Zeitschrift: bulletin.ch / Electrosuisse

Herausgeber: Electrosuisse

Band: 100 (2009)

Heft: 11

Artikel: Elektronische Dolmetscher

Autor: Wölfel, Matthias

DOI: https://doi.org/10.5169/seals-856422

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 28.11.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Elektronische Dolmetscher

Aktueller Stand der automatischen Sprach-zu-Sprach-Übersetzung

Durch die automatische Sprach-zu-Sprach-Übersetzung, mit der ein beliebiges Gespräch maschinell und simultan übersetzt werden kann, rückt man dem alten Menschheitstraum, eine fremde Sprache sprechen und verstehen zu können, ohne diese selbst zu beherrschen, erheblich näher. Doch wie funktioniert ein solches System? Und wie weit ist die Entwicklung und der dadurch erzielte Nutzen bereits heute? Der vorliegende Beitrag beschäftigt sich mit diesen Fragen, erklärt die Grundlagen, nennt Anwendungsbeispiele und gibt Aufschluss über die Qualität aktueller Systeme.

Durch die zunehmende Globalisierung und Vernetzung der Welt durch Telefon und Internet wird die Distanz zwischen Kommunikationspartnern nicht mehr in Metern

Matthias Wölfel

oder Meilen gemessen, sondern in der Grösse und Anzahl der Kommunikationsbarrieren, die überwunden werden müssen. Neben Telekommunikations- und Datenübertragungskosten sowie der Zeitverschiebung ist wohl die grösste Hürde die Überwindung der Sprachbarriere. In vielen Bereichen, wie z.B. an der Hotelrezeption, an der Fahrkartenauskunft oder beim Arztbesuch im Urlaub, ist eine menschliche Übersetzung meistens zu teuer oder nicht verfügbar. Daher werden automatisierte Verfahren benötigt, die die Sprachwissenschaftler und Entwickler vor eine gewaltige Herausforderung und faszinierende neue Probleme stellen, um dem gestiegenen Bedarf an Übersetzungen gerecht zu werden

Ein Computerprogramm kann Sprache ebenso wenig verstehen wie andere Berechnungen, die es zuverlässig durchführt. Daher muss die Funktionsweise der Sprache für den Computer durch Experten modelliert (regelbasierte Systeme) oder anhand geeigneter Trainingsdaten automatisch gefunden werden (statistische Systeme). Das Ziel ist hierbei die Gewährleistung einer natürlichen und qualitativ hochwertigen Mensch-zu-Mensch-Kommunikation, bei der die Quellinformationen erhalten bleiben und keine technischen Artefakte auftreten.

Da erste Prototypen die Realisierbarkeit solcher Systeme bereits bewiesen haben und eine kostengünstige Überwindung der sprachlichen Kluft erhebliche praktische und wirtschaftliche Vorteile birgt, erhält die automatische Sprach-zu-Sprach-Übersetzung zunehmende Bedeutung. Somit rückt man dem alten Menschheitstraum (2 prominente Beispiele sind das Pfingstwunder und Douglas Adams' «Babelfisch»), dem Verstehen und Sprechen einer fremden Sprache, ein ganzes Stück näher.

Geschichtliche Entwicklung und aktuelle Anwendungsbeispiele

Ende der 1950er-Jahre wurde die Entwicklung der automatischen Spracherkennung durch die akustische Phonetik, die die Laute der Sprache und deren Umsetzung in eine gesprochene Äusserung untersucht, vorangetrieben. So wurde bereits im Jahr 1952 von Davis, Biddulph und Balashek von den Bell Laboratories ein System zur Erkennung von getrennt gesprochenen Zahlen eines spezifischen Sprechers vorgestellt. Mitte der 1970er-Jahre wurden dann grundlegende Technologien zur Mustererkennung für die Spracherkennung adaptiert. Diese schablonenbasierten Ansätze wurden in den 1980er-Jahren durch statistische Verfahren verdrängt. Besonders populär ist hier ein doppelt stochastischer Prozess, das Hidden Markov Model (HMM), der in der Spracherkennung sowohl die Variabilität des Sprachsignals als auch die Struktur der gesprochenen Sprache in einem einheitlichen statistischen Rahmen

repräsentiert. Das HMM, als wichtigste Grundlage der automatischen Spracherkennung, ist vor allem wegen der ständigen Verbesserung und Weiterentwicklung bis heute ungebrochen.

Bis zum Ende der 1980er-Jahre wurde die maschinelle Übersetzung im Wesentlichen auf der Grundlage sprachlicher Regeln entwickelt. Mit dem Erfolg von stochastischen Methoden in der Spracherkennung, wie zuvor beschrieben, begann eine Gruppe von Wissenschaftlern bei IBM in Yorktown Heights, New York, sich auch stochastischen Methoden in der Übersetzung zuzuwenden. Durch die grössere Verfügbarkeit an Rechenleistung und Trainingsdaten als auch die bessere Eignung zur Übersetzung von ungrammatikalischer Sprache, wie sie insbesondere häufig in der gesprochenen Sprache vorkommt, konnten sich diese Ansätze in den kommenden Jahren gegenüber den regelbasierten durchsetzen bzw. sich im Hybridansatz ergänzen. Um die Anzahl der Sprachpaare überschaubar zu halten – aktuelle Schätzungen nennen über 6000 Sprachen, aus denen sich eine unüberschaubare Anzahl von nahezu 4 Millionen Sprachpaaren ergeben und um das Fehlen ausreichender Parallel-

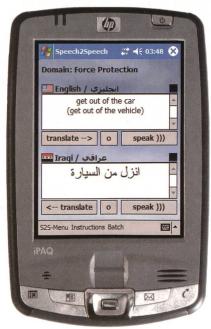


Bild 1 Ein PDA-Taschendolmetscher, entwickelt von Mobile Technologies Inc. in den USA.

texte zu kompensieren, wird zur Übersetzung vieler Sprachpaare auf eine generische Zwischensprache, die sogenannte Interlingua, zurückgegriffen.

Ende der 1980er- und Anfang der 1990er-Jahre kam es zur Entwicklung erster Systeme, die sich mit Sprach-zu-Sprach-Übersetzung (bezüglich Ausgangsmaterial bezieht sich «automatische Übersetzung» hier sowohl auf Text als auch auf Material in gesprochener Form) beschäftigten. Mit diesen konnte demonstriert werden, dass maschinelles Dolmetschen im Prinzip möglich ist. Hier sei angemerkt, dass beim Übersetzen der Ausgangstext fixiert, beim Dolmetschen aber nicht fixiert vorliegt. Somit stehen beim Übersetzen beliebig viele Ressourcen zur Verfügung, z.B. Zeit oder Wörterbücher, wohingegen beim Dolmetschen die Ressourcen, z.B. durch die kognitive Last oder Echtzeitanforderung, stark eingeschränkt sind. Hieraus entsteht bei der Humanübersetzung ein Qualitätsunterschied in der Genauigkeit von 99% bei der Übersetzung zu 80% beim Dolmetschen.

Durch die Gründung des Consortium for Speech Translation Advanced Research (C-STAR) im Jahre 1991 zur Förderung der internationalen Zusammenarbeit wurde das Interesse an der Sprach-zu-Sprach-Übersetzung weiter gestärkt. In der Mitte der 1990er-Jahre gab es eine Reihe von gross angelegten Projekten wie unter anderen das Verbmobil-Projekt, das, gefördert durch das Bundesministerium für Bildung und Forschung in Deutschland, sich zur Aufgabe gemacht hat, spontan gesprochene Sprache zu erkennen, in eine Fremdsprache zu übersetzen und die Übersetzung auszusprechen. Mit der Jahrtausendwende wurden die Forschungsrichtungen vielfältiger. So wurde zum einen weiterhin an der Verbesserung von domainbegrenzter 2-Wege-Übersetzung gearbeitet, um die Qualität, Portabilität und Zuverlässigkeit solcher Systeme zu steigern, zum anderen aber auch die Herausforderung angenommen, die Domainbeschränkung aufzuheben.

Taschendolmetscher

10

Erst das Mitführen eines elektronischen Dolmetschers erlaubt den flexiblen Einsatz vor Ort, wodurch sich neue Einsatzmöglichkeiten ergeben; z.B. im Gesundheitswesen, im Tourismus, aber auch bei humanitären Einsätzen in Krisengebieten. Da bei solchen Einsätzen oft eine geeignete Datenübertragungsmöglichkeit fehlt, ist es wünschenswert, alle Berechnungen auf dem Gerät selbst durchzuführen. Daher muss durch die beschränkte Rechenleistung und Speicherkapazität eine Domainbegrenzung in



Bild 2 Ein Taschendolmetscher auf Basis eines handelsüblichen tragbaren Computers, entwickelt von NICT/ATR in Japan.

Kauf genommen werden, die bei den zuvor genannten Beispielen durchaus akzeptabel ist.

Neben der eigentlichen Übersetzung ergeben sich bei solchen Geräten weitere Fragestellungen, insbesondere nach der Gestaltung einer sinnvollen und intuitiven

Benutzerschnittstelle mit entsprechender Rückmeldung (War die automatische Erkennung richtig? Ist die Übersetzung gut?). Mögliche grafische Benutzeroberflächen für Taschendolmetscher sind in den Bildern 1 und 2 dargestellt. Die grafische Benutzeroberfläche ist jeweils in 2 Regionen, je eine

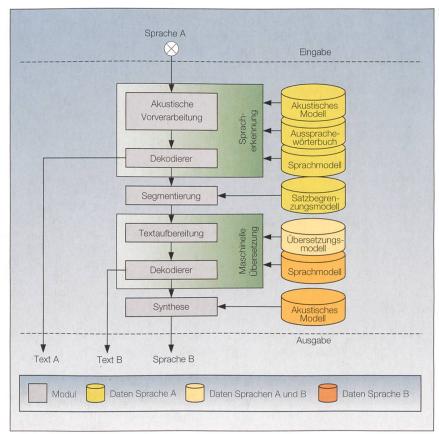


Bild 3 Systemübersicht der einzelnen Komponenten und Wissensbasen der Sprach-zu-Sprach-Übersetzung.

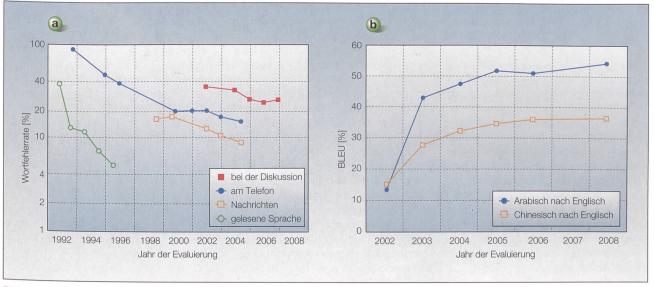


Bild 4 a) Qualität der automatischen Spracherkennung und b) der maschinellen Übersetzung im Verlauf der Zeit.

Region pro Sprache, eingeteilt. Durch eine «Push-to-talk»-Taste wird die Erkennung aktiviert und der erkannte Text im oberen Fenster angezeigt. Die Übersetzung erscheint im unteren Fenster, und der Lautsprecher gibt die Übersetzung als synthetisierte Sprache aus. Damit der Benutzer ein Gefühl bekommt, ob die Übersetzung brauchbar ist, liefert das System von Bild 1 eine Rückübersetzung (der Text in Klammern).

Universaldolmetscher

Die Übersetzung von parlamentarischen Reden, Vorträgen oder Nachrichten verlangt nach Systemen, die ohne Domainbegrenzung zuverlässig funktionieren. Durch den Verzicht auf die Beschränkung in Bezug auf das Vokabular, die Breite des Themas und des Sprachstils wird die Erkennung und Übersetzung erheblich schwieriger, da insbesondere eine grössere Anzahl von

Wörtern und Übersetzungsvarianten, aus denen jeweils ausgesucht werden muss, eine Verwechslung wahrscheinlicher machen.

Unter den besten Bedingungen – sehr viel Trainingsmaterial und Rechnerkapazität – sind Übersetzungen von nicht domainbegrenzter, gesprochener Sprache durchaus möglich. Die Machbarkeit wurde z.B. im europäischen Projekt TC-Star aufgezeigt und die Qualität solcher Systeme mit der eines Dolmetschers verglichen. Auf die Ergebnisse und die sich daraus ergebenden Schlussfolgerungen wird im Abschnitt «Qualität aktueller Systeme» genauer eingegangen.

Funktionsweise der Sprach-zu-Sprach-Übersetzung

Die Grundbausteine, mit denen Übersetzungsprogramme arbeiten, sind Wort-zu-Wort- bzw. Wortgruppen-zu-WortgruppenBeziehungen, die durch zusätzliche Regeln oder Statistiken (z.B. zur Konjunktion oder Wortumstellung) zur Kombinierbarkeit zu Sätzen angereichert sind. Daher muss jede gesprochene Äusserung zuerst einmal transkribiert werden. Nach der Übersetzung des Textes muss die Wortfolge wieder in ein Audiosignal gewandelt werden. Somit besteht eine Sprach-zu-Sprach-Übersetzung aus den 3 Kernkomponenten

- automatische Spracherkennung,
- maschinelle Übersetzung (heute insbesondere statistische Übersetzung) und
- Sprachsynthese,

um das Sprachsignal der Ausgangssprache in ein Sprachsignal der Zielsprache zu wandeln. Um eine möglichst gute Qualität zu garantieren, sollten diese Komponenten allerdings nicht getrennt betrachtet werden. Ein Diagramm der Sprach-zu-Sprach-Übersetzung ist in Bild 3 dargestellt. Da sich Sätze in verschiedenen Sprachen sowohl in der Wortstellung als auch Wortanzahl stark unterscheiden können und Wörter mehrere Übersetzungsmöglichkeiten haben, reicht es nicht, Einzelwörter wie bei einem Wörterbuch zu betrachten. Im Gegensatz zur regelbasierten Übersetzung mit der sogenannten morphologischen Zerlegung, bei der die einzelne Wortform auf eine Grundform zurückgeführt wird, werden bei der statistischen Übersetzung alle Wortformen aufgenommen und eine statistische Beziehung der Wortform hinterlegt. Um den Inhalt besser berücksichtigen zu können, werden bei der statistischen Übersetzung zusätzlich auch ganze Wortfolgen verwendet. Die festgelegten Regeln oder gelernten Statistiken sind für Sätze oder Teilsätze formuliert. Diese Satzgrenzen sind meistens bei gesprochener Sprache nicht festgelegt,

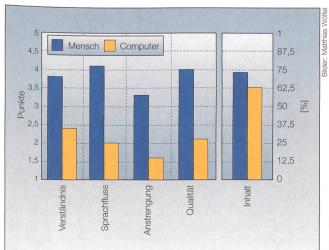


Bild 5 Vergleich der automatischen Sprachzu-Sprach-Übersetzung mit der menschlichen Übersetzung anhand verschiedener Kriterien.

Bulletin SEV/AES 11/2009

und somit müssen sinnvolle Satzgrenzen oder Teilsatzgrenzen automatisch gefunden werden (sowohl im Training als auch im Test).

Die für den Menschen akzeptable Latenz bei Sprach-zu-Sprach-Übersetzung ist kleiner als 6 Wörter, woraus folgt, dass die gefundenen Teilsätze eine Länge von 6 Wörtern nur selten überschreiten sollten. Für jedes Wort und jede Wortgruppe wird im Übersetzungsmodell festgelegt, welche Kombinationen wie wahrscheinlich sind. Ausgehend von der Struktur des ursprünglichen Satzes und den ausgewählten Übersetzungen für die einzelnen Wörter oder Wortgruppen, wird die Satzstruktur der Übersetzung durch Umgruppierung und Evaluation anhand des Zielsprachmodells aufgebaut und eventuell danach resynthetisiert.

Qualität aktueller Systeme

Um die Qualität verschiedener Verfahren vergleichen zu können, benötigt man Qualitätskriterien. Im Vergleich zum menschlichen Urteil werden, insbesondere in der Systementwicklung, automatische Verfahren bevorzugt, um die Turn-around-Zeit zu verkürzen und die Kosten zu senken. Eine Auswahl dieser Kriterien wird kurz beschrieben:

- Wortfehlerrate ist die Standardmetrik zur Messung der Qualität einer automatischen Spracherkennung und ist definiert als der normalisierte Fehler zwischen der Hypothese des Erkenners und der Referenztranskription der Ausgangssprache.
- BLEU und NIST sind Standardmetriken zur Evaluierung von automatischen Übersetzungssystemen. Beide Metriken basieren auf der Idee, die Distanz (z.B. Wortstellung, Unterschied in der Übersetzungslänge) zwischen einer oder mehreren Referenzübersetzungen und der automatischen Übersetzung zu vergleichen.
- Echtzeitfaktor beschreibt das Verhältnis zwischen der Dauer der Verarbeitungszeit und des Eingangssignals. (Für ein Echtzeitsystem muss der Echtzeitfaktor im Durchschnitt kleiner 1 sein.)
- Latenz beschreibt die Verzögerung eines Systems zwischen Ein- und Ausgabe, gemessen in Zeit oder Worten.

Bild 4 gibt die Qualitätsentwicklung, bewertet durch die Multimodal Information Group (www.itl.nist.gov/iad/mig) am NIST, von State-of-the-art-Spracherkennungsund Übersetzungssytemen, in Wortfehlerrate bzw. BLEU wider. Aus den Grafiken ist nicht nur ein Qualitätsunterschied bei der Erkennung, abhängig von der Domain, und der Übersetzung, abhängig vom Sprachpaar, zu entnehmen, sondern auch die kon-

tinuierliche Verbesserung (Schwankung abhängig von der Schwierigkeit des Testsets) der Systeme.

Durch zahlreiche Experimente wurde herausgefunden, dass es eine entgegengesetzte, nahezu lineare Korrelation zwischen Wortfehlerrate und BLEU gibt. Es wurde festgestellt, dass eine Wortfehlerrate von unter 10% zu einer angemessenen Übersetzung im Vergleich zur Übersetzung der Referenztranskriptionen führt.

Zwar konnte gezeigt werden, dass die zuvor beschriebenen automatischen Qualitätskriterien gut mit dem menschlichen Urteil korrelieren, jedoch sind aktuelle automatische Massnahmen nur bis zu einem gewissen Grad in der Lage, qualitative und semantische Unterschiede zu beurteilen. So werden bisher noch Fehler, die die semantische Bedeutung des gesamten Satzes zerstören, mit Fehlern, bei denen die semantische Bedeutung erhalten bleibt, gleichgesetzt. Daher ist eine menschliche Evaluation der Ergebnisse – insbesondere des Endergebnisses - unumgänglich. Eine solche Studie wurde von Hamon et al. (2007) für Übersetzungen von Reden im Europäischen Parlament vom Englischen ins Spanische durchgeführt [1]. Die Ergebnisse, in Bezug auf 5 verschiedene Aspekte, sind in Bild 5 zusammengefasst.

Während das Verständnis, der Sprachfluss, die Anstrengung (wie stark man sich konzentrieren muss) und die Textqualität hinter dem menschlichen Dolmetscher deutlich hinterherhinkt, ist in Bezug auf den Transfer des Inhalts ein Niveau erreicht, das dem des Dolmetschers schon sehr nahekommt. Der Verlust des Inhalts bei der automatischen Sprach-zu-Sprach-Übersetzung ist auf Fehler in der Erkennung und der Übersetzung zurückzuführen, wohingegen Fehler des Dolmetschers darauf zurückzuführen sind, dass er gelegentlich Informationen weglassen muss, um dem Sprachfluss weiter folgen zu können.

Ausblick

Zwar hat die Leistungsfähigkeit heutiger Spracherkennungs- und Übersetzungssys-

teme bei Weitem noch nicht die menschliche erreicht, jedoch ist für stark eingeschränkte Bereiche die Qualität der Spracherkennung und -übersetzung bereits so gut, dass erste kommerzielle Systeme auf dem Markt sind und sich in naher Zukunft als nützliche Helfer, z.B. auf dem Handy, bewähren werden. Im Bereich der universell einsetzbaren Systeme sind die Forschungsergebnisse ermutigend. Dieser Forschungsbereich bleibt aber weiterhin eine generell schwierige Materie, wo nur langsame Fortschritte zu erwarten sind, sofern es nicht zu einem unvorhersehbaren Durchbruch kommt. Somit bleiben qualitativ hochwertige Dolmetsch- und Übersetzungsaufgaben weiterhin eine Aufgabe von Menschen.

Die Sprachsynthese hat schon heute ein Niveau erreicht, bei dem die synthetische Sprache gut verständlich ist. Daher konzentriert sich die Forschung auf eine weitere Steigerung der Natürlichkeit, z.B. durch die Nachahmung der Intonation, die bisher kaum berücksichtigt wurde, und dem Gebiet der Sprachtransformation, um die synthetisierte Zielsprache der Sprecher der Ausgangssprache akustisch anzugleichen.

Referenz

 Hamon, Mostefa und Choukri: End-to-end evaluation of a speech-to-speech translation system in TC-STAR. Tagungsband Machine Translation Summit XI, 2007.

Weiterführende Literatur

Wölfel und McDonough: Distant Speech Recognition. Wiley & Sons, 2009.

Köhn: Statistical Machine Translation. Cambridge University Press, 2009.

Angaben zum Autor

Dr.-Ing. Matthias Wölfel hat an der Universität Karlsruhe (TH), Deutschland, der University of Massachusetts, USA, und der Carnegie Mellon University, USA, studiert. Sein Interesse umfasst die Mensch-Computer-und die Mensch-Computer-Mensch-Interaktion im Spannungsfeld zwischen Kunst und Wissenschaft, dem er im ZKM Zentrum für Kunst und Medientechnik, Karlsruhe, nachgeht.

Center for Art and Media, Lorenzstrasse 19, DE-76135 Karlsruhe, woelfel@zkm.de

Résumé

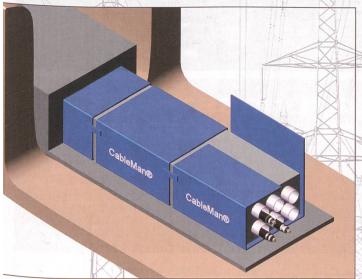
Des interprètes électroniques

Etat actuel de la traduction automatique de la parole. Avec la traduction automatique de la parole permettant de traduire un entretien quelconque en simultanée au moyen d'une machine, on s'approche du rêve bien ancien de pouvoir parler et comprendre une langue étrangère sans la maîtriser soi-même. Mais comment un tel système fonctionnet-il? Où en est le développement et quelle en est actuellement l'utilité? L'article examine ces questions, expose les bases, cite des exemples d'application et renseigne sur la qualité actuelle de tels systèmes.



CableMan[®], gebaut für den Energietransport der Zukunft

ohrblock im Kabelgraben, abgeschirmt mit CableMan



Die herausragenden Merkmale von CableMan

- hohe Schirmwirkung gegen elektromagnetische Felder
- · eliminiert bis zu 99% der Emissionen
- Ausführungen sowohl für neue als auch für bestehende Erdkabel und Rohrblöcke
- einfache, schnelle und platzsparende Montage
- erfüllt die NISV 2000 auch in extremsten Situationen
- Typengeprüft durch die FKH

CFW EMV-Consulting AG

Nordstrasse 24 CH-9410 Heiden Telefon ++41-71-891 57 41 Telefax ++41-71-891 57 43

info@cfw.ch www.cfw.ch



Magnetfeld Technologie



Many electrical products in Europe are not safe – **we have the solution** ...

ENEC is **the** independent high quality European safety mark for electrical products that demonstrates compliance with EU requirements

Electrosuisse
Luppmenstrasse 1, CH-8320 Fehraltorf
Phone +41 44 956 13 14
testing@electrosuisse.ch

www.electrosuisse.ch

electrosuisse »



Gut, dass wir auf Magnetfelder spezialisiert sind.

- Messung
- Berechnung
- Planung / Ausführung von Abschirmungen

Egal, ob Sie gesetzliche Grenzwerte, oder – im Sinne der Vorsorge – tiefstmögliche Werte für niederfrequente Magnetfelder einhalten müssen, bei uns erhalten Sie Lösungen aus einer Hand.

Wir bieten neben Beratung und Planung von Massnahmen auch die fachgerechte Montage von Abschirmungen.

Zudem führen wir normenkonforme Messungen und Berechnungen von Anlagen durch.









Flächenabschirmung



Kompaktabschirmung

Ihr kompetenter Partner rund um Magnetfelder.



Systron EMV GmbH Bubikonerstr. 45a · 8635 Dürnten Tel. 055 250 53 20 · Fax 055 250 53 21 www.systron.ch



rittmeyer KOMPETENT AUS ERFAHRUNG

Wirtschaftlichkeit und Sicherheit für Mensch und Umwelt – Wir stimmen diese Anforderung optimal aufeinander ab und lösen für Sie diese zentrale

Aufgabe beim Bau und der Modernisierung von Wasserkraftwerken.

www.rittmeyer.com