

Einfluss von Bandbreite und räumlicher Sprachwiedergabe auf die kognitive Anstrengung bei Telefonkonferenzen in Abhängigkeit von der Teilnehmeranzahl

Janto Skowronek, Alexander Raake

Deutsche Telekom Laboratories, Technische Universität Berlin, Deutschland, Email: janto.skowronek@telekom.de

Einleitung

Telefonkonferenzen mit mehr als zwei Personen werden oft als anstrengend empfunden, da insbesondere das Auseinanderhalten von Sprechern erschwert ist. Frühere Untersuchungen, z.B. [1, 2], haben gezeigt, dass räumliche Sprachwiedergabe und hohe Bandbreite des übertragenden Sprachsignals diese Zuordnung erleichtern können. Die Analyse in [2] wirft jedoch die Frage auf, inwieweit die Anzahl der Teilnehmer den effektiven Nutzen solcher Systemeigenschaften beeinflusst. Daher wurde ein Hörversuch durchgeführt, in dem Versuchspersonen Aufnahmen von Telefonkonferenzen beurteilten, welche bezüglich der Anzahl der Konferenzteilnehmer sowie den Systemeigenschaften Bandbreite und räumlicher Wiedergabe variiert wurden. Dieser Beitrag beschreibt die Versuchsvorbereitungen, in denen Szenarien als Gesprächsgrundlage entwickelt und entsprechende Sprachaufnahmen durchgeführt wurden. Anschließend werden die Versuchsdurchführung erläutert, die Ergebnisse diskutiert und entsprechenden Schlussfolgerungen getroffen.

Testszenarien

Inhaltliche Grundlage der Aufnahmen waren die in der Qualitätsbeurteilung von Telefonverbindungen oft verwendeten strukturierten Testszenarien. Ursprünglich für Zweipersonengespräche entworfen, wurden diese Szenarien für Dreipersonengespräche weiterentwickelt [2]. Sie beinhalten stichpunktartige Informationen, welche die Gesprächspartner austauschen müssen, um zu einem Resultat zu kommen. Die Vereinbarung von Ort und Zeit für eine Besprechung wäre ein Beispiel hierfür.

Um den Einfluss der Anzahl der Gesprächspartner untersuchen zu können, müssen die Testszenarien untereinander vergleichbar bleiben, während die Anzahl der Gesprächspartner variiert wird. Gleichzeitig sollten die Szenarien aber auch eine Skalierung der kognitiven Anstrengung in Abhängigkeit der Gesprächspartneranzahl realisieren, so wie man es bei realen Telefonkonferenzen beobachten kann: je mehr Teilnehmer desto mehr Diskussionen. Daher wurden die Dreipersonenszenarien aus [2] durch systematisches Hinzufügen von Informationseinheiten modifiziert. Mit jedem weiteren Gesprächspartner wurde ein weiterer Tagesordnungspunkt (TOP) hinzugefügt und die Informationen, die in diesem TOP ausgetauscht werden sollten, wurden jeweils anders auf die Gesprächspartner aufgeteilt. Es wurden 13 Szenarien aus dem Geschäftsleben realisiert: Vier Dreipersonengespräche und jeweils drei Zwei-, Vier- und Sechspersonengespräche.

Aufnahmen

Zu den Aufnahmen wurden sechs männliche professionelle Sprecher eingeladen. Als Aufnahmeraum diente ein großer vollseitig reflexionsarmer Raum der Technischen Universität Berlin ($120m^2$ Grundfläche). Die Sprecher wurden mit dem Gesicht zur Wand so platziert, dass ihr Abstand zueinander so groß wie möglich war. Durch diese Sitzanordnung konnten sich die Sprecher nicht direkt sehen und etwaiges akustisches Übersprechen der Mikrofone wurde minimiert. Die Sprecher benutzten hochqualitative Sennheiser Headsets, welche über RME Audiokarten mit zwei Laptops verbunden waren; ein Laptop diente der Aufnahme, der andere als Konferenzsimulation. Für die Konferenzsimulation wurde die SoundScape-Renderer Software [3] verwendet, die mittels binauraler Wiedergabe die Gesprächspartner eines jeden Sprechers gleichmäßig auf einem Halbkreis im virtuell-akustischen Raum verteilte. In einer anschließenden Editierung der Aufnahmen wurde die Vergleichbarkeit der Gespräche untereinander optimiert und deren Länge gekürzt, so dass sie in einem Hörversuch einfacher einzusetzen waren. Ergebnis dieses Prozesses sind 13 Telefonkonferenzgespräche mit einer Gesamtlänge von ca. 97 Minuten.

Versuchsdurchführung

Dreizehn Freiwillige (5 w, 8 m, Alter 26-40) aus den Telekom Laboratories nahmen am Versuch teil. Dieser Personenkreis wurde so ausgewählt, dass alle Versuchspersonen bereits Erfahrungen mit Mehrpersonen-Telefongesprächen gemacht hatten.

Jedem Versuchsteilnehmer wurde ein Dreipersonengespräch in einer Trainingsphase dargeboten; die anderen zwölf Aufnahmen dienten als Teststimuli. Dabei wurden die jeweils drei Zwei-, Drei-, Vier- und Sechspersonengespräche mit drei unterschiedlichen Wiedergabequalitäten verwürfelt. Die Wiedergabequalitäten unterschieden sich in der Frequenzbandbreite und in der räumlichen Wiedergabe; die drei gewählten Konditionen waren Narrowband-Mono (NB-M), Fullband-Mono (FB-M) und Fullband-Spatial (FB-3D).

Der Fragebogen enthielt einen Erinnerungs- und einen Beurteilungsteil. Im Erinnerungsteil kreuzten die Versuchspersonen für zehn Zitate aus dem Gespräch an, welcher von zwei möglichen Sprechern dieses gesagt hatte (*REC*). Und sie gaben an, wie sicher sie sich bei ihrer Antwort waren (*CONF*). Im Beurteilungsteil wurde gefragt, wie schwer es war, zu erkennen, welcher Gesprächspartner spricht (*SPKID*) und wieviel Konzentration es erforderte, um der Konferenz zu folgen (*CONC*).

Ergebnisse

Bild 1 zeigt die Mittelwerte und die 95%-Konfidenzintervalle der vier abgefragten Größen *REC*, *CONF*, *SPKID* und *CONC* in Abhängigkeit der verschiedenen Wiedergabequalitäten (*SysCond*) und Anzahlen der Gesprächspartner (*NumPart*), Tabelle 1 die entsprechenden ANOVA-Ergebnisse. In Bild 1 ist der Ska-

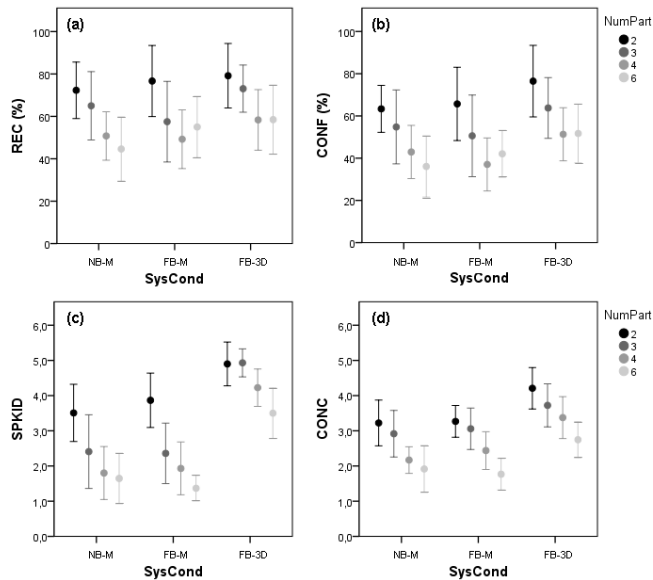


Abbildung 1: Mittelwerte und Konfidenzintervalle für *REC*, *CONF*, *SPKID* und *CONC* in Abhängigkeit von *SysCond* und *NumPart*.

Tabelle 1: ANOVA Ergebnisse für *REC*, *CONF*, *SPKID* und *CONC* in Abhängigkeit von *SysCond* und *NumPart*. Angezeigt sind F-Wert und Signifikanzniveau p .

Variable	Faktor	F	p
REC	SysCond	1,997	0,157
	NumPart	19,754	0,000
CONF	SysCond	6,616	0,005
	NumPart	49,556	0,000
SPKID	SysCond	32,277	0,000
	NumPart	27,482	0,000
CONC	SysCond	14,753	0,000
	NumPart	25,160	0,000

lierungseffekt in Abhängigkeit von der Anzahl der Gesprächspartner innerhalb einer Wiedergabequalitätsstufe deutlich zu sehen. Die ANOVA-Ergebnisse bestätigen dieses durch einen signifikanten Einfluss von *NumPart* auf alle vier Messgrößen. Das heisst, die entworfenen Szenarien erreichen in der Tat den gewünschten Effekt einer zunehmenden kognitiven Anstrengung.

Bezüglich des Einflusses der Wiedergabequalitäten sind die Resultate weniger deutlich. Das Maß *REC* zeigt keine signifikante Abhängigkeit von der Wiedergabequalität, die Daten erscheinen hier als zu rauschhaft. Dieses schwächt die Aussagekraft des Experimentes dahingehend, dass *REC* als ein eher objektives Mass anzusehen ist, weil es im Gegensatz zu den anderen Maßen keinen Beurteilungsschritt durch die Versuchspersonen erfordert.

Die anderen Maße *CONF*, *SPKID* und *CONC* sind

dagegen gemäß ANOVA signifikant durch die Wiedergabequalität beeinflusst, dieser Trend ist besonders deutlich im Bild 1 bei *SPKID* zu sehen. Positive Veränderungen sind vor allem zwischen Mono- und räumlicher Wiedergabe (FB-M, FB-3D) zu vermerken, insbesondere bei *SPKID* und *CONC*. Im Gegenzug werden kaum Veränderungen durch die Erhöhung der Bandbreite erzielt. Die Werte zwischen NB-M und FB-M sind annähernd konstant; ein Effekt der auch in [2] gefunden wurde.

Bezüglich einer möglichen Interaktion zwischen Anzahl Gesprächspartner und Wiedergabequalität ergaben die Interaktionsterme in den ANOVAs keine signifikanten Werte, jedoch sind einige Trends in Bild 1 zu sehen. Bei *SPKID* ist z.B. der Gewinn durch die räumliche Wiedergabe bei mehr als zwei Gesprächspartnern wesentlich größer als der Gewinn bei genau zwei Gesprächspartnern. Dann aber ist dieser Zugewinn annähernd gleich, unabhängig davon ob es drei, vier oder sechs Gesprächspartner sind. *REC* und *CONF* zeigen einen Trend, dass es bei voller Bandbreite (FB-M, FB-3D) keinen Unterschied zwischen vier und sechs Gesprächspartnern gibt, während bei Schmalbandwiedergabe (NB-M) die Sechspersonengespräche noch schwieriger sind als die Vierpersonengespräche.

Schlussfolgerungen

Die Ergebnisse zeigen sehr deutlich eine Abhängigkeit der Messgrößen von der Anzahl der Gesprächspartner. Mit Ausnahme einiger qualitativer Beobachtungen gelang es jedoch nicht, quantitative Aussagen über das Zusammenspiel von Wiedergabequalität und Anzahl der Gesprächspartner treffen zu können. Anscheinend ist die hier umgesetzte experimentale Methode nicht ausreichend sensitiv. So ist eines der Maße nicht durch die Wiedergabequalität beeinflusst, und die anderen zeigen keine signifikanten Interaktionen zwischen Wiedergabequalität und Anzahl der Gesprächspartner. Daher ist als nächster Schritt eine Überarbeitung der experimentalen Methode angedacht, um die Sensitivität bezüglich des Erinnerungsteils (insbesondere *REC*-Maß) sowie bezüglich der Wiedergabequalitäten zu verbessern.

Literatur

- [1] J. J. Baldi, "Effects of Spatial Audio on Memory, Comprehension, and Preference during Desktop Conferences", Proceedings of the ACM CHI 2001 Human Factors in Computing Systems Conference, pp.166-173, 2001.
- [2] A. Raake, C. Schlegel, K. Hoeldtke, M. Geier, J. Ahrens, "Listening and conversational quality of spatial audio conferencing", AES 40th International Conference, Tokyo, 2010.
- [3] M. Geier, J. Ahrens, S. Spors, "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods", Proceedings of 124th AES Convention, Amsterdam, 2008.