfallenden Datenpakete zu übertragen. Demzufolge kann unser qualitativ hochstehender Sprachkonferenzdienst über das Internet angeboten und gesteuert werden.

Im Gegensatz zu unserer Architektur leiden bestehende Sprachkonferenzsysteme entweder an einer benutzerunfreundlichen Bedienschnittstelle oder können keine zufriedenstellende Sprachqualität anbieten. Demgegenüber hat die Implementation unseres Prototyps gezeigt, dass es möglich ist, mit heute verfügbaren Mitteln sowohl eine benutzerfreundliche Bedienschnittstelle als auch eine hohe Sprachqualität zur Verfügung zu stellen. Zum heutigen Zeitpunkt sind uns keine anderen Systeme bekannt, die diesen Anforderungen genügen.