WHITE PAPER - PRIMO 2015

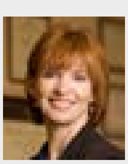
IMPROVED SPEECH UNDERSTANDING WHEN IT MATTERS MOST

ABSTRACT

Oticon's hearing technology is well known for excellent sound quality. Speech Guard E is a specialized multi-channel compression system created to preserve the natural details of the speech waveform and maximize sound quality. It has been shown that when sound is processed it has the potential to negatively affect sound quality (Arehart et al 2010). Yet, compression is essential to maintain audibility while protecting the wearer from disturbingly loud sounds.

Oticon's Speech Guard E compression system maintains near linear sound processing as much as possible. This gives better sound quality. Two feedforward, sensing systems are used simultaneously. This allows more linear attack and release time constants to be adapted to incoming sound levels within each frequency channel of the compressor. Our Speech Guard E preserves natural amplitude dynamics without sacrificing audibility or protection for loud sound.

Research in hearing impaired children at Vanderbilt University Medical Center demonstrated a 6-8% improvement in speech recognition when in noise and when simultaneously performing a physical task (Angelo et al, 2014). The University of Arizona found Speech Guard E provided an average benefit of 20% compared to syllabic compression in more complex contexts for both adults and children (Pittman et al 2014). The improvements seen for speech is postulated to be allowing the brain to hear all the details in speech. When our brains can have access to detail in speech, listening is likely to be less effortful. This contributes to BrainHearing™.



Terri E. Ives, Sc.D., Au.D.
Senior Research Audiologist
Oticon A/S

oticon
PEOPLE FIRST

BACKGROUND

Let's think about what hearing technology must accomplish. Its primary purpose should be to make speech audible, clear and understandable. A phoneme is the smallest sound unit in speech which can carry or change the meaning of a word. Hearing technology should be capable of accurately following the fast speech of a classic "fast talker" with over 20 phonemes per second, as well as the average talker at 10 phonemes per second. It should also comfortably capture the variety of life's sounds, from the softest leaf rustle to the loudest plate crashing to the floor. The level of incoming voices can also vary a lot. Phonemes vary within the speech of a single person by approximately 30 dBSPL (Boothroyd, 1993). When you add in variations between talkers, speech effort and distance, these level differences can vary by 53 dBSPL (Boothroyd, Erikson & Medwedsky 1993, 1994) or even greater (Pearson, Bennett & Fidell, 1977).

Hearing Technology must accurately capture **speech** at a rate of up to **20 phonemes per second** over a range of **53 dBSPL or more!**

Now add quick, loud sounds of kitchen plates banging, car horns blasting or a dog barking; and you begin to understand how much the sound level can vary during conversations in our dynamic world. Hearing technology has to fit all this sound comfortably into the reduced hearing range of the wearer. If hearing technology uses a linear approach to fitting the hearing range, no single volume setting would be appropriate because the difference between the highest and lowest signal levels (the 'dynamic range') is too large. This is the reason hearing technology now uses compression.

Compression removes the user's need to constantly adjust the volume, essentially doing the volume change for them. Compression reduces the large dynamic range of the world's sound levels and fits all this sound into the smaller available dynamic range of the wearer. The compressor does this by turning down (or 'compressing') uncomfortably loud sound and turning up what was inaudible so it can be now be heard. Every compressor implements these adjustments in different ways. Why? Because even though compression is good for comfortable audibility, the way it is done can have side effects. The SpeechGuard E method was created to maximize the benefits of compression while minimizing its side effects.

One of the most difficult situations for hearing technology to conquer is when what you want to hear occurs at the same time as a sudden unwanted sound. For example, you are trying to talk to your friend at a noisy playground, but a dog nearby begins barking... what happens to what she said? In this situation, everyone, regardless of hearing acuity, will experience difficulty. When you use hearing technology, advanced noise reduction strategies may reduce the ongoing playground noise occurring around you, but separating the sudden noise of the dog from the speech you want to hear from your friend is the hard part. This is where the right implementation of compression comes into the picture to help your brain separate the sounds.

In general, compression systems cannot tell if what is heard is speech or noise. The incoming sound types are all treated the same, gain is adjusted based on level. Compression then works to make sound comfortably audible within each frequency band. In this case, the compression system goes into action to diminish the loud bark and improve the sounds of speech. The speed of the compressor's response to the bark (sudden change in sound level) is called the "attack time". This happens fast so that hearing levels are kept as comfortable as possible. For comfort reasons, most major hearing technology compression systems use fast attack times (20 ms or less). Once a loud sound is gone, the device must return sound levels to normal audibility quickly without introducing distortion. The "release time" is the time it takes for the compressor to return to a normal amplification level.

Compression distorts sound. The **key** is, like medication, **limit the negative** side effects and **maximize the benefits** so you hear your best.

The smallest sounds that make a difference in speech (a phoneme) typically last as short as 20-40 ms in duration with a typical length of 100-200 ms. Fast-acting compressors use a release time ranging from 10-75 ms. Slow-acting compressors have release times greater than 200 ms. There are positive and negative side effects of compression release times. The key to compression, just like medication, is to limit the negative side effects and maximize the benefits so you hear your best.

What are the benefits and side effects of fast-acting compression?

Fast-acting compression (often called syllabic compression) acts on each phoneme. An example of where this would be helpful is a word such as "pat". The /p/ is a soft phoneme, so it needs extra amplification; the /a/ vowel phoneme is louder so it requires less amplification; and it is followed by another soft phoneme, /t/, requiring more amplification. This all happens very quickly and a fast compressor allows you to hear each individual phoneme.

Because fast-acting compression is increasing the gain of most soft sounds while decreasing the gain of the loud sounds it reduces the differences we hear between each of the phonemes. This gives speech an unnatural quality and distorts the details and cues received in natural speech. Because all soft sound, not just soft phonemes, are amplified, the compressor can sound "noisy" because the low level sounds in the background are also made audible.

What are the benefits and side effects of slow-acting compression?

When speech and other sounds are relatively stable in level, slow compression is preferred by listeners because it sounds more natural. Slow compression better preserves the natural differences in intensity between phonemes. In general, people perform better in noise conditions where the speech and noise are separated (Moore et al, 2010).

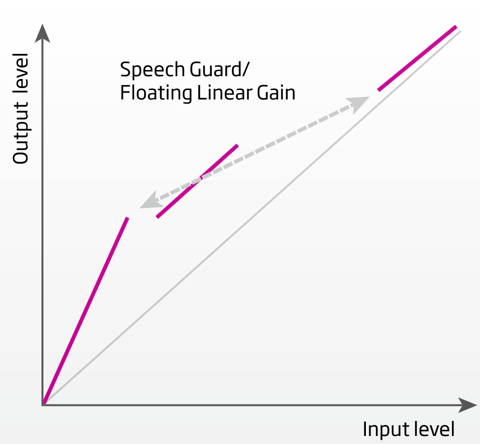
Slow-acting compression does have some possible negative side effects. It can let through transient loud sounds which may be uncomfortable, or react too slowly after a loud sound causing you to miss part of what is said.

So what type of compression is best?

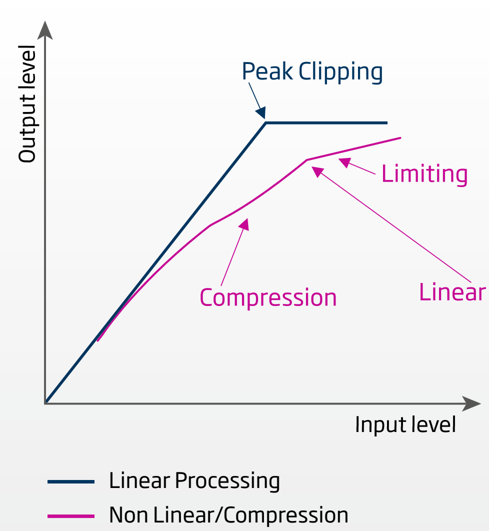
Previous research studies have not determined a leader between the various methods of compression, with no one method shown to give a consistent improvement in speech intelligibility (Gatehouse, Naylor & Eberling 2006). The research literature may be mixed on what single type of compression outperforms the others based on age or hearing loss, but a wearer's level of cognitive function appears to be one critical deciding factor (Rönnberg et al, 2003, 2008, 2011; Ng et al 2013). The more intact a person's cognitive function the faster the time constant they can effectively use for improving speech understanding. But even though they can use fast time constants effectively, will it be something they prefer to listen to? So why would I be talking about another compression method if there is no difference between them? Great question!

Satisfaction with **hearing technology** is **80.6%** -- as much as athletic shoes, iPhones and BMW's!

First, we know that sound quality significantly impacts the desire to wear hearing technology. If you like the sound of your hearing technology, you will likely wear it. Sound quality is related to 6 of the 10 most important factors for overall satisfaction with hearing technology (Kochkin 2010). According the MarkeTrak VIII survey, new hearing technology has a high satisfaction rating of 80.6%. That implies people like their hearing devices as much as athletic shoes – 80%; iPhone -average 81%; and BMW – 80% (American Consumer Satisfaction Index Reports 2014 a, b, c).



Speech Guard applies gain based on a 12 dB linear segment that slides along the typical compression function based on input levels and personalization .



Oticon devised Speech Guard E to combine the best features of both slow and fast compression to give improved audibility with high sound quality (Stone & Moore 2008; Souza 2002; Boike & Souza 2000). Speech Guard E uses fast-acting compression only when necessary and to keep sound comfortable. Otherwise, the system uses slow acting, essentially linear compression which gives the natural sound quality and pitch perception patients prefer (Stone et al 2012). Speech Guard E protects the speech envelope, which has been demonstrated to be a critical cue, for instance, when segregating sound sources (Stone & Moore 2008). This helps keep voices distinct from one another and separating them more clearly from competing background noise. The details that are imbedded in the speech signal and become critical for understanding when listening situations are difficult or complex are better preserved. Speech Guard E appears to decrease listening effort for both children and adults (Pittman 2014; Foo et al 2007; Gatehouse et al 2006; Gatehouse et al 2003).

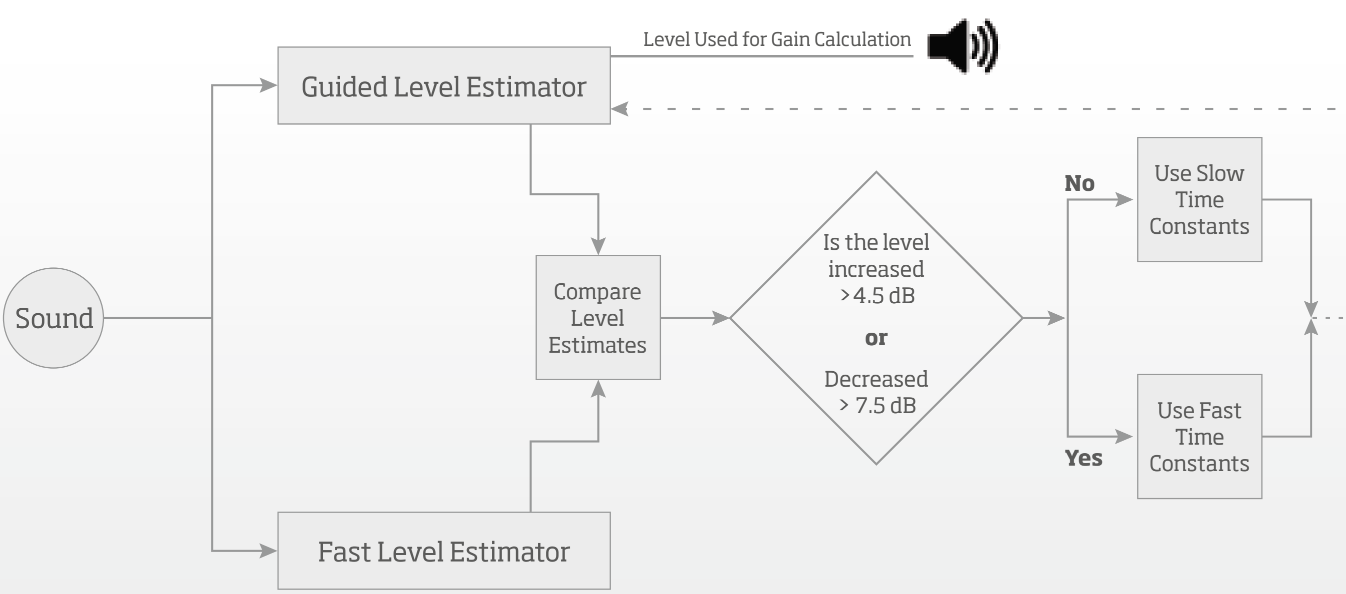
Speech Guard E significantly **improves speech** recognition when **overlapped with noise**.

The benefits of Speech Guard E are not only improved audibility and protection from loud sound, but significantly improved speech recognition when noise overlaps with speech. Therefore, our brain is helped in using its ability to recognize a sound in order to make sense of it.

How does Speech Guard E combine the best features of compression?

Speech Guard E does this by utilizing two simultaneous level sensors for monitoring the auditory scene, a guided level estimator and a fast level estimator. The guided level estimator is taking an average of the level of sound occurring over a longer fixed time period. This sensor can be thought of as being similar to the slow setting on a sound level meter. When you take sound measurements using "slow" in a dynamic sound environment, the meter moves slower up or down based on the averaged overall level. Another way to think of the guided level estimator is to imagine a large boat sitting on the ocean. The individual small waves are not felt, but the overall up and down of ocean swells are easily followed.

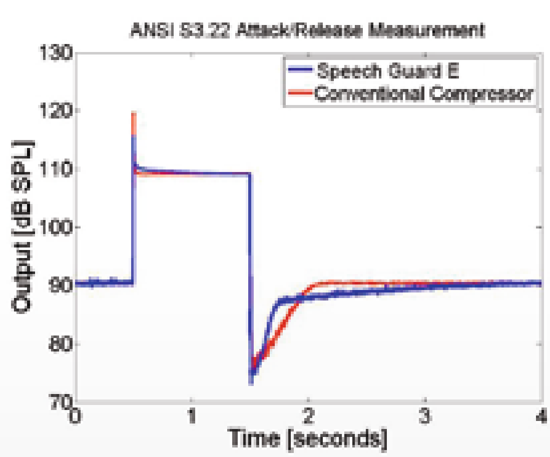
The fast level estimator is taking an average of the level of sound occurring in a short time period. This sensor acts more like the "fast" setting on a sound level meter. It reacts quickly to the dynamic environment, causing the meter to bounce wildly up and down with each tiny change in sound level. Using the boat analogy, the fast estimator is reacting to each tiny wave as if you were now in a tiny, one-person dingy on the ocean.

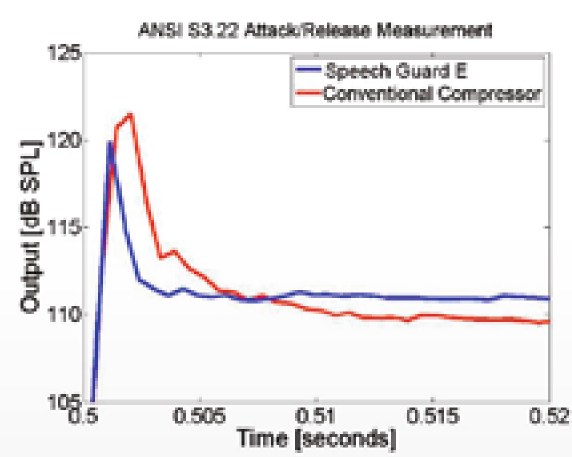


These level estimation systems work together to decide if the ongoing changes in the auditory scene are small or large abrupt changes. The hearing device is determining output of each level estimator at a rate of over 1 sample per millisecond. The output levels of the two estimation systems are also continually compared to each other. If the output level of the fast estimator has not changed upward by more than 4.5 dB or downward by more than 7.5 dB compared to the output of the guided level estimator, then the output of guided level estimator determines the time constants of the amplifier. If the comparison of the two estimator outputs exceeds this range, then the fast-estimator determines the time constants. This tolerance factor (4.5 and 7.5 dB) is being calculated in each of the frequency regions to determine the compression time constants to apply. This increases the time that the system is operating more linearly, particularly with speech. If the sound environment falls within the 12 dB variation, slow time are constants applied. The resulting gain is perceived as more constant. For a typical speech signal, the level will stay stable and only fluctuates during pauses. Therefore, slow time constants and constant gain will be applied during speech. During the pauses fast time constants will apply and the constant gain will be adapted to a new level.

In addition to the determination of the time constants for each band based on tolerance between the guided level estimator and the fast level estimator, the upward and downward slopes of the time constant are optimized to cause a linear behavior for small input signal fluctuations and a strongly non-linear behavior for larger input fluctuations. Speech Guard in the Agil product does not have this feature and uses a 9 dB window tolerance rather than 12dB as in Speech Guard E.

The time constant settings are not only adjusted based on the level estimators, but also on the personalization identity chosen. Attack times vary slightly with each identity but are in general a little faster than 10 ms. Release times cannot be measured using the IEC or ANSI standardized methods. This is because Speech Guard E operates with an initially very fast release followed by a slow return to the input steady state level. Therefore, release times are estimated based on expected perception which is in line with the point where the time constant release curves from fast to slow. Based on this estimate, release time constants can vary for the high frequency compression band from approximately 80 ms for the Energetic identity to 400 ms for the Calm identity. As the frequency of the compression band decreases the speed of the release slows with the Calm setting for the low frequency band of compression using the slowest perceived release time constant.





Attack- and release time measurements for Speech Guard E and an instrument with a conventional slow-acting compressor measured using a sinusoid at 1600 Hz as standardized in ANSI 3.22. The right graph is a magnified version of the left graph focusing on the attack time.

Back to the boat analogy, the level estimation monitor systems in Speech Guard E would allow the boat to follow the natural swells and waves of the ocean, only reacting quickly to protect you from a sudden "rogue" wave which would capsize your boat. A large increase in level is sensed by the fast sensor as being very different from the slow sensor averaged level. This tells Speech Guard E to activate fast compression, keeping the sudden event more comfortable. The same happens if there is a sudden large drop in level (which happens frequently if speech pauses). This sudden drop is detected as a large difference between the two monitor systems. Speech Guard E reacts fast to make the sound which dropped in level audible. Once the sound level is made more audible by the fast release, the compression release is then slowed to diminish the distortion of sound that can occur with fast release times. This way the compressor reacts fast enough to not lose what is said while keeping good sound quality. This is a unique feature to Speech Guard E.

Speech Guard E is set to maintain slow compression whenever the auditory scene is not changing by more than 12 dB. Thus, small but fast changes in the dynamic environment don't activate the fast compressor and introduce distortion of speech, because the fast level monitor is not significantly different from the guided level. Slow compression is maintained as much as possible at near linear levels. When the speech and noise are spatially separated this improves speech comprehension (Moore et al, 2010). When speech and a sudden loud noise overlap speech comprehension is also improved (Pittman et al. 2014). Speech Guard E lets you maintain the natural swells and lulls in conversation, experience the little nuances of what is said and still keep you comfortable and protected from large sudden changes.

Speech Guard E is a key element to Oticon's reputation for excellent sound quality. The key is a dual feedforward (looking ahead) comparative sensor design which allows hearing technology to:

- Preserves all the details of speech. This allows BrainHearing™ to occur.
- Gives comfortable listening in dynamic situations.
- Gives audibility when speech suddenly drops in level, so you can even hear soft spoken speech.
- Reduces noise when speech pauses.
- Maximizes sound quality.

Oticon Speech Guard E protects the natural sound of speech and improves speech recognition even when the acoustic environment is challenging. This BrainHearing™ feature can give our brains what we need to understand will less effort.

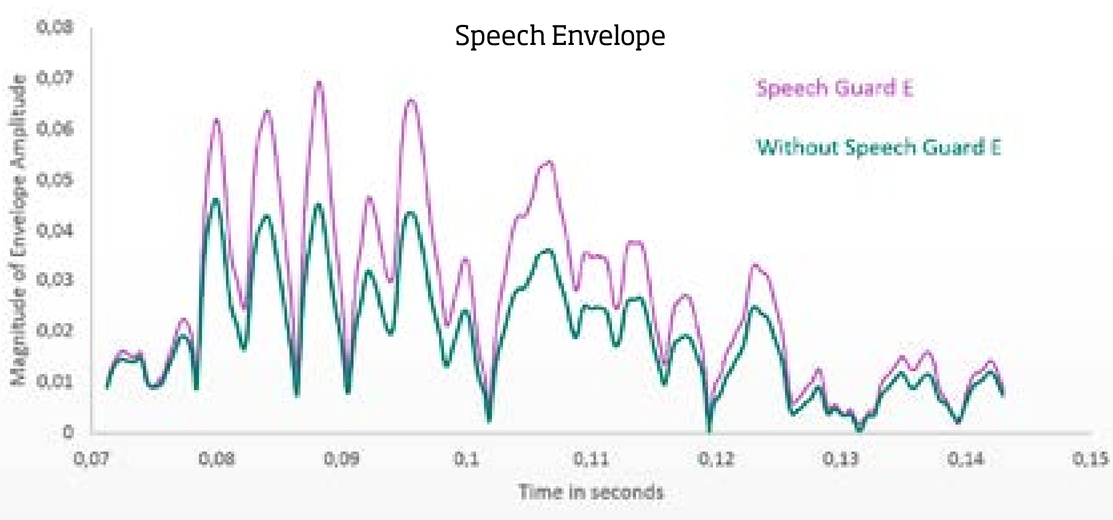


Figure 5. The details (modulations) of speech are shown with and without Speech Guard E. Figure represents the Hilbert Envelope of the sound 'æ' in 'appetizing' at the output of a Gammatone filter at 3500 Hz. The input signal is the sentence "An appetizing store in lower Manhattan" recording on a mannequin (HATS) using an Alta Pro and Nera Pro. All other processing was equalized between devices.

References

1. American Consumer Satisfaction Index (2014 a, October 21). ACSI nondurable products report 2014. Retrieved from http://www.theacsi.org/news-and-resources/customer-satisfaction-reports/reports-2014/acsi-nondurable-products-report-2014/acsi-nondurable-products-report-2014-download.
2. American Consumer Satisfaction Index (2014 b, May 20). Benchmarks by Smartphone Brand. Retrieved from http://www.theacsi.org/customer-satisfaction-benchmarks/benchmarks-by-brand/benchmarks-for-smartphones.
3. American Consumer Satisfaction Index (2014 c, August 26). Quarterly update on U.S. overall customer satisfaction-2014 results for automobiles. Retrieved from http://www.theacsi.org/news-and-resources/customer-satisfaction-reports/reports-2014/acsi-automobile-report-2014/acsi-automobile-report-2014-download.
4. Angelo, K., Angley, G. Ricketts, T., Pederson M.S., & Behrens T. (2014) Floating linear gain compression to improve speech recognition of children with moderately severe hearing loss. Poster presentation, Hearing Health Across the Lifespan, Como, Italy.
5. Arehart, K., Kates, J. Anderson, M. (2010). Effects of noise, nonlinear processing, and linear filtering on perceived speech quality. Ear Hear, 31, 420-436.
6. Boike, K.T., & Souza, P.E. (2000). Effect of compression ratio on speech recognition and speech-quality ratings with wide dynamic range compression amplification. Journal of Speech and Hearing Research, 4, 456-468.
7. Boothroyd, A. (1993). Speech perception, sensorineural hearing loss, and hearing aids. In Studebaker G and Hochberg I (Ed.), Acoustical factors affecting hearing aid performance. MA: Allyn and Bacon: Boston.
8. Boothroyd, A., Erickson, F., & Medwetsky, L. (1994). The hearing aid input: a phonemic approach to assessing the spectral distribution of speech. Ear and Hearing, 15, 432-442.
9. Foo, C., Rudner, M., Rönnberg, J., & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. J Am Acad Audiol, 18, 553-566.
10. Gatehouse, S., Naylor, G., & Elberling, C. (2003) Benefits from hearing aids in relation to the interaction between the user and the environment. Int J Audiol, 42(Suppl. 1), S77–85.
11. Gatehouse, S., Naylor, G. & Elberling, C. (2006). Linear and nonlinear hearing aid fittings: Patterns of benefit. Int J Audiol, 45, 130_152.
12. Kochkin, S. (2010) MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing. The Hearing Journal, 63 (10), 19-27.
13. Pittman, A.L., Pederson, A.J., Rash M.A. (2014) Effects of fast, slow and adaptive amplitude compression on children’s and adult’s perception of meaningful acoustic information. Journal of the American Academy of Audiology (accepted).
14. Moore, B.C., Füllgrabe, C., & Stone, M.A., (2010) Effect of special separation, extended bandwidth and compression speed on intelligibility in a speech-competing task. Journal of the Acoustical Society of America, 128(1), 360-371.
15. Ng, E.H.N., Rudner, M., Lunner, T., Pederson, M.S., & Rönnberg, J. (2013) Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. International Journal of Audiology, Early Online, 1-9.
16. Pearson, K.S., Bennett, R.L. & Fidell, S. (1977) Speech levels in various noise environments. Environmental Health Effects Research Series. EPA-600/1-77-025. U.S. Environmental Protection Agency, Washington, DC 20460, USA
17. Rönnberg, J. (2003) Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. International Journal of Audiology, 42(s1), 68-76.
18. Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). International Journal of Audiology, 47 (Suppl. 2), S99-S105.
19. Rönnberg, J., Rudner, M. & Lunner, T. (2011) Cognitive hearing science: The legacy of Stuart Gatehouse. Trends in Hearing, 15(3), 140-148.
20. Souza, P.E. (2002) Effects of compression on speech acoustics, intelligibility, and sound quality. Trends Amplification, 6, 131-165.
21. Stone, M.A. & Moore B.C. (2008) Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task. Journal of the Acoustical Society of America, 123, 1063-1076.
22. Stone, M.A., Anton, K. & Moore, B.C. (2012) Use of high-rate envelope speech cues and their perceptually relevant dynamic range for the hearing impaired, J. Acoust. Soc. Am., 132(2), 1141-51

People First

People First is our promise
to empower people
to communicate freely,
interact naturally and
participate actively

18162UK / 02.15 Printed on 100% recycled paper

oticon
PEOPLE FIRST