

Wie interaktiv sind Telefongespräche?

Florian Hammer

Forschungszentrum Telekommunikation Wien ftw., 1220 Wien, Österreich, Email: hammer@ftw.at

Einleitung

Packet-basierte Übertragung von Sprache (Voice-over-IP, VoIP) bringt zwei wesentliche Faktoren mit sich, die die Qualität der Übertragung beeinflussen: Paketverluste und Übertragungsverzögerung. Dabei repräsentiert die Übertragungsverzögerung eine besondere Beeinträchtigung, falls kein elektrisches oder akustisches Echo auftritt: Die Verzögerung ist nicht „hörbar“ im Sinne einer wahrnehmbaren Veränderung des übertragenen Sprachsignals. Stattdessen wirkt sich die Laufzeit auf die Struktur eines Gesprächs aus. Um eine Verminderung der Sprachqualität eines VoIP-Systems durch Latenzzeiten zu vermeiden, wird von der Standardisierung (ITU-T) ein Maximalwert von 150 ms Einweg-Verzögerung empfohlen. Jüngste Studien [1] haben allerdings gezeigt, dass sich höhere Verzögerungszeiten nur geringfügig auf die von Benutzern wahrgenommene Gesamtqualität auswirken.

Warum interessieren wir uns nun für die Interaktivität von Telefongesprächen? Wir betrachten die Interaktivität als eine instrumentell messbare Größe, die die Struktur der Gespräche darstellt. Sie soll als Hilfsparameter dienen, um den Einfluss der Verzögerungszeit auf die wahrgenommene Qualität genauer studieren zu können.

Konversations-Interaktivität

Da der Begriff „Interaktivität“ in einem sehr breiten Kontext verwendet wird, erweitern wir in unseren Studien die Bezeichnung auf „Konversations-Interaktivität“. Somit ist eine Unterscheidung von andern Bereichen möglich.

Als Grundlage für unsere Beschäftigung mit der Konversations-Interaktivität verwenden wir ein einfaches Konversationsmodell, das aus vier Zuständen besteht:

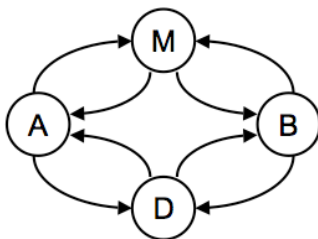


Abbildung 1: Konversationsmodell mit vier Zuständen.

Die Zustände „A“ und „B“ repräsentieren jene Situationen, in denen jeweils nur Teilnehmer A bzw. Teilnehmer B sprechen. Zustand „M“ (mutual silence) steht für den Fall, dass keiner spricht und Zustand „D“ (double talk) für den Fall, dass beide gleichzeitig sprechen. Die direkten Übergänge zwischen „A“ und „B“ bzw. „M“ und „D“ sind sehr seltene Ereignisse und werden in diesem Modell daher

nicht berücksichtigt. Als Gesprächsmodellparameter betrachten wir im Folgenden die Wahrscheinlichkeiten, dass ein Gespräch in einem bestimmten Zustand verweilt (Zustandswahrscheinlichkeiten), und die entsprechenden Verweilzeiten. Die grundlegenden Elemente der Konversationsstruktur sind in [2] beschrieben. Bezüglich der Beschreibung der Konversations-Interaktivität betrachten wir in dieser Untersuchung die folgenden beiden Ansätze:

Die *Rate des Sprecherwechsels* [min^{-1}] gibt an, wie oft ein Gespräch zwischen den beiden Teilnehmern hin- und herwechselt.

Die *Konversationstemperatur* ist ein Interaktivitätsmaß, das auf den Zustandsverweilzeiten basiert. Sind die Verweilzeiten klein, so steigt die Gesprächstemperatur, wird ein Gespräch beispielsweise von einem Teilnehmer dominiert (hohe Verweilzeit), so verringert sich die Gesprächstemperatur, und somit die Konversations-Interaktivität. Die Temperaturmetrik ist in [1] genau beschrieben.

Im Allgemeinen hängt die Konversations-Interaktivität von mehreren Faktoren ab: von der Art/dem Kontext des Gespräches (z.B. privat, beruflich), von den Eigenschaften der Teilnehmer (z.B. Alter, Erfahrung, Geduld, etc.) und von den Eigenschaften der benutzten Telefonverbindung. In diesem Artikel wollen wir uns auf den Kontext und die Übertragungsverzögerung beschränken und anhand von drei Szenarien untersuchen, wie sich die Verzögerung auf die Konversations-Interaktivität auswirkt.

Szenarien

Das erste Szenario wurde einer Untersuchung von Kitawaki [3] entnommen und adaptiert. Dabei werden vorgegebene Listen von Zufallszahlen möglichst schnell reihen- oder spaltenweise überprüft. Dieses Szenario (RNV, Random Number Verification) legt eine sehr strikte Gesprächsstruktur fest und ist mit einem schnellen Frage-Antwort-Spiel vergleichbar.

Der Short-Conversation-Test (SCT) ist ein Standardszenario für die Messung von Konversationssprachqualität. Dieser Test besteht aus Aufgabenstellungen wie der Bestellung einer Pizza oder der Buchung eines Hotelzimmers und ermöglicht ausgewogene, freie Gespräche von kurzer Dauer (ca. 2-3 Minuten) auf Basis realistischer Vorgaben.

Als drittes Szenario wurden freie Gespräche über vorgegebene Themen, z.B. den letzten Urlaub oder der Organisation einer Party, gewählt. Die Zieldauer wurde mit sieben Minuten vorgegeben. Dieses Szenario (FC, Free Conversation) bietet die Möglichkeit, die Struktur von Gesprächen zu untersuchen, deren Verlauf, abgesehen vom Thema, nicht durch die Aufgabenstellung beeinflusst ist.

Benutzertests

Um den Einfluss der Übertragungsverzögerung auf die Konversations-Interaktivität messen zu können, haben wir am ftw. Benutzertests durchgeführt. Der Versuchsaufbau bestand aus zwei PCs mit VoIP-soft-clients (Gnome-Meeting, G.711 codec), die über einen Delay-Emulator miteinander verbunden waren. Zwölf Paare von Versuchspersonen führten eine Reihe von Gesprächen, die auf den drei oben beschriebenen Szenarien basierten. Dabei wurden pro Szenario je drei Gespräche unter verschiedenen Einweg-Laufzeitbedingungen (200, 350 und 500 ms) geführt. Die Gespräche wurden aufgenommen und die aktiven Sprachanteile wurden für die weitere Auswertung markiert.

Resultate

Die Zustandswahrscheinlichkeiten und Verweilzeiten der drei Szenarien in Abhängigkeit der Übertragungsverzögerung sind in Abbildung 2 dargestellt. Bezüglich der Wahrscheinlichkeiten weist das RNV-Szenario den höchsten Anteil an gemeinsamer Stille („M“) auf. Dieser Anteil steigt von 37 % bei 200 ms auf 43 % bei 350 ms an. Gegensprechen tritt hier nur in geringem Ausmaß auf (um 5 %). In den SCT- und FC-Szenarien bewegen sich die Wahrscheinlichkeiten, dass entweder Sprecher „A“ oder „B“ spricht, bei 33 %. Der Anteil der gemeinsamen Stille ist bei den SCTs wesentlich höher als bei den FCs, was auf die stärker vorgegebene Gesprächsstruktur in der Aufgabenstellung zurückgeführt werden kann.

Die Verweilzeiten für die Zustände „A“ und „B“ der SCT- und FC-Szenarien sind ungefähr gleich (um 1,4 s), während diese beim RNV-Szenario im Vergleich nur mehr die Hälfte (um 0,7 s) betragen. Eine verlängerte Zeitverzögerung bewirkt einen stetigen Anstieg der Verweilzeit für gemeinsame Stille („M“) im RNV-Szenario, während sich bei den SCT- und FC-Konversationen dieser Wert nicht wesentlich ändert. Bei den freien Konversationen ergeben sich die höchsten Verweilzeiten für Gegensprechen, die sich jenen der gemeinsamen Stille annähern und über die Verzögerungszeit konstant bleiben.

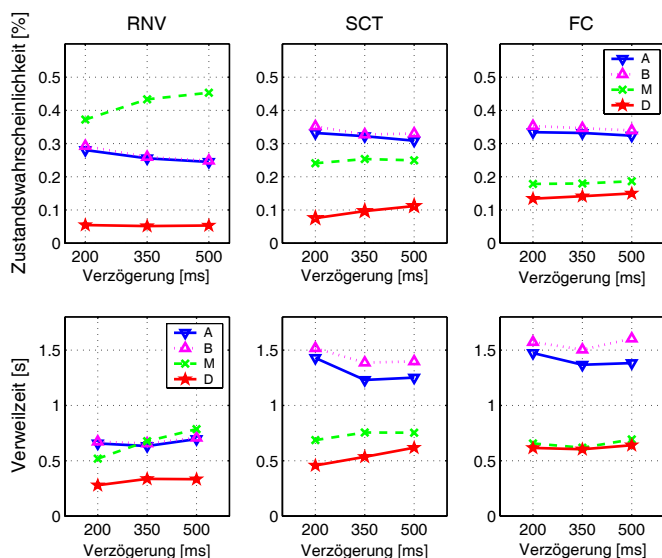


Abbildung 2: Mittlere Zustandswahrscheinlichkeiten und Verweilzeiten.

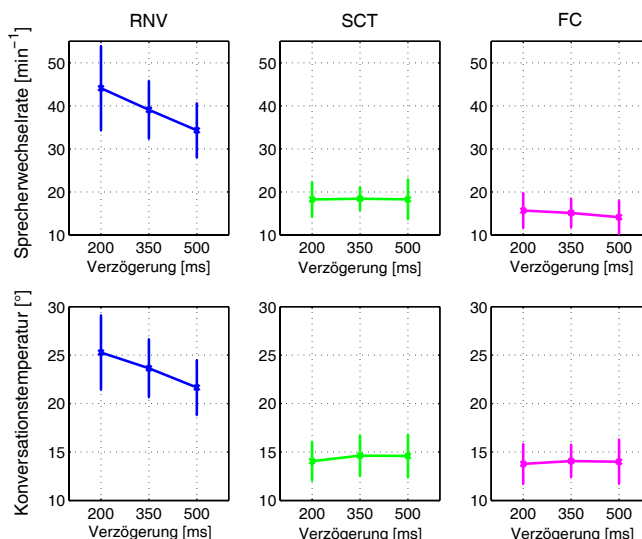


Abbildung 3: Mittlere Sprecherwechselraten und Konversationstemperaturen und deren Standardabweichungen

Die mittleren Sprecherwechselraten und Konversationstemperaturen sind in Abbildung 3 dargestellt. Ein Vergleich der Sprecherwechselraten zeigt, dass diese im RNV-Szenario mit steigender Verzögerung signifikant sinken. In den anderen beiden Szenarien ändert sich die Sprecherwechselrate kaum. Dies lässt sich wiederum durch die starre Struktur der RNV-Aufgaben erklären, durch die das Gespräch ständig zwischen den beiden Teilnehmern wechselt. Die Werte der Konversationstemperaturen zeigen das gleiche Bild wie die der Sprecherwechselraten. Das RNV-Szenario ist bei weitem am interaktivsten, und das SCT-Szenario scheint ein wenig interaktiver als die freien Konversationen und nähert diese gut an.

Zusammenfassung

Eine detaillierte Untersuchung des Einflusses der Übertragungsverzögerung auf die Sprachqualität erfordert die Differenzierung von Szenarien und eine genaue Studie der Gesprächsstrukturen. In diesem Artikel haben wir drei Szenarien verglichen und deren Konversations-Interaktivität untersucht.

Im nächsten Schritt wird die Konversations-Interaktivität der wahrgenommenen Sprachqualität gegenübergestellt.

Literatur

- [1] F. Hammer, P. Reichl and A. Raake, „The Well-Tempered Conversation: Interactivity, Delay and Perceptual VoIP Quality“, IEEE Int. Conf. on Communications (ICC), Seoul, Korea, May 2005.
- [2] F. Hammer, P. Reichl and A. Raake, „Elements of Interactivity in Telephone Conversations“, Int. Conf. Spoken Language Processing (ICSLP/INTERSPEECH 2004), Vol. 3, Jeju Island, Korea, Oct 2004.
- [3] Kitawaki und Itoh, “Pure Delay Effects on Speech Quality in Telecommunications”, IEEE Journal on Selected Areas of Communications, Vol. 9, No. 4, pp. 586-593, May 1991.