Comparative study between Speaksee and alternative transcription apps

Goal: At the end of 2022 a comparative study with Speaksee equipment was conducted. This comparative study was conducted by clinical physicist audiologist Jan-Willem Wasmann of the RadboudUMC in Nijmegen. During this comparative study, Speaksee's speech-to-text performance was compared with two other state-of-the art systems, called NALScribe (Apple speech recognition) and Google Live Transcribe. During an earlier study in 2020, the RadboudUMC conducted the same tests for AVA, Earfy and Speechy. These research results in the comparison will be included.

Working method: The Dutch audiological speech tests which were performed, were the speech in noise tests, i.e., the PLOMP and Matrix test. Also, the Dutch and English transcription accuracy was tested with the WER test.

Results: In the PLOMP test Speaksee had the lowest (best) signal-to-noise ratio -11dB in separated noise. In the other five apps, speech volume could not be made lower than noise. Speaksee also achieved the lowest SRT+1dB in frontal noise. In the MATRIX test, the lowest SRT measured was again for Speaksee +2 dB. Speaksee had no difference between the result with a fixed noise or speech level. In the WER, of the six different systems, Speaksee achieved the lowest (best) error rate, this for both Dutch (2%) and English (3%). While Speaksee only makes 2 errors per 100 words, other apps in Dutch have a 10 times higher rate.

Conclusion: For both speech-in-noise tests and transcription accuracy, Speaksee achieves the best results. These results confirm that Speaksee is the best app for situations with background noise. The SRT in separated noise of Speaksee equals with the SRT of a well-hearing person in background noise. Speaksee also makes less errors in its transcription compared to alternatives, both in Dutch and English dialogue.

Speaksee
SEE WHAT YOU CAN'T HEAR

Introduction

Despite the use of hearing aids, people with hearing loss experience difficulties with speech understanding, especially in group conversations. If background noise is present or if the room has poor acoustics, it becomes almost impossible to follow a conversation. Think of noisy restaurants, social gatherings, meetings, etc. This makes it difficult to participate in society and can put this population at risk of social isolation[1].

Speaksee aims to enable this population to actively and fully participate in group conversations and experience pleasure in communicating with others. The Speaksee Microphone Kit enables participation in (group) conversations by converting speech to text, regardless of background noise or distance from the speaker. Each speaker is represented in a different colour. We do this using beamforming microphones.

In 2022, a pilot with the Speaksee equipment was conducted, in which 29 deaf and hard of hearing people participated. The deaf and hard of hearing people who participated in the pilot mostly experienced added value from using Speaksee. The supervising audiologists requested a study of how this technology works and how Speaksee performs compared to alternative transcription apps. This comparative study was conducted by clinical physicist audiologist Jan-Willem Wasmann of Radboud UMC.

The comparative study involved the following: comparison of speech-to-text performance of Speaksee versus two other state-of-the-art systems (NALscribe and Google Live Transcribe); description of the user experience of various employees of the Audiology Center at the Radboudumc with Speaksee in the consulting room.
Based on the findings of the comparative study, this white paper was written by Karlien Vanpoecke, product specialist and master of audiology at Speaksee.

System description

Speaksee Microphone Kit consists of a base station and three microphones. Each microphone has its own colour. Each speaker wears a microphone. The microphone hangs from a cord at the level of the speaker's chest. The microphones use beamforming directed toward the mouth. The beamforming ensures that speech is picked up stronger than ambient noise. The speech is sent via the base station to Speaksee's cloud environment. Speaksee's cloud environment has servers that analyse the speech signal to I) recognize and convert speech into text and II) accurately distinguish different speakers through colour coding. The servers send the text and speaker identification to a user's smartphone, tablet or laptop, allowing the user to read along with the conversation in real time. The base station and the three microphones are battery-powered to allow mobility. The base station can also charge the microphones on the go.

Measurement setup

The microphone of the smartphone (iPhone and Samsung) was placed at 1 meter distance to make a comparison with a (human) listener at 1 meter distance which is usual in audiology. For the measurements at 1 m distance, the microphone was placed on a tripod, with the microphone positioned at "ear level". The (telephone) microphone pointed toward the speaker. In normal Speaksee use, the microphones are positioned hanging by the cord about 30 cm from the speaker's mouth. Speaksee's equipment is also designed and optimized for use of the microphones at 30 cm. Therefore, this usage situation was used in the studies. The same Internet connection was used for all measurements. All data ran through a TP Link Mi-Fi single band 2.4 GHz router with 4G SIM card. The router was paired with the equipment (always one active device per measurement) and placed close to the equipment. The strength of the Internet connection was monitored during the measurements.

Results

During the comparative study, several studies such as PLOMP, MATRIX and WER, were conducted with Speaksee and alternatives, with the purpose of identifying the quality of transcription.
In June 2020, measurements at Radboud UMC were conducted to evaluate the speech-to-text performance of different speech-to-text apps on standard audiometric tests[2]. At that time, Speaksee was not on the market yet and therefore not part of the research. The results of the other apps (Earfy, AVA and Speechy) in 2020 also were included in these results.

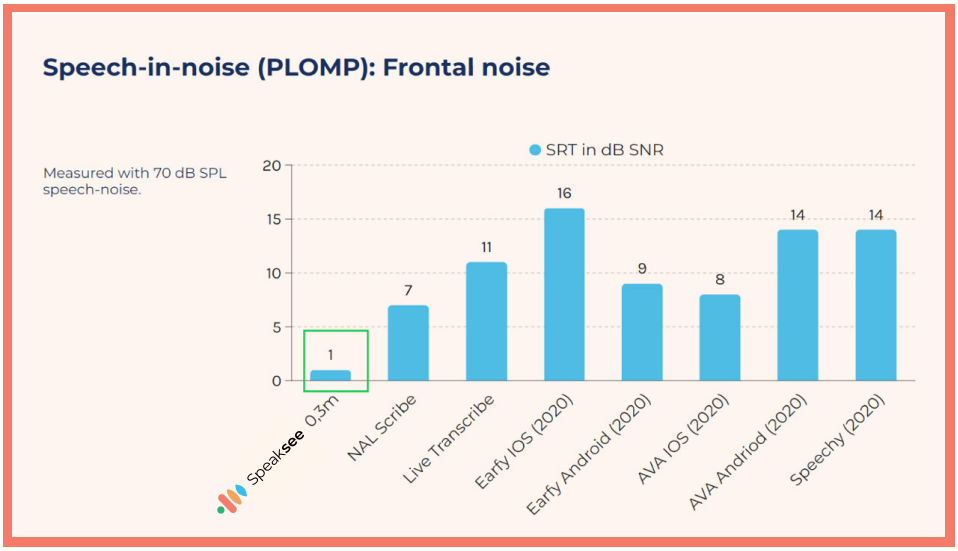

Figure 1: PLOMP Test: Frontal Noise. Comparison between Speaksee and alternative apps. Measured at a constant noise level of 70 dB SPL

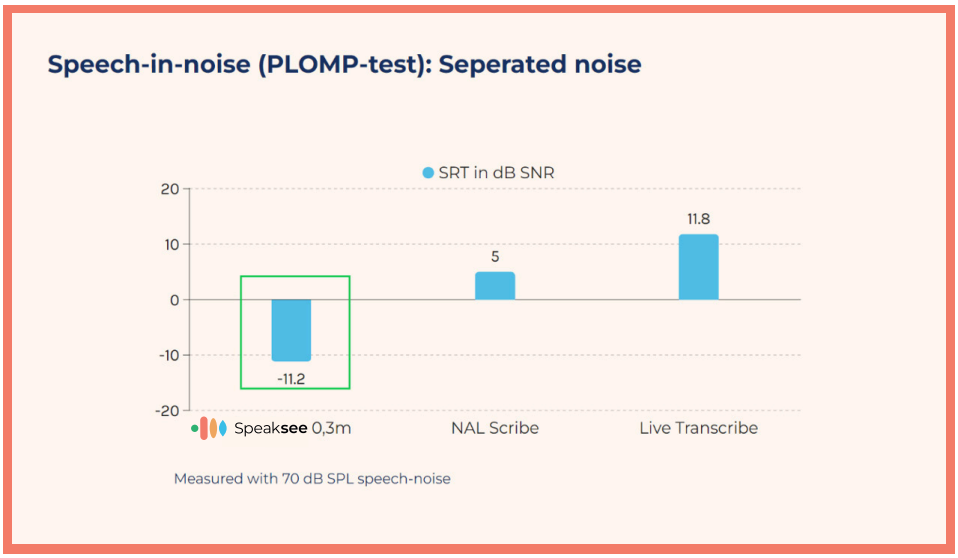

Figure 2: PLOMP Test: Seperated Noise. Comparison between Speaksee and alternative apps. Measured at a constant noise level of 70 dB SPL

Sentences-in-noise: Speaksee works better than other apps with background noise

PLOMP-Test

To investigate speech understanding in background noise, the Dutch PLOMP test was used. This involves 13 sentences of 8-9 syllables presented in noise. The noise has the same spectrum as speech. A sentence is scored correctly when it is displayed completely correctly in the app. If words are missing, incorrect or added, the sentence is scored as incorrect. If a sentence was scored correctly, the speech level became lower; if it was scored incorrectly, the speech level became louder.

The Speech Reception Threshold (SRT) is determined. This is the noise level in dB SNR (signal-to-noise ratio) at which 50% of the sentences are properly reproduced. A continuous speech noise level of 70 dB SPL was used.

The PLOMP test was administered in two different conditions:

- Frontal noise: in which speech and noise come from the same direction, angle of 0 degrees
- Separate noise: where the speech and the noise come from the opposite direction, angle of 180 degrees

Note: The measurement in frontal noise is a less representative measurement of reality, since in everyday life speech and noise usually do not come from the same direction. The measurement in separated noise is more representative of reality to measure speech in background noise.

Frontal noise, see Figure 1:

The frontal noise results show that when using Speaksee, the SRT is at +1 dB. The speech level can be made almost as low as the noise level, and 50% of the sentences are still rendered well. With the other apps, the speech level must be made much louder than the noise, before 50% is properly rendered.

Separated noise, see Figure 2:

The separated noise results show that when using Speaksee, the SRT is at -11 dB SNR. The speech level can be made much lower than the noise level, and then 50% of the sentences are still rendered well. With the other apps, the speech level cannot be made softer than the noise.

Matrix test

The sentence material from the PLOMP test is from the 80s, and no longer fully representative of the modern Dutch language. Therefore, the Matrix test was also conducted. The Matrix test contains more common Dutch words and is scored per correct word rather than per sentence. The sentences from the Matrix test consist of 5-word sentences namely, name, verb, numeral, adjective and a noun e.g.: 'Anneke wins three big boxes'.

Testing was done with a fixed speech level (fixed speech level, see Figure 3), in which the strength of speech noise was varied, and with a fixed speech noise level (fixed noise level, see Figure 4), in which the strength of speech was varied.

The first sentence is presented with a signal-to-noise ratio (SNR) of 0 dB. For subsequent presentations, the speech or noise level is adjusted according to the patient's prior response. This is automatically done by the software. If the test person correctly repeats three to five words of the words presented, the speech or noise level of the next presentation is reduced. If the test person repeats less than three words correctly, the speech or noise level of the next presentation is increased.

Speaksee performs best compared to the other apps, with no difference between the result with a fixed noise or speech level for Speaksee. This seems to indicate that varying speech and/or noise level does not affect Speaksee's performance, which is not the case with Live Transcribe. The result with NALscribe with fixed speech level is missing because it was not possible to take it down.

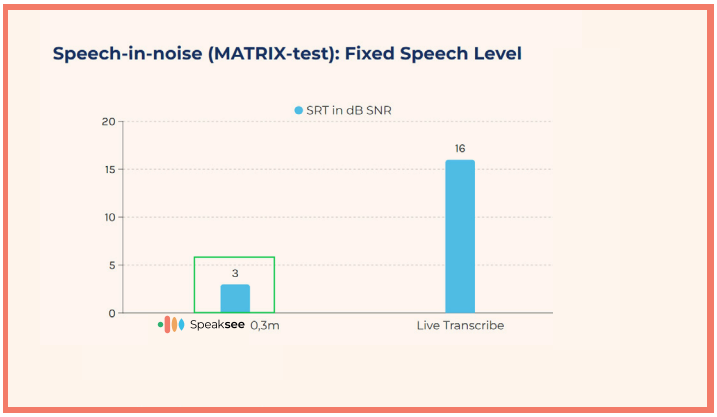

Figure 3: Matrix- Test: Fixed Speech Level. Speaksee compared to alternatives. Measured at 70 dB SPL speech level. No data is available for NAL-Scribe.

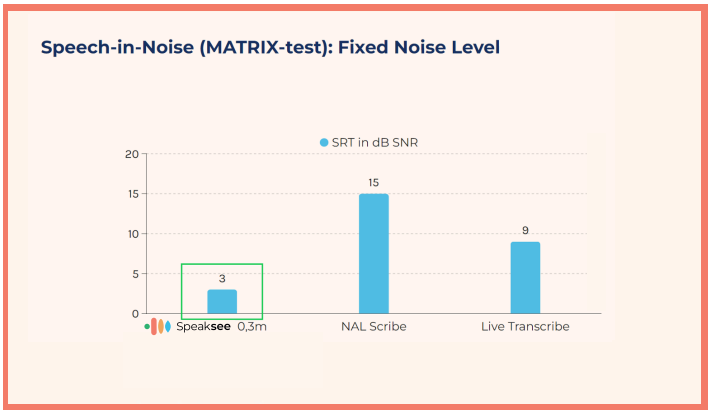

Figure 4: Matrix Test: Fixed Noise Level.
Speaksee compared to alternatives.
Measured in 70 dB SPL speech noise.

Speaksee translates less words incorrectly than other apps in dialogue

Word Error Rate- Test

The most commonly used measure of automated speech recognition performance is the Word Error Rate (WER). WER is the ratio of errors in a transcript to the total number of words spoken. The lower the WER, the more accurate the transcription.

For example, a WER of 10% means that the transcript is 90% accurate, meaning that 1 out of every 10 words are translated incorrectly.

The WER is calculated by adding up the number of missing, incorrect and added words in the transcription and dividing this by the total number of spoken words. In the technical validation, Dutch and English dialogues were played and the errors in the transcript were counted. These dialogues mimicked a 1-on-1 conversation.

The Dutch dialogue was an introduction video from RadboudUMC with a female voice speaking clearly and at a normal pace.
https://www.youtube.com/watch?v=zBJBD1-ePRw .

For the English dialogue, part of an advanced English tutorial was played. This video featured a conversation between a male and female voice.
https://www.youtube.com/watch?v=JtMgw2rCYSo&t=1s.

The Dutch dialogue consisted of 256 words, while the English dialogue consisted of 248 words.

WER Dutch-language dialogue, see Figure 5

The results from the WER test show that Speaksee makes the least mistakes in transcribing Dutch compared to the other apps. Whereas Speaksee makes only 2 errors per 100 words, is this 10 times higher for the other apps.

WER English dialogue, see Figure 6

Speaksee also scores the best on the WER in English compared to other apps. Speaksee makes only 3 errors per 100 words.

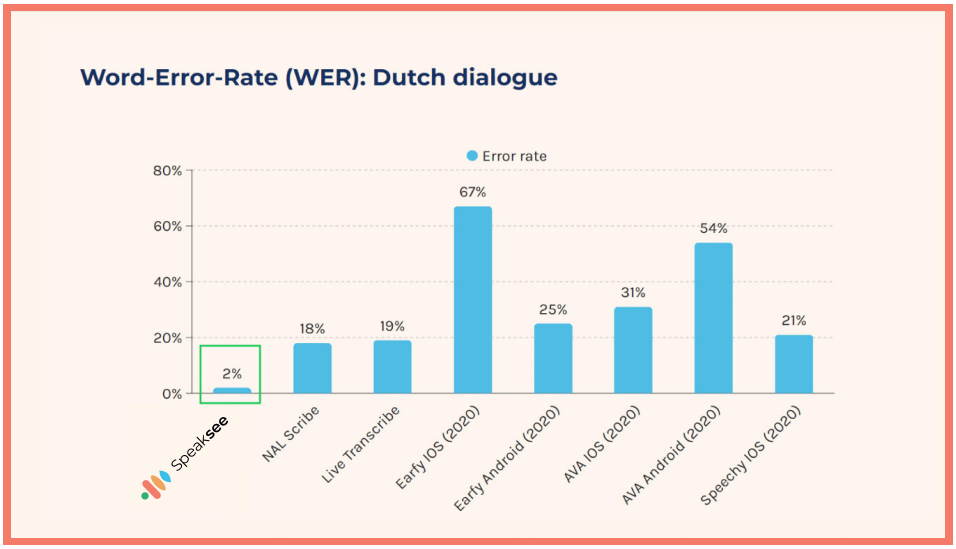

Figure 5: Word-Error-Rate: Dutch dialogue. Comparison between Speaksee and alternatives. Displayed in %.

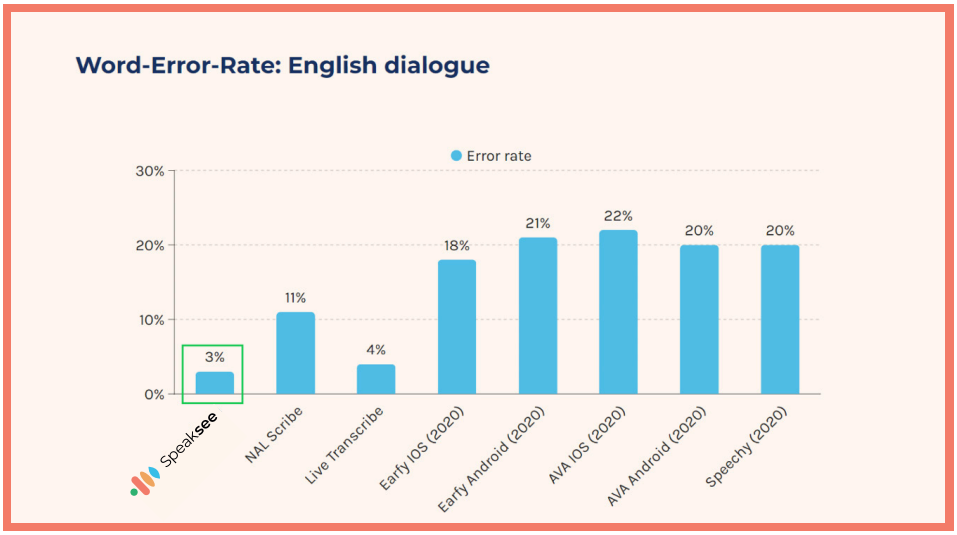

Figure 6: Word-Error-Rate: English-language dialogue. Comparison between Speaksee and alternatives. Displayed in %.

Discussion
Comparison with normal hearing people and hearing aid users

PLOMP and MATRIX test

Speaksee achieves the best scores on both speech-in-noise tests, so on the PLOMP and MATRIX test. Based on this, we can conclude that Speaksee is currently the best performing system in background noise compared to alternatives.

Normal hearers have an SRT at an SNR of -8 to -10 dB[3]. In separated noise on the PLOMP test, Speaksee achieves an SRT of -11 dB. With this result, Speaksee equals the SRT of a person with good hearing in background noise. Speaksee is the only app where speech level could be made lower than noise. The measurement in separated noise is a representative measurement of reality to measure speech in background noise, since speech and background noise do usually not come from the same direction. This indicates that hearing impaired individuals would experience very significant benefits from using Speaksee in background noise.

In frontal noise on the PLOMP test, Speaksee achieves an SRT of +1 dB. With this, Speaksee achieves a better result than hearing aid users with moderate to severe hearing loss. Kaandorp et al. (2015)[4] found an average SRT of +2dB on Dutch sentences in noise by keyword scoring for hearing aid users with moderate to severe hearing loss. For unilateral CI users, the SRT is +8 dB. Kaandorp et al. (2016)[5] found a significant difference of 1.0 dB in favour of a keyword scoring procedure over

whole sentence scoring. Alternative apps achieve an SRT of 7 dB or higher. Alternative apps score equal or worse than a unilateral CI user in background noise.

WER Dutch and English

Speaksee has the highest accuracy in transcription compared to the alternatives. This is evident from the WER for Dutch and English. Whereas Speaksee makes only 2 errors per 100 words in Dutch dialogue, is this 10 times higher for the other apps. Speaksee also makes the least errors in the English dialogue. Only Live Transcribe achieves a percentage for English that comes close to Speaksee.
Persons with severe hearing loss who use a cochlear implant achieve an error rate of about 20-40%. Hearing aid users with severe hearing loss achieve lower scores. For these groups, the use of Speaksee could likely provide significant benefits[67].

User experiences employees RadboudUMC (taken from comparative study report)

Both audiologists and rehabilitation therapists at RadboudUMC tried out the Speaksee system in the consulting room during the period from December 2022 - January 2023. The experiences are mostly positive. The equipment is easy to use, sufficiently reliable and good to use in the consulting room with hearing impaired people. This was done primarily in patients with very severe hearing loss in which less than 80% speech understanding was achieved with the assistance of hearing aids and/or cochlear implant.

For some patients there is a wow factor in the consultation room, while other patients are initially less open to new technology or prefer not to use screens. In the consultation room, the equipment works properly. The texts appear fluent and are sufficiently readable due to punctuation and formatting. In general, we see good speaker identification (correct colour) as long as everyone wears a microphone and maintains sufficient distance. The quality of the transcription is fine and in line with the WER in the dialogue measurements of this study.

Radboudumc states: "In case of problems, Speaksee's help desk is easily accessible." Radboudumc did have some small suggestions to improve the readability on a tablet screen. They also experience problems with the battery. For now, this is solved by charging the Speaksee set half an hour before use. We also have ideas to improve the use of Speaksee in hospitals. In time, we see the possibility of using Speaksee as a translation tool to communicate with non-native parents in the Amalia Children's Hospital."

Conclusion

The comparative study with Speaksee and alternatives, NAL Scribe and Google Live Transcribe, shows that Speaksee outperforms the other apps in transcribing speech in background noise. This is evident from the results of the PLOMP test and Matrix test. In the PLOMP test, Speaksee outperforms the other apps in rendering sentences-in-noise, both in frontal noise and separated noise. In separated noise, Speaksee even performs at the same level as a hearing-impaired person, and Speaksee is the only app where speech levels can be made lower than noise levels. In the Matrix test, Speaksee also outperforms the other apps in rendering correct words in noise. In this test the speech level can be made almost as low as the noise level, while this is much higher in the other apps.
The WER tests show that Speaksee is the most accurate app for transcribing both Dutch and English dialog, with only 2 errors per 100 words in Dutch dialogue and 3 errors per 100 words in English dialogue. This is significantly better than the performance of other apps where the error rate in Dutch dialogue is 10 times higher compared to Speaksee. Persons with very severe hearing loss who use a cochlear implant or hearing aid achieve an error rate of 20-40% or higher.

In conclusion, since Speaksee is the best app for conversations in background noise and has the highest accuracy in transcription, many hearing-impaired and deaf individuals could benefit from using Speaksee. With the Speaksee Microphone Kit, these individuals would be able to follow conversations in noisy environments better, such as a restaurant, and actively participate in the conversation.

Auteurs

Karlien Vanpoecke:

Karlien Vanpoecke is a product specialist and Master audiologist at Speaksee. She earned a master's degree in audiology from the Catholic University of Leuven in Belgium in 2020. In 2019, she also earned a master's degree in speech therapy at the Catholic University of Leuven.

Jari Hazelebach:

Jari Hazelebach is CEO and co-founder of Speaksee. He has deaf parents and their issues in following conversations inspired him to start Speaksee. He studied business administration at Erasmus University.

Speaksee

Speaksee is
a Dutch company based in
Rotterdam.

[1] Shukla, A., Harper, M. S., Pedersen, E., Goman, A. M., Suen, J. Y., Price, C., Applebaum, J., Hoyer, M., Lin, F. R., & Reed, N. S. (2020, 10 maart). Hearing Loss, Loneliness, and Social Isolation: A Systematic Review. Otolaryngology-Head and Neck Surgery; SAGE Publishing. https://doi.org/10.1177/0194599820910377

[2] Pragt, L., Van Hengel, P. W. J., Grob, D., Wasmann, J. A. (2022, 16 februari). Preliminary Evaluation of Automated Speech Recognition Apps for the Hearing Impaired and Deaf. Frontiers in digital health; Frontiers Media. https://doi.org/10.3389/fdgth.2022.806076

[3] Plomp, R., & Mimpen, A. M. (1979, 1 november). Speech-reception threshold for sentences as a function of age and noise level. Journal of the Acoustical Society of America; Acoustical Society of America. https://doi.org/10.1121/1.383554

[4] Kaandorp, M. W., Smits, C., Merkus, P., St, G., & Festen, J. M. (2015, 1 januari). Assessing speech recognition abilities with digits in noise in cochlear implant and hearing aid users. International Journal of Audiology; Informa. https://doi.org/10.3109/14992027.2014.945623

[5] Kaandorp, M. W., De Groot, A. M. B., Festen, J. M., Smits, C., & Goverts, S. T. (2016, 3 maart). The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. International Journal of Audiology; Informa. https://doi.org/10.3109/14992027.2015.1104735

[6] Blamey, P. J., Artières, F., Başkent, D., Bergeron, F., Beynon, A. J., Burke, E. A., Dillier, N., Dowell, R. C., Fraysse, B., Gallego, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. M., . . . Lazard, D. S. (2013, 1 januari). Factors Affecting Auditory Performance of Postlinguistically Deaf Adults Using Cochlear Implants: An Update with 2251 Patients. Audiology and Neuro-otology; Karger Publishers. https://doi.org/10.1159/000343189

[7] Flynn, M. C., Dowell, R. C., & Clark, G. M. (1998, 1 april). Aided Speech Recognition Abilities of Adults With a Severe or Severe-to-Profound Hearing Loss. Journal of Speech Language and Hearing Research; American Speech–Language–Hearing Association. https://doi.org/10.1044/jslhr.4102.285