Höranstrengung als Messverfahren für die Evaluierung von Near-End Listening Enhancement Algorithmen

Arne Pusch[1], Jan Rennies[1], Henning Schepker[2], Simon Doclo[1,2]
[1] *Fraunhofer IDMT, Projektgruppe Hör-, Sprach- und Audiotechnologie, Oldenburg*
Email: arne.pusch@idmt.fraunhofer.de
[2] *Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Universität Oldenburg*

Einleitung

In verschiedenen Anwendungen wird Sprache über Lautsprecher mit dem Ziel wiedergegeben, Informationen anderen Pesonen mitzuteilen. Hierzu zählen z.B. die Durchsagen in einem Bahnhof oder Telefongespräche. Erfolgt die Wiedergabe im Beisein von Störgeräuschen, wird die Verständlichkeit erschwert. Dieser erschwerten Verständlichkeit kann entgegengewirkt werden, indem die Sprache auf geeignete Weise vorverarbeitet wird. Verschiedene Strategien dieser Vorverarbeitung wurden in Form sogenannter Near-End Listening Enhancement Algorithmen von verschiedenen Forschungsteams entwickelt (z.B. [3], [8] [10]). Als Evaluierung wurde bisher hauptsächlich die Messung der Sprachverständlichkeit verwendet (z.B. [2]). Diese ist zwar sehr etabliert und bietet daher eine gute Vergleichbarkeit, hat jedoch einen großen Nachteil. So kann nur mit niedrigen Signal-Rausch-Abstand (SNR) gemessen werden. Ist der SNR zu hoch, wird bereits die unverarbeitete Sprache vollständig verstanden und es ist kein Mehrwert durch den Algorithmus messbar [5] [7]. In vielen Anwendungsszenarien treten solche niedrigen SNR jedoch nur selten auf. Es ist daher von Interesse, ob NELE-Algorithmen auch einen Vorteil bieten, wenn realistische, also bessere SNR auftreten. Die Messung der Höranstrengung hat das Potential auch sensibel für diese SNR zu sein. Aus diesem Grund wurde untersucht, inwiefern sie sich als Alternative zur Messung der Sprachverständlichkeit eignet. Hierfür wurde für einen zuvor entwickelten NELE-Algorithmus sowohl Höranstrengung als auch Sprachverständlichkeit über einen großen SNR-Bereich gemessen.

Methoden

Verwendeter NELE-Algorithmus

Als zu untersuchender NELE-Algorithmus wurde AdaptDRC verwendet (genauere Beschreibung: [8]) Dieser Algorithmus verwendet eine Blockverarbeitung mit einer Fensterlänge von 20 ms. Jeder Block durchläuft zwei Verarbeitungsstufen. In der ersten Stufe wird eine Änderung des Frequenzspektrums vorgenommen. Hierzu wird zunächst das Sprachsignal und das Umgebungsgeräusch in acht Oktavbänder von 125 Hz bis 16 kHz unterteilt. Anhand der Pegel dieser Bänder wird eine vereinfachte Version des SII (Speech Intelligibility Index) berechnet. Der SII gibt vor, wie die einzelnen Bänder gewichtet werden. Beträgt der SII 1 wird keine Gewichtung vorgenommen. Beträgt der SII 0 erfolgt eine Gewichtung, welche die gleiche Leistung in allen Bändern zu Folge hat. Bei Werten zwischen diesen Extremen findet ein kontinuierlicher Übergang zwischen keiner Gewichtung und einem flachen Oktavband-Spektrum statt. In der zweiten Stufe wird ein Kompressor mit dem Ziel angewandt, leise Passagen für eine bessere Verständlichkeit anzuheben. Das Maß der Anhebung ist dabei abhängig vom SNR. Die Leistung des breitbandigen Ausgangssignal des Algorithmus entspricht der des Eingangssignals, sodass keine breitbandige Pegelerhöhung erfolgt.

Probanden

An beiden Experimenten nahmen dieselben elf Probanden teil. Alle Probanden waren normalhörend mit einer durchschnittliche Hörschwelle im Audiogramm von < 25 dB HL. Das Alter der Probanden war zwischen 24 und 36 Jahren (Median 27 Jahre).

Stimuli und Equipment

Das Sprachmaterial entstammt dem Oldenburger Satz-Test [9]. Es besteht aus Sätzen mit jeweils fünf Wörter der Struktur: Name – Verb – Zahl – Adjektiv – Objekt (z.B.: „Peter kauft drei weiße Tassen.“). Für jedes Wort gibt es zehn Alternativen, welche zufällig ausgewählt werden. Die Wörter sind semantisch nicht vorhersehbar. Die Sätze wurden entweder unverarbeitet wiedergegeben oder mit dem NELE-Algorithmus AdaptDRC [8] bearbeitet. Der Sprachpegel war auf 60 dB SPL fixiert. Es wurden zwei verschiedene Störgeräusche verwendet. Ein stationäres Rauschen, dessen Langzeitspektrum dem der Sprache entspricht (speech-shaped noise: SSN) und ein Cafeteria-Geräusch, welches fluktuierende Eigenschaften besitzt. Der Pegel der Störgeräusche wurde variiert, um den gewünschten SNR zu erzielen. Die SNRs für die Messung der Sprachverständlichkeit wurden so gewählt, dass jeweils ca. 20 bzw. 80% verstanden wurde. Für die Messung der Höranstrengung wurde ein breiter SNR-Bereich von -15 bis 10 dB gewählt. Die SNR wurden für beide Experimente anhand von Pilotmessungen ausgewählt. Die Signale wurden in Matlab gemischt, anschließend D/A gewandelt (RME ADI-8 PRO), verstärkt (DT HB7) und diotisch über Kopfhörer (Sennheiser HD650) in einer schallbedämpften Kabine wiedergegeben.

Messverfahren

Alle Probanden starteten mit der Messung der Höranstrengung. Hierbei wurde den Probanden ein Satz

in einer Schleife wiedergegeben und die Aufgabe gestellt, die Anstrengung für das Verstehen des Satzes zu bewerten. Dabei wurden die Probanden instruiert, den Satz mindestens einmal vollständig anzuhören. Die verwendete Fragestellung lautete: „ Wie anstrengend ist es für Sie, die Sprache zu verstehen?“. Die Probanden gaben ihre Einschätzung auf einer 13-stufigen Skala über eine GUI ab. Die Skala erstreckte sich von ‚mühelos‘ (1 Effort Scaling Categorial Unit, ESCU) bis hin zu ‚extrem anstrengend‘ (13 ESCU) [4]. Zusätzlich gab es eine 14. Kategorie (‚nur Störgeräusch‘), falls der Proband keine Sprache detektieren konnte. Die verwendete Kondition, das Start-Sample des Störgeräuschs und der verwendete Satz wurden zufällig ausgewählt. Jede Kondition wurde dabei im Laufe des Experiments sechs mal gemessen. Die Probanden hatten die Möglichkeit, ausreichend Pausen einzulegen.

Bei der Messung der Sprachverständlichkeit wurde den Probanden ein Satz einmal vorgespielt und die Aufgabe gestellt, diesen Satz mündlich zu wiederholen. Die Anzahl richtig verstandener Wörter wurden gemessen. Die Probanden erhielten kein Feedback über diese Anzahl. Anschließend erfolgte die Wiedergabe des nächsten Satzes mit der selben Kondition. Es wurden 20 Sätze pro Kondition gemessen. Das Start-Sample des Störgeräuschs und der verwendete Satz wurde zufällig ausgewählt. Nach diesen 20 Sätzen wurde die nächste Kondition gemessen. Dabei wurde die Reihenfolge der Konditionen zufällig ausgewählt. Die Probanden absolvierten zwei Listen mit jeweils 20 Sätzen mit unverarbeiteter Sprache als Training, bevor das eigentliche Experiment begann.

Ergebnisse

Sprachverständlichkeit

Die untere Hälfte von Abbildung 1 zeigt die Mittelwerte der Sprachverständlichkeitsmessung für das SSN-Störgeräusch (links) und das Cafeteria-Störgeräusch (rechts). Die Fehlerbalken entsprechen plus und minus einer Standardabweichung. Für jeden Probanden wurde eine psychometrische Funktion geschätzt, indem eine Sigmoid-Funktion an die Daten gefittet wurde [1]. Die dargestellten Kurven entsprechen den parametrisch über die Probanden gemittelten psychometrischen Funktionen. Für beide Arten des Störgeräuschs wurde eine Verbesserung der Sprachverständlichkeit durch Verwendung des Algorithmus erzielt, was durch die Verschiebung der psychometrisches Funktion nach links zu erkennen ist. Die mittlere Verschiebung an der Sprachverständlichkeitsschwelle (50 % korrekt) betrug 10,8 dB für SSN und 5,8 dB für Cafeteria.

Für die statistische Analyse wurden zunächst anhand der individuellen psychometrischen Funktion die SNR berechnet, welche zu 20, 50 und 80 % Sprachverständlichkeit führen. Da diese Daten normalverteilt waren, konnte eine dreifaktorielle ANOVA mit den Faktoren: Art des Störgeräuschs, Verarbeitung der Sprache und Wert der Sprachverständlichkeit durchgeführt werden. Das Signifikanzlevel betrug 0,05 und die Freiheitsgrade wurden Greenhouse-Geissler korrigiert. Die Analyse ergab, dass alle drei Faktoren einen signifikanten Einfluss auf den SNR hatten (Störgeräusch: $F(1,10) = 26.697$, $p<0.001$; Verarbeitung: $F(1,10) = 1049.046$, $p<0.001$; Sprachverständlichkeit: $F(1,10.001) = 108.717$, $p<0.001$). Post-hoc-Tests zeigten eine signifikante Verrginerung der SNR, die zu 20, 50 und 80 % Sprachverständlichkeit führten, aufgrund der NELE-Verarbeitung. Als einen repräsentativen SNR dafür, ab dem das Messverfahren gesättigt ist wurde der SNR ermittelt, bei dem unverarbeitete Sprache zu 95 % verstanden wird. Dieser SNR lag bei -4,3 dB für SSN und 0,1 dB für Cafeteria.

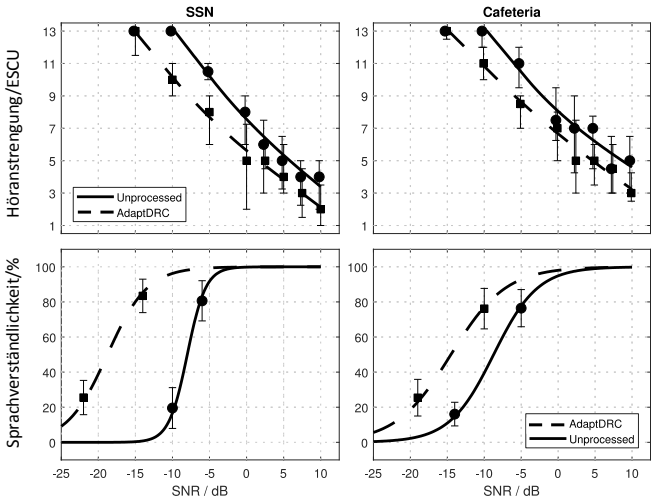

Abbildung 1: Oben: Median der Höranstrengungsskalierung. Fehlerbalken entsprechen Interquartilsabständen. Unten: Mittelwert der Sprachverständlichkeit. Fehlerbalken entsprechen ±1 Standardabweichung.

Höranstrengung

Die Symbole der oberen Hälfte von Abbildung 1 zeigen den Median der Höranstrengungsbewertung aller Probanden. Die Fehlerbalken entsprechen den Interquartilsabständen. Die Linien entsprechen einer an die Mediane gefitteten Funktion (die Fitting-Funktion enstammt aus [4]). Für die unverarbeitete Sprache wurden die Daten für den SNR von -15 dB nicht berücksichtigt, da bereits bei -10 dB eine Bewertung von 13 ESCU erreicht wurde, die Messung also bereits gesättigt war. Insgesamt nahm die Höranstrengung mit steigendem SNR ab. Die Höranstrengung war für die verarbeitete Sprache immer geringer als ohne Verarbeitung. Für SNR kleiner 0 dB konnte die Höranstrengung um 2 bis 3 ESCU, für SNR größer 0 dB um 1 bis 2 ESCU reduziert werden. Ähnlich wie bei der Analyse der Sprachverständlichkeit wurden aus den individuellen Funktionen die SNR berechnet, die zu den einzelnen ESCU Werten gehören (Schrittweite 1 ESCU). Aufgrund der beschriebenen Sättigungseffekte bei 13 ESCU und der geringen Anzahl an Daten unterhalb 4 ESCU, wurde für die Analyse nur der Bereich zwischen 4 und 12 ESCU verwendet. Mit diesen Daten wurde eine dreifaktorielle

ANOVA mit den Faktoren: Art des Störgeräuschs, Verarbeitung und Höranstrengungsbewertung durchgeführt. Die Analyse ergab, dass die Faktoren Verarbeitung ($F(1,10)=52.196$, $p<0.001$) und Höranstrengungsbewertung ($F(1.088,10.882)=80.886$, $p<0.001$) signifikant waren, der Faktor Störgeräusch hingegen nicht ($F(1,10)=2.098$, $p<0.178$). Die Interaktion von Verarbeitung und Art des Störgeräuschs war nicht signifikant, was auf einen ähnlichen Gewinn durch den Algorithmus für beide Störgeräusche hinweist. Die Interaktion von Höranstrengungsbewertung und Art des Störgeräuschs war ebenfalls signifikant, was darauf hindeutet, dass der Gewinn abhängig von der Position auf der psychometrischen Funktion war.

Für die Post-hoc-Tests wurden die Störgeräusche separat betrachtet und gepaarte t-Test durchgeführt, um zu ermitteln, bei welchen Werten der Höranstrengung ein signifikanter Gewinn durch den Algorithmus erzielt wurde. Überprüft wurde dabei, ob durch die Verarbeitung der Sprache ein signifikant unterschiedlicher (d.h., geringerer) SNR zur selben Bewertung der Höranstrengung führte. Aufgrund des multiplen Testens, wurde das Signifikanzlevel auf $0{,}05/18 = 0{,}0028$ angepasst. Für SSN waren die Unterschiede bis auf 4 ESCU ($p=0{,}005$) signfikant ($p<0{,}001$). Für Cafeteria waren die Unterschiede zwischen 6 und 11 ESCU signifikant ($p\leq 0{,}002$), während sie für 12 ESCU ($p=0.008$), 5 ESCU ($p=0{,}007$) und 4 ESCU ($p=0{,}039$ nicht signifikant waren.

Bei den SNR, bei dem die Messung der Sprachverständlichkeit Deckeneffekte zeigte (95 % Sprachverständlichkeit) lag eine Höranstrengungsbewertung von 10,2 ESCU für SSN und 7,5 ESCU für Cafeteria vor.

Diskussion

Einige Konditionen dieser Studie wurden in vorherigen Messungen mit anderen Probanden untersucht. Zu diesen Messungen zeigen sich lediglich geringe Abweichungen [6][5][8]. Dies deutet darauf hin, dass beide Messverfahren reproduzierbare Ergebnisse liefern. Die Messung der Sprachverständlichkeit zeigt auf, dass ab einem SNR von -4 dB (SSN) bzw. 0 dB (Cafeteria) ca. 95 % der unverarbeiteten Sprache verstanden wird. D.h. ab diesen SNR ist kein Nutzen von NELE-Algorihmen mit diesem Messverfahren mehr zu identifizieren. Im Gegensatz dazu liegt bei diesen SNR eine Bewertung der Höranstrengung vor, welche 7 ESCU nicht unterschreitet. Hier ist demnach noch ca. die Hälfte der Höranstrengungs Skala übrig, um Aussagen über höhere SNR machen zu können. Die niedrigste Höranstrengungsbewertung, bei der noch eine signifikante Verbesserung durch die Vearbeitung der Sprache erzielt werden konnte lag bei ca. 5,5 dB SNR für unverarbeitete Sprache (bei beiden Störgeräuschen). Dieser SNR ist über 10 dB höher als der SNR bei dem die Sprachverständlichkeit für SSN gesättigt ist, bzw. ca. 5,5 dB bei Cafeteria. Für noch höhere SNR führte die Verarbeitung der Sprache zu keiner signifikanten Reduzierung der Höranstrengung. Dies passt zu der Eigenschaft des AdaptDRC Algorithmus, das Maß der Verarbeitung bei guten SNR zu reduzieren.

Bei sehr schlechten SNR hingegen ist die Messung der Höranstrengung gesättigt, während die Messung der Sprachverständlichkeit noch signifikante Ergebnisse liefert. Insgesamt ist demnach die Messung der Höranstrengung sehr gut geeignet für mittlere und hohe SNR und Sprachverständlichkeit für niedrige SNR. Auch wenn in dieser Studie nur ein einzelner Algorithmus untersucht wurde, gibt es keine Hinweise dafür, dass die Verfahren bei anderen Algorithmen weniger geeignet sein sollten.

Zusammenfassung

In dieser Studie wurde untersucht, inwiefern sich die Messung der Höranstrengung zur Evaluierung von NELE-Algorithmen eignet. Hierfür wurde ein NELE-Algorithmus mittels einer kategorialen Messung der Höranstrengung und zusätzlich mit der etablierten Methode der Messung der Sprachverständlichkeits untersucht. Es zeigte sich, dass die Verarbeitung der Sprache sowohl zu einer geringeren Bewertung der Höranstrengung, als auch zu einer erhöhten Sprachverständlichkeit führt. Die Analyse beider Messverfahren ergab, dass die Messung der Höranstrengung sensibel für höhere SNR ist als die Messung der Sprachverständlichkeit. So ist die Sprachverständlichkeit für SSN (speech-shaped noise) ab -4 dB gesättigt, da bereits die unverarbeitete Sprache zu ca. 95 % verstanden wird. Mit der Messung der Höranstrengung lassen sich statistisch signifikante Verbesserungen durch Einsatz des Algorithmus bis zu einem SNR von ca. 5,5 dB nachweisen. Dieser erweiterte Messbereich enthält SNR, welche häufig in typischen Anwendungszenarien auftreten. Besteht Interesse an der Evaluierung von NELE-Algorithmen für solche SNR, eignet die Messung der Hörantrengung daher als gute Alternative bzw. Ergänzung.

Danksagung

Diese Studie wurde unterstützt durch die Deutsche Forschungsgemeinschaft (DFG) durch Forschergruppe FOR132 Ïndividualized Hearing Acoustics", Exzellenzcluster 1077 "Hearing4All" und Forschungsstipendium RE 4160/1-1.

Literatur

[1] Brand, T. and Kollmeier, B. (2002), Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests, J. Acoust. Soc. Am., 111(6), p. 2801–2810.

[2] Cooke, M., Mayo, C., and Valentini-Botinhao, C. (2013). Intelligibility enhancing speech modifications: The Hurricane Challenge, in Proceedings of Interspeech, Lyon, France, pp. 3552-3556.

[3] Kleijn, W. B., Crespo, J. B., Hendriks, R. C., Petkov, P., Sauert, B., and Vary, P. (2015). Optimizing speech intelligibility in a noisy environment: A unified view, IEEE Signal Process. Mag. 32(2), 43-54.

[4] Krueger, M., Schulte, M., Zokoll, M., Wagener, K., Meis, M., Brand, T., and Holube, I. (2017). Relation between listening effort and speech intelligibility in noise, Am. J. Audiol. 26, 378-392.

[5] Rennies, J., Drefs, J., Hülsmeier, D., Schepker, H., and Doclo, S. (2017). Extension and evaluation of a near-end listening enhancement algorithm for listeners with normal and impaired hearing, J. Acoust. Soc. Am. 141, 2526-2537.

[6] Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation, J. Acoust. Soc. Am. 136, 2642-2653.

[7] Schepker, H., Haeder, K., Rennies, J., and Holube, I. (2016). Perceived listening effort and speech intelligibility in reverberation and noise for hearing-impaired listeners, Int. J. Audiol. 55, 738-747.

[8] Schepker, H., Rennies, J., and Doclo, S. (2015). Speech-in-noise enhancement using amplication and dynamic range compression controlled by the speech intelligibilility index, J. Acoust. Soc. Am. 138, 2692-2706.

[9] Wagener, K., Brand, T., and Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III: Evaluation des Oldenburger Satztests (Development and evaluation of a German sentence test Part III: Evaluation of the Oldenburg sentence test), Z. Audiol. 38(3), 86-95.

[10] Zorila, T.-C., and Stylianou, Y. (2014). On spectral and time domain energy reallocation for speech-in-noise intelligibility enhancement, in Proceedings of Interspeech, Singapore, 328 pp. 2050-2054.