SmartWeb Datenworkshop Protokoll

Moritz Kaiser
IPSK
Ludwigs-Maximilians-Universität München
ariser@phonetik.uni-muenchen.de

10. August 2004

Der Datenworkshop befasste sich mit den Sprachdaten, die vom IPSK aufgezeichnet werden sollen. Hierzu wurden die Wünsche gesammelt, die von den verschiedenen Projektpartnern geäußert wurden und versucht, einen Konsens über den Umfang und die Art der Daten, die aufgezeichnet werden sollen, zu finden.

Unter anderem wurden folgende Probleme besprochen:

- Welche Geräte werden zur Aufnahme benutzt?
- Wie viele Sprachkanäle sollen aufgezeichnet werden?
- Wie sollen die Daten verschriftet werden?
- Wie können möglichst dialogähnliche Aufzeichnungen gemacht werden?
- Welche Themen sollen in den Aufzeichnungen fokussiert werden?

Inhaltsverzeichnis

Protokoll

Moderation Silke Steininger, Florian Schiel

Protokollant Moritz Kaiser

Anwesend Anton Batliner (FAU)
Irene Cramer (UdS)
Stefan Schacht (UdS)
Andreas Dirschl - (BMW)
Ralf Decke (BMW)
Florian Gallwitz (Sympalog)
Robert Porzel (EML)
Anselm Blocher (DFKI)
Hannes Mögele (LMU)

Sitzungsort IPSK, LMU, Schellingstr. 3, München

Datum 22. Juli 2004 10:10 – 17:00

Tagesordnung

---

1 Erste Ergebnisse zu UMTS und Handies

Moritz Kaiser stellt kurz die Ergebnisse der bisher stattgefundenen Recherchen des IPSK zum Thema UMTS-Netz und mobile Endgeräte vor.

1.1 Informationen über UMTS

- UMTS ist laut T-Mobile zunächst nur in Ballungsräumen verfügbar, weil nur mit dieser Strategie die Vorgaben des Bundes bezüglich Versorgung der Bevölkerung zu erreichen sind. München ist hier relativ gut versorgt.
- Mit T-Mobile wurden mögliche Aufnahmeorte erörtert. Geeignet sind das Olympiastadion, der Hauptbahnhof und für bewegte Aufnahmen auch ruhige Wohngebiete am Stadtrand.
- Endgeräte werden von Siemens, Motorola, Nokia und Sony-Ericsson hergestellt. Siemens und Nokia sind im Handel erhältlich, Siemens stellt kostenlos Testgeräte zur Verfügung. Das IPSK hat bereits zwei Siemens U15 mit der neuesten Softwareversion.
- Die Datenübertragungsraten bei paketvermittelnden Diensten, wie Telefonie und Internet liegen bei 64 kbit/s upstream und 384 kbit/s downstream. Für Videokonferenzen (sog. switched-circuit-Verbindung) liegt die Datenrate bei 64 kbit/s in jede Richtung. Theoretisch kann die Datenrate verändert werden. Es gibt eine Norm mit 128 kbit/s upstream.

1.2 Testergebnisse mit Siemens U15 und Logitech Headset

Die Mobiltelefone wurden mit einer normalen D1 Karte in Betrieb genommen. Es wurde von Logitech ein Bluetooth-Headset beschafft, das sich mit dem Siemens U15 betreiben lässt. Es liegen erste Erfahrungen vor.

- Der Wechsel von UMTS zu GSM bei schlechter werdendem WCDMA-Band wird von den Geräten im Betrieb automatisch vorgenommen. Dies lässt sich momentan noch mit kurzen akustischen Unterbrechungen im Subsekundenbereich erkennen. Der Wechsel zu GSM kann aber durch manuelle Auswahl des Frequenzbandes unterbunden werden.
- Die Sprachqualität des Bluetooth-Headsets ist relativ gut, Probleme entstehen aber durch Windgeräusche, die zwar durch den mitgelieferten Windschutz reduziert aber nicht unterbunden werden.
- Für die Nutzung des UMTS-Netzes ist eine D1- oder D2-SIM-Karte ausreichend, die Gebühren sind gleich, es muss nichts freigeschaltet oder konfiguriert werden.
- Am IPSK wurde ein ISDN-Server so konfiguriert, dass testweise Sprache über den Telefoniekanal aufgezeichnet werden kann. Es werden einfache Prompts ausgegeben und über definierte Zeitspannen aufgezeichnet.

Es werden noch einige Testaufnahmen vom ISDN-Server abgespielt. Hier treten vor allem die Windgeräusche deutlich zu Tage und etwaige Probleme mit der Bluetoothverbindung. Im Gegensatz zu GSM, wo bei Netzverlust kurze Pausen entstehen, treten bei UMTS Kompressionsartefakte auf, wie man sie von zu stark und fehlerhaft komprimierten Mpeg 2 Layer 3 Spuren kennt. Akustisch äußert sich das durch „Zwitschern“ und „Gurgeln“. Ursache ist vermutlich die Tatsache, dass der Codec bei schlechter werdendem Kanal die Kompression erhöht und die Datenrate senkt.

2 Bekannte Wünsche zu den technischen Rahmenbedingungen

Es wird unterschieden zwischen der Ausstattung des Fußgängers und des Motorradfahrers.

2.1 Wünsche zur Technik beim Fußgängerszenario

- UMTS Smartphone von Sony-Ericsson (FAU)
- Bluetooth Headset (FAU)
- Über UMTS-Kanal Headset und eingebautes Mikro aufzeichnen und beide Kanäle noch einmal unkomprimiert. (FAU)

- Headset und eingeb. Mic nur über UMTS aufzeichen (T-Systems)
- Videokanal mit 15 fps und 320x240 Pixel (FAU, UdS)

2.2 Wünsche zur Technik beim Motorradszenario

- Das UMTS-Handy soll per Bluetooth mit dem Motorrad (Gateway) verbunden sein und von dort den Audiokanal erhalten.
- Der Sprachinput kommt über ein Helmmikrofon zum Motorrad (Gateway).
- Zusätzlich sollen zwei unkomprimierte Kanäle aufgezeichnet werden. Einer aus dem Helmmikrofon und einer aus einem zusätzlichen Mikrofon.

3 Diskussion über die technischen Rahmenbedingungen

Die Diskussion über die Technik ist mit den anderen Tagesordnungspunkten stark verwoben, weswegen der Diskussionsablauf nur ungefähr wiedergegeben werden kann.

3.1 Diskussion zum Aufnahmegerät, Bedientechnik, Headset usw.

Die FAU wünscht sich das Z1010 von Sony-Ericsson als Aufnahmegerät, weil zu diesem Smartphone ein gutes Developmentkit existiert, und die Videoaufzeichnung damit möglich sein soll. Es kommt zu einer Diskussion über das Thema Videoaufzeichnung.

⊕ Notwendig für Erkennung von On-/Off-View, On-/Off-Talk (FAU, Sympalog)

⊕ Notwendig für die Erkennung der Lippenbewegungen (UdS)

⊖ Aufzeichnung ist nach jetzigem Kenntnisstand sehr schwierig und aufwändig

⊖ Eine Annotation ist aus Zeit- und Kostengründen unmöglich.

- Florian Schiel plädiert dafür, lieber rechtzeitig mit den Sprachdaten anzufangen und freie Kapazitäten für Video zu nutzen.
- Stefan Schacht schlägt eine Offline-Aufzeichnung von Videodaten vor.

⇒ Es wird mit der Sprachaufzeichnung begonnen, parallel wird die Machbarkeit von Videoaufzeichnungen studiert.

Beschluss: Keine exakte Festlegung bei Videoaufzeichnungen

> Es können keine exakten Festlegungen zu den Videoaufzeichnungen getroffen werden, außer, dass zunächst die Machbarkeit geprüft wird, und dann wahrscheinlich eine Offlineaufzeichnung mit 320 x 250 Pixeln in 12 bis 15 fps durchgeführt wird.

Die Teilnehmer sind sich nach kurzer Besprechung einig, dass der Typ des Mobiltelefons bei Aufnahme über das Headset keine große Rolle für die Tonqualität spielen wird. Es empfiehlt sich deshalb, das U15 verwenden, weil keine zusätzlichen Besorgungs- und Organisationsaufgaben anfallen.

Beschluss: Aufnahmegerät wird das Siemens U15

> Es wird mit dem U15 aufgenommen, da das Vorhandensein des Gerätes das Prozedere beschleunigt.

Beschluss: Wenn BMW andere Handies verwenden will, stellen sie diese selbst.

Beschluss: Aufnahmen zu Fuß werden mit Headset gemacht

> Die Benutzung des Telefons erfolgt ausschließlich über das Headset, andere Modi zur Sprachaufzeichnung werden nicht berücksichtigt, bzw. erst zu einem späteren Zeitpunkt.

Beschluss: Aufgenommen wird mit verschiedenen Headsets

> Mehrere Headsets sollen besorgt und wechselweise verwendet werden, um eine möglichst große Variation bei den darin verbauten Mikrofonen zu erreichen.

Es wird kurz über die Anforderungen an die Qualität gesprochen. Die Entwickler der Spracherkenner stellen klar, dass schlechte Übertragungsqualität in den Aufnahmen die Fähigkeiten der Erkenner nicht wirklich verbessert. Netzaussetzer mit nennenswertem Informationsverlust machen die Erkenner schlechter.

Beschluss: Bei den Aufnahmen muss die Netzabdeckung optimal sein

> Damit möglichst viele verwertbare (aus Sicht der Spracherkenner) Aufnahmen entstehen, sollen keine Aufnahmeorte mit schlechtem Netz verwendet werden. Es soll „garbage in - garbage out“ Problematik beim Erkenner vermieden werden.

Beschluss: Es wird nur UMTS benutzt

> Da der Netzwechsel auf GSM auch von Aussetzern begleitet ist, und das Projekt insgesamt auf die Nutzung von Breitbanddiensten ausgerichtet ist, wird auf Bandswitching verzichtet und nur über UMTS aufgezeichnet.

3.2 Diskussion über die aufzuzeichnenden Audiokanäle

Nach momentanem Wissensstand kann nur ein Audiokanal über UMTS übertragen werden. Es müssen also alle zusätzlichen Kanäle lokal aufgezeichnet werden. Florian Schiel schlägt hierzu einen Harddiskrecorder vor, der als MP3-Box von Creative Labs erhältlich ist. Das Gerät ist in der Lage, zwei Kanäle unkomprimiert mit 16 bit über mehrere Stunden aufzuzeichnen. Es zeichnet allerdings mit vorgegebener Abtastrate auf, die nicht reduzierte werden kann. Es werden mehrere Vorschläge zu möglichen Kanälen gemacht, unter anderem:

- Headset-Mikro über UMTS

- Eingebautes Mikro über UMTS
- Headset-Mikro direkt auf Datenträger
- Eingebautes Mikro direkt auf Datenträger
- Zusatzmikro am Gerät auf Datenträger

Die Verwendbarkeit des eingebauten Mikrofons ist nicht geklärt, weshalb sich dieser Kanal nicht mit Sicherheit realisieren lässt.

Beschluss: Headset-Mikro wird über UMTS aufgezeichnet

> Es wird definitiv ein Kanal vom Headsetmikrofon über UMTS auf dem Sprachserver aufgezeichnet

Beschluss: Headset-Mikro wird direkt aufgezeichnet (HD)

> Das IPSK setzt alles daran, das Headsetmikro auf einem Harddiskrecorder aufzuzeichnen

Beschluss: Zusatzmikro wird direkt aufgezeichnet (HD)

> Ein weiteres Mikrofon wird auf einem Harddiskrecorder aufgezeichnet, möglicherweise kann das eingebaute Mikrofon dazu verwendet werden

Beschluss: Motorradkanäle werden geklärt (LMU $\longleftrightarrow$ BMW)

> Die vom Motorrad aufzuzeichnenden Kanäle, die über das BT-Headset hinausgehen, werden von BMW und LMU direkt abgesprochen, z.B. mögliche Aufzeichnung der Umgebungsgeräusche zur Erfassung des Lombardeffekts

Beschluss: Für Motorrad wird auch GSM verwendet

Beschluss: Harddiskrecorder wird von LMU gestellt

> Für die Aufzeichnung aller Nicht-UMTS-Kanäle wird der Harddiskrecorder der LMU verwendet.

Nach der Beschlussfassung über die Kanäle wird noch einmal das Problem der Videoaufzeichnungen besprochen. Die LMU wird versuchen, Streamingaufzeichnungen auf einem Server zu realisieren, bzw. wenn das nicht funktionieren sollte, die Möglichkeiten, lokale Aufzeichnungen in den Speicher des Mobiltelefons durchzuführen, testen. Außerdem wird über die Inhalte der Videodatensammlung diskutiert. Es ist fraglich, ob Videoaufnahmen, wenn sie durchgeführt werden, für das Lippenablesen geeignet sein werden. On-View / Off-View wird wahrscheinlich erkennbar sein, ebenso wie On-head / Off-head. Schwierigkeiten dürfte auch On-talk / Off-talk bereiten.

Beschluss: Videoaufzeichnungen werden in die Sprachdatensammlung integriert

> Die zukünftigen Videoaufnahmen werden in die Sprachdatensammlung integriert, das heißt, sie werden unter gleichzeitiger Aufzeichnung der Sprache und gleichen Bedingungen aufgenommen

Beschluss: Videoaufnahmen mit späterem Endgerät.

Es muss in jedem Fall auf das Ziel hingearbeitet werden, die Videoaufnahmen mit dem späteren Prototypen durchzuführen

Beschluss: On-/Off-view/-head sollen auf Video sein.

Beschluss: Die Aufnahmen werden zu je 50% mit Motorrad und Handheld gemacht

Das Datenvolumen der Aufnahmen, die auf dem Motorrad erzeugt werden, sollte ungefähr gleich dem der Aufnahmen vom Handheld sein.

4 Diskussionen über technische Probleme und deren Lösungsansätze

4.1 Differenzen bezüglich des Gerätes

Dieser Punkt wurde nicht analog zur Tagesordnung bearbeitet, sondern innerhalb des vorhergehenden Überpunktes mitabgehandelt. Zur Beschlusslage siehe die Beschlusszusammenfassung im Anhang.

4.2 Probleme mit dem UMTS-Netz

Siehe vorhergehenden Punkt.

5 Wünsche zu den Inhalten/Modalitäten der Datensammlung

5.1 Vorüberlegungen der LMU zur Datensammlung vor dem Datenworkshop

Hier werden nur kurz die Überlegungen der LMU bezüglich einiger Rahmenbedingungen dargestellt. Diese waren Gegenstand der Diskussion auf dem Datenworkshop und sind in den Beschlüssen aufgegangen. Die Liste aus der Präsentationsvorlage wird nur der Vollständigkeit halber angeführt.

Es folgt eine Auflistung von kleineren Stichpunkten:

- KEINE WOZ-Dialoge! (Kosten!)
- Vpn rufen mit Handy einen Telefonserver an.
- Bekommen „Situative Prompts“ aus zwei Hauptgruppen.
- Ein Thema (z.B. Weg von U-Bahn zu Stadion) wird mehrmals abgefragt und dabei um unterschiedliche Formulierungen gebeten.

- Aufnahmen finden in Gruppen im Feld statt.
- Nachbearbeitung: Synchronisation, Transliteration.

Was unter situativen Prompts zu verstehen ist, sollen zwei Beispiele erläutern

- Prompt: *Stellen Sie sich vor, Sie kommen gerade aus der U-Bahn und wissen den Weg zum Stadion nicht. Verwenden Sie Ihre eigenen Worte, um Smartweb um Hilfe zu bitten.*
- Antwort: „Äh... Ok. Also... Smartweb, bitte zeig mir doch mal - gib mir bitte eine Wegbeschreibung von der U-Bahn zum Stadion. U-Bahn Olympiapark bis Olympiastadion."

Das zweite Beispiel zwingt den User zu mehr „Turns", man könnte die Methode *Kontrollierte Dialogfragmente* nennen:
Die Versuchsperson wird mit zwei Stimmen konfrontiert, einem Operator, der Instruktionen zum Versuch gibt und einer *Promptstimme*, die sozusagen SmartWeb simuliert.

- Stellen Sie sich vor, Sie sind Zuschauer in einem Vorrundenspiel. Da Sie sich über das Reglement nicht ganz sicher sind, wollen Sie einfach wissen, welche Duelle im Achtelfinale stattfinden. Fragen Sie Smartweb nach Informationen.
- Smartweb! Wer spielt im Achtelfinale?
- *Es stehen noch nicht alle Mannschaften fest, die im 8tel-Finale spielen werden.*
- Versuchen Sie jetzt, wenigstens eine Teilinformation zu erhalten.
- Welche Mannschaften stehen schon fest?
- Bitte formulieren Sie die gleiche Anfrage noch einmal in anderen Worten.
- Welche Mannschaften spielen sicher im Achtelfinale?
- SmartWeb sagt Ihnen eine Liste der Mannschaften. Verlangen Sie eine Wiederholung der Aufzählung.
- Bitte nochmal.

5.2 Inhalte der Datensammlung, die von den Partnern gewünscht waren

Folgende Wünsche betreffen die Art der Aufnahmen, und weitere nichttechnische Rahmenbedingungen:

- On/Off-Talk, On/Off-View, On/Off-Head (FAU)
- Nicht nur Anfragen, sondern auch (Art der) Antworten (EML)

- Kompletter Dialog mit Schneiden (UdS)
- Vermeidung von Paraphrasierung/untypischer Sprache (EML)
- Trigger word: „Smartweb bitte“ (FAU, EML) vs. On-Talk Detektion (?)
- Benutzern frei lassen, welche Infos sie erfragen wollen (EML)

Die detaillierten Wünsche zu den Inhalten *vor* dem Workshop werden der Vollständigkeit halber wiedergegeben:

- Offene Domäne (Bsp. „Fuji?“) (DFKI, EML)
- Allgemeine Fußballthemen (FAU, UdS)
- Aktuelle Fußballthemen (FAU, UdS, EML, Sympalog)
- Alle WM-Austragungsorte, Repräsentative Menge der WM- Mannschaften (FAU, Sympalog)
- Fußgänger-Navigation (T-Systems, DFKI, UdS)
- Freundeskreis, Verabredungen, Treffen (DFKI)
- Meta-Komm., Dialognavigation, Fehlerkorrektur (T-Systems, DFKI)
- Ungewöhnliche Namen (DFKI)
- Provozierte OOVs (FAU, Sympalog)
- Namen einzeln und in Listen (FAU, Sympalog)

6 Diskussion über Modalitäten und Inhalte der Datenerfassung

Da zwischen den Modalitäten und den Inhalten große Interdependenzen bestehen, fand eine Diskussion über beide Aspekte der Daten in Einem statt. Wichtige Argumente werden hier erwähnt. Viele Teilnehmer sehen das Fehlen von echten WOZ-Dialogen (Wizard of Oz) als sehr problematisch an, weil sich Vpn bei Prompts wesentlich anders verhalten, als in einem realen Dialog. Andererseits werden die Nutzer von Smartweb später auch nicht mit einer realen Person am Telefon kalkulieren und sich ebenso unberechenbar verhalten. Die LMU führt an, dass WOZ-Aufnahmen schlicht unbezahlbar sind, zum Einen, weil eine zusätzliche Person notwendig ist, um den Dialog zu führen und zum Anderen, weil die Prompts auch transliteriert werden müssen. Nach der Vorstellung der situativen Prompts und der kontrollierten Dialogfragmente werden Zweifel geäußert, ob diese auch „funktionieren“, also von der Versuchsperson so befragt werden, wie sich die LMU dies vorstellt. Es ist aus Sicht der Beteiligten notwendig, *vor* Beginn der Aufnahmen diese Prompts aus echten Dialogen zu entwickeln und zu testen.
Die simulierten Dialoge können aber nicht alle notwendigen Worte zum Training der

Spracherkenner liefern. So müssen Ordinalzahlen u.ä. erzwungen werden. Hier entsteht eine umfangreiche Liste von Inhalten, die nicht mit zufälligen Dialogfragmenten erzeugt werden können, sondern „abgefragt“ werden müssen.

6.1 Ablauf der Datensammlung und weitere Rahmenbedingungen

Aus der Diskussion ergibt sich folgende Vorgehensweise:

Beschluss: Es wird eine Vorstudie mit echten Dialogen durchgeführt

> Diese Vorstudie wird mit einer realen Person als „Sprachserver“ durchgeführt, der eine Suche im Internet vornimmt. Um den Ablauf des Informationsabrufs nachvollziehen zu können, wird ein Protokoll über die abgerufenen Webseiten angelegt.

Beschluss: Zwischenergebnisse werden mit den Partnern diskutiert

> Die Ergebnisse der Vorstudie werden zeitnah mit den Partnern besprochen und daraus die situativen Prompts entwickelt. Die Partner liefern weitere Informationen

Beschluss: Die Prompts werden von den Partnern bewertet

> Vor Beginn der Datensammlung wird von den Partnern das OK eingeholt

Innerhalb der Diskussion um die Modalitäten wird auch der Ort für die Aufnahmen thematisiert. Zentraler Punkt der Ortsfrage ist das Maß an Hintergrundgeräuschen.

Beschluss: Outdooraufnahmen sollen bei geringem Wind erfolgen

> Bei viel Wind sind die Störgeräusche so stark, dass Spracherkenner nicht mehr mit den Daten trainiert werden können

Beschluss: Indooraufnahmen sollen in ruhiger Umgebung erfolgen

> Dies führt auch zu einem besseren Training der Erkenner. Tolerabel sind Hintergrundgespräche, also gedämpfte Kaffeehausumbgebung.

Beschluss: Outdooraufnahmen sollen bei moderatem Lärm erfolgen.

> Dies soll bei 2/3 der Aufnahmen gelten. Der Rest kann mit starkem Straßenlärm oder einzelnen Störunterbrechung belastet sein.

Beschluss: Unverständliche, beschädigte Aufnahmen aussortieren

> Dies soll auch dem garbage-in garbage-out Phänomen vorbeugen. Diese Aufnahmen sollen aber nur separiert und nicht weggeworfen werden.

Beschluss: Die Aufzeichnungsorte für Motorrad werden zwischen LMU und BMW geklärt.

Beschluss: Aufzeichnungsformat für Video ist MPEG4

6.2 Inhalte, die in der Datensammlung landen sollen

Beschluss: Wortschatz aus Fifawebseiten verwenden

Beschluss: Große Variabilität in den Formulierungen

Beschluss: Versuchspersonen sollen evtl. vorher Webseiten sehen.

> Ein Priming mit Webseiten soll in manchen Fällen vorher verhindern dass typische „Äh, ich weiß nicht, was ich jetzt fragen soll“ -Fragen gestellt werden und bestimmte Inhalte sicher abgedeckt werden.

Beschluss: Provoziertes Offtalk (10-20%) soll entstehen

> Der Versuchsleiter soll die Vpn gezielt unterbrechen, damit Offtalk stattfinden kann.

Beschluss: Mitschneiden des Gesamtdialogs wenn möglich auch Server-seitig.

> Dies soll die Auswertung erleichtern, steht und fällt aber mit den technischen Möglichkeiten.

Beschluss: Allgemeine Fußballfragen, Frage nach aktuellen Fußballinfos werden provoziert.

Beschluss: Fußgängernavigation, touristische Anfragen, ÖPNV wird provoziert.

Beschluss: Sog. Streitthema, Off-Talk mit Versuchsleiter, wird provoziert.

> Das Streitthema ist meist eine einfache inhaltliche Anfrage, die zur Klärung einer Streitfrage gestellt wird. Es wird aber auch zur Erzeugung von Off-Talk verwendet werden.

Beschluss: Ja-Nein-Fragen, Wiederholunganforderungen müssen provoziert werden.

Beschluss: Anfragen zur Community

> Die Community soll über Fragen erforscht werden.

Beschluss: Austragungsorte, Ländernamen (Varianten), ungew. Namen, werden gepromptet.

Beschluss: Mindestens 250 Listen (3-6 Einträge) von „OOV“ aufnehmen

Beschluss: Es wird versucht, offenere Fragen zu prompten.

Beschluss: Alle Domänen sollen inhaltlich gemischt werden.

6.3 Inhalte der Datensammlung, die spezifisch für das Motorrad sind

Beschluss: Es werden fahrzeugspezifische Fragen gestellt.

Beschluss: Es werden Fragen zu allgemeinen points of interest gestellt.

Beschluss: Spezielle Kommandoworte für das Gefährt werden provoziert.

Beschluss: Situative Prompts zur Wetter- und Verkehrslage werden gestellt.

Dies ist einer der wichtigsten Punkte für BMW.

7 Festlegung des Sprecherprofils

Die Diskussion um die Sprecherprofile der Handhelduser verläuft sehr kurz, weshalb hier keine weiteren Details und Argumente angeführt werden.

Beschluss: Alter der Sprecher: 18-45 Jahre

Beschluss: Geschlecht der Sprecher: 60% Männer, 40% Frauen

Beschluss: Beruf/Bildung gemischt

Beschluss: Sprecher besitzen eigenes Handy

Beschluss: Fußballinteressierte Sprecher

Beschluss: Sprachdialogerfahrung ist für die Sprecher unwichtig

Beschluss: Der Sprecher spricht fließend deutsch

Für die Sprecher am Motorrad gelten nur geringfügig abweichend Daten, sie ergeben sich aber aus den zur Verfügung stehenden Fahren von BMW.

Beschluss: Alter der MR-Fahrer liegt durchschnittlich 40

Beschluss: Geschlecht der Fahrer: max möglich % Frauen (?)

Beschluss: Erfahrene Motorradfahrer sind nötig.

Beschluss: Sprachdialogerfahrung der Fahrer ist irrelevant.

8 Festlegung der Transliteration

In diesem Zusammenhang schlägt die LMU zunächst eine Verschriftung mit einer Teilmenge der Regeln aus SmartKom vor. Es wird aber auch um die Änderung einiger Regeln gebeten, um ein Parsen mit regulären Ausdrücken zu erleichtern. Einige Wünsche zum Offtalk werden angebracht.

Beschluss: Basis der Verschriftungsregeln ist eine echte Teilmenge der Smartkomregeln

Durch diese Entscheidung können Tools aus SmartKom und Erfahrungen übernommen werden. Die Verschriftung wird durch Weglassen von Prosodieinformationen wesentlich vereinfacht.

Beschluss: Gelabelt werden soll: OOT, ROT, <ähm>..., <P>, <Z>

Beschluss: Weitere Labels für: Neologismen, Buchstabierungen, Fremdspr., Abbrüche, Wortunterbrechungen, Aussprachevarianten

Beschluss: Beispiele für Transliteration werden rechtzeitig an die Partner geschickt.

Beispiel: *s038_pfu_002_AAS: <hm> wann gibt es wann spielt denn Deutschland wieder*

Beschluss: Für Video wird keine andere TRL verwendet.

Beschluss: Die Videodaten selbst werden nicht gelabelt.

9 Aufzeichnung der Metadaten / Inhalte der Protokolle

Die zu notierenden Metadaten bestehen aus einem Recordingprotokoll und einem Sprecherprotokoll

9.1 Inhalt des Recordingprotokolls

Da hinter dem Recordingprotokoll kein nennenswerter Aufwand steht, kann allen Wünschen entsprochen werden. Diskussionsbedarf besteht nur an den Stellen, wo unscharfe Aussagen die Information verwischen und wertlos machen würden.

Beschluss: Beginn/Ende Recording

Beschluss: Örtlichkeit (bei Bewegung Anfang, Endpunkt)

Beschluss: Hintergrundgeräusche Vier Kategorien (von VL und Labeller)

Beschluss: Wetter: Regen, trocken, warm, kalt

Beschluss: Mot.: Verkehr, Strecke

Beschluss: Besondere Umstände (Freitext)

Beschluss: Requisiten der Vpn

Beschluss: Einstellungen am Aufnahmegerät (Gain, ...)

Beschluss: Handytyp, Headsettyp

Beschluss: Biosignale werden zwischen LMU, FAU und EML vereinbart.

9.2 Inhalt des Sprecherprotokolls

Beschluss: Alter des Sprechers

Beschluss: Geschlecht des Sprechers

Beschluss: Bundesland Grundschule (Dialekt) des Sprechers

Beschluss: Muttersprachen (Sprecher, Mutter, Vater)

Beschluss: Piercings, sichtbar + Mund

Beschluss: Raucher

Beschluss: Brille, Bartart (Vollb., Schnauzer, Backenbart, Kinnbart, Koteletten), Glatze

Beschluss: Schulabschluss des Sprechers

Beschluss: Sprachdialogsystemerfahrung des Sprechers

Beschluss: Mot.: Testfahrer, normaler Fahrer

Beschluss: Besonderheiten (Freitext, Sprachfehler, unkoop.)

A Entscheidungen

B Nachträgliche Entscheidungen vom Gesamtworkshop 26.-27.7.2004

- Beim Labeling sollen Windgeräusche gekennzeichnet werden.
- Für die Datensammlung am Motorrad sollen folgende weiteren Inhalte erfasst/provoziert werden:
  - Einzelziffern
  - Zweistellige Zahlen
  - Auswahlordinalien aus Listen bis fünf Einträge (z.B.: erster, drittes, fünfte)