

IMPACT OF HEARING IMPAIRMENT ON FRICATIVE INTELLIGIBILITY FOR ARTIFICIALLY BANDWIDTH-EXTENDED TELEPHONE SPEECH IN NOISE

Patrick Bauer, Jennifer Jones, and Tim Fingscheidt

Institute for Communications Technology, Technische Universität Braunschweig
Schleinitzstr. 22, D-38106 Braunschweig, Germany
{bauer, fingscheidt}@ifn.ing.tu-bs.de

ABSTRACT

Because of its limited bandwidth, telephone speech is poorly intelligible. Artificial bandwidth extension (ABWE) reconstructs the missing frequencies aiming at, e.g., higher intelligibility. It was recently demonstrated that hearing-impaired persons wearing a hearing aid benefit from ABWE-enhanced telephone speech. However, it is unclear, whether persons without hearing impairment also take profit from ABWE in the same test conditions and if so, to what extent. This paper presents a subjective listening test with normal-hearing subjects based on meaningless German syllables simulating narrow-band (NB), ABWE-enhanced and wideband (WB) telephone speech in two noisy listening conditions. The test results reveal a clear impact of hearing impairment on the ABWE capability to improve telephone intelligibility. For a signal-to-noise ratio (SNR) of 0 dB, subjects with and without hearing impairment similarly benefit from ABWE. At 20 dB SNR, hearing-impaired subjects take even more profit in contrast to normal-hearing subjects.

Index Terms— hearing impairment, telephone intelligibility, fricatives, speech enhancement, artificial bandwidth extension

1. INTRODUCTION

The acoustic bandwidth of telephone speech is still widely limited to a narrowband (NB) frequency range of about 300 Hz to 3.4 kHz [1]. Wideband (WB) speech services ranging from 50 Hz to 7 kHz are already provided in some mobile and IP-based networks [2, 3], however, a WB call will only be established, if both conversational partners access such a network using WB-capable telephones. In a mixed case, the call is set up in the NB mode. Missing frequencies below 300 Hz are mainly responsible for a degradation of speech *quality* [4], whereas the spectral gap above 3.4 kHz leads to reduced speech *intelligibility* [5]. Thus, proper names without context often have to be spelled by means of a telephone alphabet.

Several approaches on artificial bandwidth extension (ABWE) have been proposed that aim at enhancing NB speech by estimation and reconstruction of missing frequencies [6]. Some ABWE techniques extend both spectral gaps, i.e., the upper and lower one [7, 8]. For a low-band extension, however, the pitch of the given speaker needs to be reconstructed accurately to avoid annoying artifacts [9]. Unfortunately, pitch estimation – particularly in noise – is still quite challenging [7, 10]. Due to the small dimension of their loudspeakers, mobile devices cannot sufficiently represent the lower frequencies anyway. Therefore, this paper only addresses the high-band extension based on [11], intending to basically improve *intelligibility*.

This work was funded by the German Research Foundation (DFG) under grant no. FI 1494/4-1.

Speech sounds with considerable high frequency content, such as plosives and fricatives [12, 13], are poorly intelligible over the telephone. This is not the case for vowels and sonorant consonants having their main energy contribution at low frequencies. Fricatives /s/ and /f/ can hardly be distinguished from each other on the telephone, having very low spectral content in the baseband, i.e., 0...4 kHz. ABWE algorithms tend to confuse them, which results in artifacts [14, 15]. We therefore proposed a phonetically trained ABWE [16] in order to reduce these undesired effects, leading to an improved intelligibility of meaningless English syllables in [17].

With increasing age the capability to understand NB telephone speech decreases [18, 19]. Though being individually adapted to the specific hearing impairment of their wearers, hearing aids are not capable of fully compensating for the bandwidth limitations that arise in a phone call, particularly not in noisy environments. Simply increasing the volume in order to elevate the speech level fails, because the noise level would be amplified as well [20]. Hence, it is particularly important to provide hearing-impaired persons with WB speech [21, Sec. 6.2.2]. This has been recently confirmed by a study in [22], which presents a subjective listening test based on meaningless German syllables [23, 24], employing hearing-impaired subjects that were monaurally fitted with a hearing aid [25]. It reveals a significantly improved intelligibility of ABWE-enhanced telephone speech.

The impact of both hearing impairment and signal-to-noise ratio (SNR) on intelligibility of ABWE-enhanced telephone speech, however, remains unclear. To clarify these dependencies, we performed a new subjective listening test, but this time on normal-hearing subjects. To allow comparison, it is based on the same algorithms and logatome data as in [22], simulating NB, ABWE and WB telephone speech in two noisy listening conditions.

The paper is structured as follows: Sec. 2 describes the setup of a subjective listening test on normal-hearing subjects evaluating the intelligibility of meaningless German syllables. Experimental results are presented in Sec. 3 and compared to a former study with hearing-impaired subjects. Finally, conclusions are drawn in Sec. 4.

2. EXPERIMENTAL SETUP

In order to investigate the impact of hearing impairment on the intelligibility of ABWE-enhanced telephone speech vs. both a NB and WB reference, a subjective listening test on meaningless German syllables was performed. We followed a former study with hearing-impaired persons [22], but this time subjects had normal-hearing abilities. At first, Sec. 2.1 explains the preparation of telephone speech conditions simulating an automotive listening environment. The listening test setup is then described in Sec. 2.2.

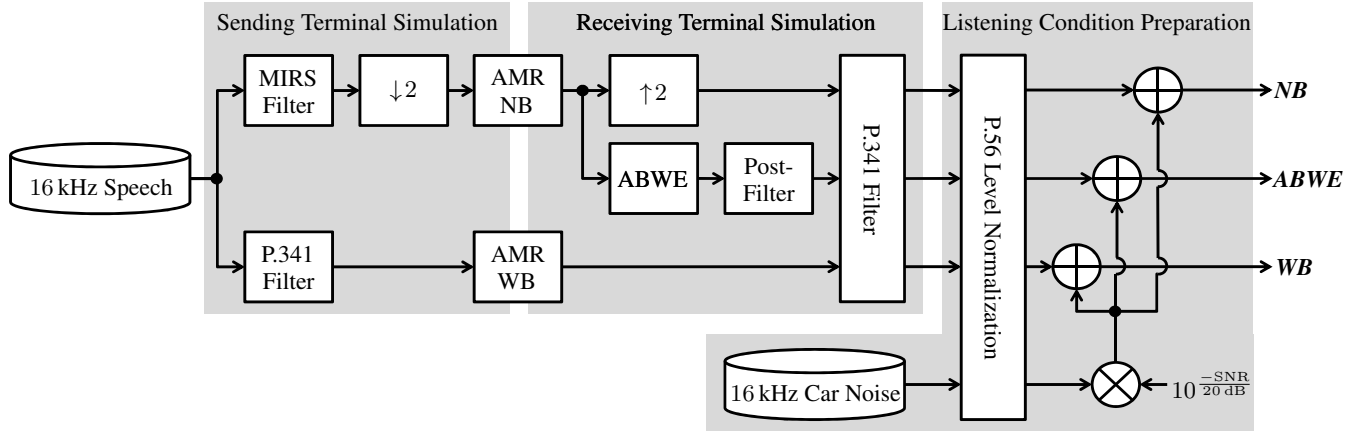


Fig. 1. Preparation of telephone speech conditions simulating an automotive listening environment at the near-end side.

2.1. Simulation of telephone speech conditions

In advance of the subjective listening test, the speech material had to be prepared adequately, as depicted in Fig. 1. It shows the pre-processing steps to prepare an artificially bandwidth-extended telephone speech condition (*ABWE*), simulating an automotive listening environment at the near-end side. The corresponding NB and WB conditions (*NB*, *WB*) thereby serve as the expected lower- and upper-bound intelligibility reference, respectively. In order to simulate the transmission characteristics of NB- and WB-capable sending terminals, MIRS [26, Annex D] and P.341 [27] filter masks were used to weight the 16 kHz sampled speech data [28], respectively.

The P.341-filtered signal was then directly transcoded via the adaptive multirate wideband (AMR-WB) speech codec at bitrate 12.65 kbps [29]. The MIRS-filtered signal was decimated to 8 kHz sampling rate, and transcoded via the adaptive multirate narrowband (AMR-NB) speech codec at bitrate 12.2 kbps [30]. On the one hand, the AMR-NB-transcoded signal was subject to ABWE processing [16]. Subsequent lowpass post-filtering reduced high-frequency whistling artifacts that may arise from ABWE [22]. On the other hand, the AMR-NB-transcoded signal was upsampled and interpolated to 16 kHz sampling rate (*NB* condition).

The characteristic of a WB-capable receiving terminal was then simulated by means of P.341 weighting for all telephone speech conditions, i.e., *NB*, *ABWE*, and *WB*. This allows for a fair comparison between the *NB* condition and the others, since an MIRS weighting of the *NB* condition instead would have unnecessarily degraded its intelligibility by characterizing the receiving terminal only with a NB capability.

The active speech level of all conditions was then equally normalized to -26 dBov [31]. Furthermore, 16 kHz sampled car noise was used to simulate an automotive environment at the near-end side (i.e., listening condition preparation). The root mean square (RMS) noise level was therefore normalized to -26 dBov [31] and then scaled by a factor to adjust a specific SNR. Finally, the scaled noise signal was individually added to the speech signal of each condition.

2.2. Subjective listening test setup

The performed subjective listening test is based on the same meaningless German vowel-consonant-vowel (VCV) syllables as used for [22]. It combined the vowels /a/, /I/ and /U/ with the unvoiced fricatives /s/, /f/, /S/, /x/ and /C/. Since /x/ and /C/ are allophones of the

same phoneme, /x/ was paired with the vowels /a/ and /U/, while /C/ was only paired with the vowel /I/ [22], as it is naturally the case in German language. The remaining fricatives were paired with all vowels. This resulted in 12 different syllables with identical vowels at the beginning and the end.

The speech samples included two male and two female voices. In order to allow for the simulation as outlined in Sec. 2.1, an initial decimation from 44.1 kHz to 16 kHz was done. Subsequently the three test conditions *NB*, *ABWE* and *WB* were generated for all data amounting to a total of 144 samples.

Each sample was processed to yield an SNR of 0 dB and 20 dB, respectively. The entire data sets of 0 dB and 20 dB SNR were separately examined in random order by 12 German subjects each. To familiarize the subjects with the test environment, a preliminary test phase of eight clean WB samples was provided beforehand. For usability purposes, the data sets of the main test were randomly divided into four subsets of 36 samples.

The subjects were asked to identify only the center consonant of each speech sample from a given selection of four answers, without differentiation between /x/ and /C/. A single repeated replay was allowed for each speech sample.

The hardware setup consisted of a laptop PC, an RME Fireface 400 external sound card and high-quality AKG K 271 MK II headphones. The audio samples were presented in a diotic manner. Adjustments of the sound level to individual comfort were permitted once during the familiarization phase.

3. EXPERIMENTAL RESULTS

Experimental results evaluated in terms of the phoneme error rate (PER) with respect to the center fricatives /s/, /f/, /S/, /x/, and /C/, as well as to the overall mean are demonstrated in Fig. 2. Standard deviations with 95 % confidence interval relating to the mean PER scores of the 12 subjects are also given. The left-hand side of Fig. 2 shows results of the new subjective listening test that was performed with normal-hearing subjects according to Sec. 2, while the right-hand side depicts results of a former subjective listening test [22]¹ on hearing-impaired subjects wearing a hearing aid. SNR conditions of

¹Please note that the results in [22, Tab. 1] were derived by taking into account the correction-for-guessing formula [32]. In this paper, we decided to abstain from this correction, keeping in mind a chance-level PER of 75 %, since four classes had to be discriminated (i.e., /s/, /f/, /S/, and /x/ or /C/).

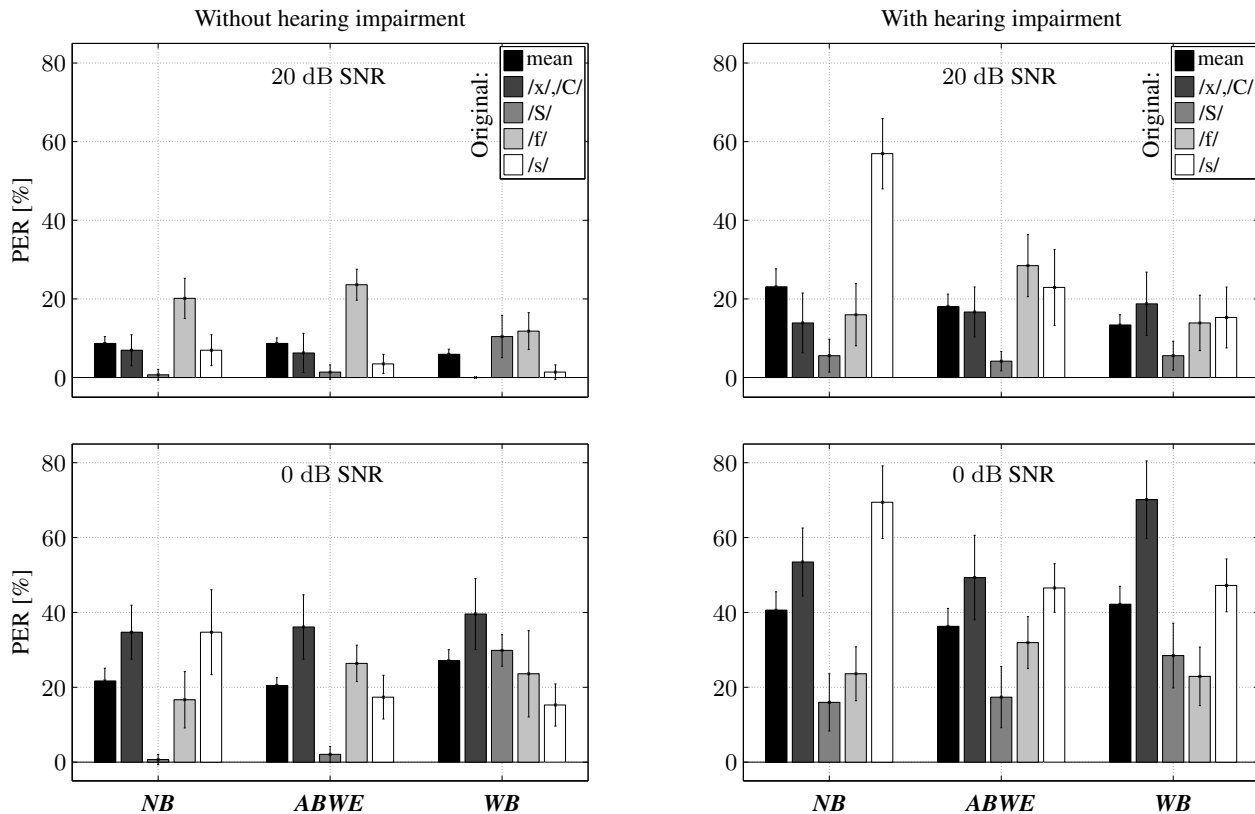


Fig. 2. Results of subjective listening tests on subjects without (left) and with (right) hearing impairment in terms of the phoneme error rate (PER) for *NB*, *ABWE*, and *WB* at 20 dB (top) and 0 dB (bottom) SNR, respectively.

20 dB and 0 dB are given on top and bottom, respectively. In general, we observe a much better intelligibility at an SNR of 20 dB as compared to 0 dB – both for normal-hearing and hearing-impaired subjects.

For normal-hearing subjects (left side) we observe that at 0 dB SNR, the *NB* condition shows a slightly lower PER than *ABWE* for the fricatives /x/, /C/ and /S/. The difference even increases to 9.7 % PER for fricative /f/. However, *NB* performs very poor on fricative /s/ with 34.7 % PER, whereas *ABWE* performs much better halving the PER to 17.4 %. In total, *ABWE* reduces the PER by 5.5 % relative to *NB*.

The test results for hearing-impaired subjects (right) in general show a much higher PER level for both SNR values. Furthermore, the benefit of *ABWE* over *NB* gets more significant. At 0 dB SNR, fricative /s/ is improved by 22.9 % absolute PER. But also /x/ and /C/ take some profit. The overall PER of *ABWE* is significantly reduced by 10.6 % relative to *NB*.

At 20 dB SNR, *ABWE* even achieves a relative PER improvement of 21.6 % for hearing-impaired subjects. Again fricative /s/ benefits most, with 34 % PER absolute below *NB*. In contrast, subjects without hearing impairment do not seem to take profit from *ABWE* for higher SNR, as shown at the top of Fig. 2 in the left graph. Obviously, the PER level is in general very low. In case of the *NB* condition, the potential to further improve the overall PER of 8.7 % is too small. Even the performance on the critical fricative /s/ is quite good, with only 6.9 % PER. Interestingly, it is much better than on fricative /f/. Note that *ABWE* indeed achieves an absolute improvement of 3.5 % PER on /s/ vs. *NB*, however, being

compensated by the degradation of /f/. For the remaining fricatives, there is almost no PER difference between *NB* and *ABWE*. A former study with normal-hearing German subjects based on English VCV syllables pointed out a PER reduction of 8.9 % for *ABWE* relative to *NB* at 20 dB SNR [17]. It involved four comparable unvoiced center fricatives and their voiced counterparts. By ignoring voiced/unvoiced errors, *ABWE* achieved a relative PER improvement of 12.5 %. In fact, normal-hearing subjects still seem to take profit from the *ABWE* condition for higher SNR values, but only if the telephone conversation is held in a foreign language. Otherwise, the recognition task is too simple and does not offer enough potential for further improvements vs. *NB*.

Focussing on the *WB* condition, all mentioned subjective listening tests consistently revealed the lowest overall PER results compared to *NB* and *ABWE* at 20 dB SNR. At 0 dB SNR, however, both graphs at the bottom of Fig. 2 surprisingly show the highest overall PER for *WB*. Obviously, the fricatives /x/, /C/ and /S/ are mainly responsible for that. This is exemplarily confirmed by the fricative confusion matrices in terms of the phoneme recognition rate (PRR) in Fig. 3 based on the experiment with normal-hearing subjects at 0 dB SNR. In contrast to *NB* and *ABWE* depicted in the top and center graphs of Fig. 3, respectively, the *WB* condition illustrated at the bottom leads to a relatively strong confusion of /S/ with the critical fricatives /s/ and /f/. Also /x/ and /C/ are often confused with them. The PRR of *WB* on /s/ and /f/, however, is still reasonably high in comparison to *NB* and *ABWE*.

This relatively poor *WB* performance originates from the (correct!) use of different sending-side filter masks – i.e., P.341 and

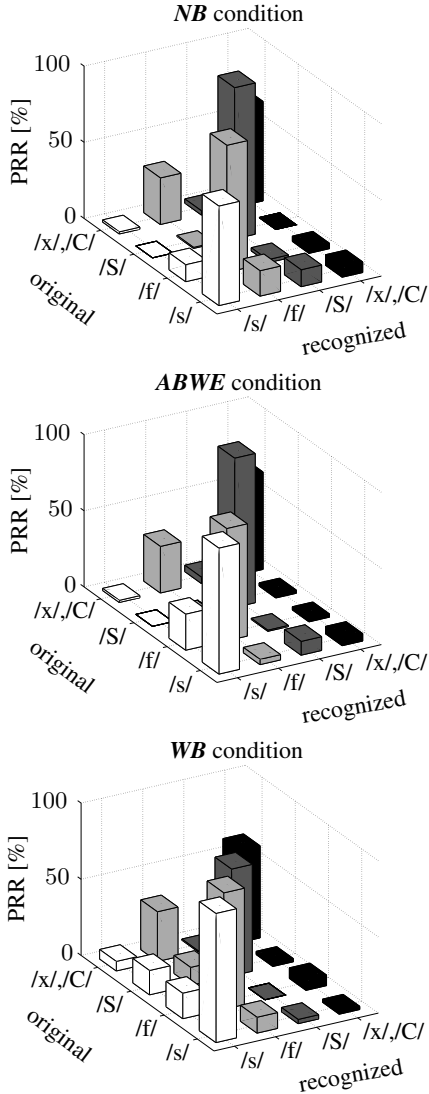


Fig. 3. Fricative confusions in terms of phoneme recognition rate (PRR) from top to bottom for **NB**, **ABWE**, and **WB**, given the experiment without hearing impairment at 0 dB SNR (Sec. 2).

MIRS – to perform a condition-specific spectral weighting in Sec. 2.1 (see Fig. 1, sending terminal simulation) *in combination with* the final P.56 level normalization that is equally applied to all conditions (see Fig. 1, listening condition preparation). To analyze this effect, we performed a simple experiment: Fig. 4 shows the long-term spectral magnitude of the concatenated VCV test speech data after having been filtered by P.341 (green, solid) and MIRS (red, dashed), respectively, at the sending side (circles). Furthermore, it shows the effect of the P.56 level normalization, with the P.341- and MIRS-filtered signals being directly normalized to an active speech level of -26 dBov (asterisks, “+ P.56”).

As expected, the MIRS weighting mask implies a strong attenuation of frequencies below 1.5 kHz and above 3.5 kHz. In between, it slightly amplifies the frequencies above 2 kHz. In contrast, the P.341 weighting mask is completely flat from about 50 Hz until 7 kHz, so it neither causes an attenuation nor an amplification in the WB frequency range. The peak of the P.341-filtered signal

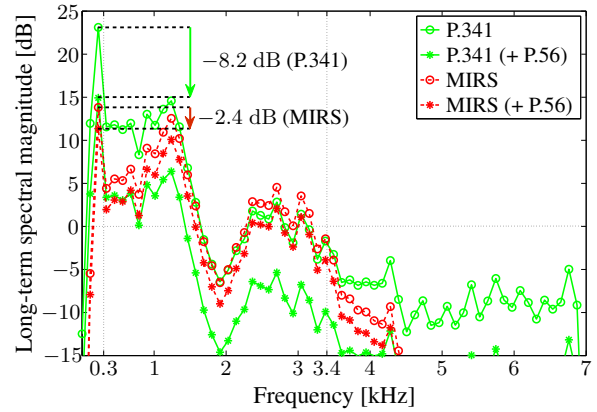


Fig. 4. Long-term spectral magnitude of VCV test speech data after sending-side P.341/MIRS weighting w/ and w/o P.56 normalization.

at 23.1 dB points out that its main energy content is located below 300 Hz. After normalization, the peak remains at 14.9 dB, i.e., the normalized P.341-filtered signal is attenuated by 8.2 dB. Due to the attenuation of the MIRS filter at low frequencies, the peak of the MIRS-filtered signal is already shrunk to 13.8 dB. Hence, the normalization only causes an attenuation of 2.4 dB, resulting in a peak at 11.4 dB. Consequently, the normalized P.341-filtered signal is 5.8 dB more attenuated, dropping significantly below the normalized MIRS-filtered signal between 1 and 4 kHz.

When finally adding the noise signal in Fig. 1, the **WB** condition, referring to the normalized P.341 curve, is therefore masked stronger than **NB** and **ABWE**, which both refer to the normalized MIRS curve. Of course, the masking effect is more relevant for 0 dB than for 20 dB SNR, which explains the higher PER of **WB** for 0 dB SNR in comparison to **NB** and **ABWE**. We further assume that this effect is predominant for fricatives /x/, /C/ and /S/, because they are more present in the respective frequency range than /f/ or /s/ [12, 13]. Anyway, the test results reveal a clear impact of both hearing impairment and listening SNR on the ABWE capability to improve telephone intelligibility: Subjects with and without hearing impairment similarly benefit from ABWE at 0 dB SNR, whereas hearing-impaired subjects take significantly more profit for 20 dB SNR.

4. CONCLUSIONS

Today, telephone speech bandwidth is still widely limited, leading to poor intelligibility particularly in noise. Being employed at the receiving side, artificial bandwidth extension (ABWE) reconstructs the missing frequencies. In order to investigate the impact of hearing impairment and listening signal-to-noise ratio (SNR) on the telephone intelligibility, subjective listening tests with normal-hearing subjects as well as hearing-impaired subjects wearing a hearing aid have been performed. They are based on meaningless German syllables that were prepared to simulate narrowband (NB), ABWE-enhanced and wideband (WB) telephone speech in two noisy listening conditions. The test results reveal for ABWE an improved fricative intelligibility at 0 dB SNR compared to the NB and WB reference. At 20 dB SNR, the benefit of ABWE even increases relative to NB speech for subjects with hearing impairment. Consequently, normal-hearing, but particularly hearing-impaired persons take a significant intelligibility profit from using ABWE during telephone conversations.

5. REFERENCES

- [1] A.H. Inglis, "Transmission Features of the New Telephone Sets," *Transactions of the American Institute of Electrical Engineers*, vol. 57, no. 10, pp. 606–612, Oct. 1938.
- [2] T. Fingscheidt, "The Silent Speech Bandwidth Revolution in Mobile Telephony," *IEEE Speech and Language Processing Technical Committee Newsletter*, Aug. 2012.
- [3] S. Ferraz de Campos Neto and K. Jarvinen, "Wideband Speech Coding Standards and Wireless Services [Guest Editorial]," *IEEE Communications Magazine*, vol. 44, no. 5, pp. 56–57, May 2006.
- [4] W. Krebber, *Sprachübertragungsqualität von Fernsprech-Handapparaten*, Ph.D. thesis, VDI Fortschrittsberichte, Reihe 10, Nr. 357, 1995.
- [5] N.R. French and J.C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," *Journal of the Acoustical Society of America*, vol. 19, no. 1, pp. 90–119, Jan. 1947.
- [6] B. Iser, W. Minker, and G. Schmidt, *Bandwidth Extension of Speech Signals*, vol. 13, Springer-Verlag US, 2008.
- [7] M.R.P. Thomas, J. Gudnason, P.A. Naylor, B. Geiser, and P. Vary, "Voice Source Estimation for Artificial Bandwidth Extension of Telephone Speech," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, U.S.A., Mar. 2010.
- [8] K. Kalgaonkar and M.A. Clements, "Sparse Probabilistic State Mapping and Its Application to Speech Bandwidth Expansion," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, Apr. 2009.
- [9] C.-F. Chan and W.-K. Hui, "Quality Enhancement of Narrowband CELP-Coded Speech via Wideband Harmonic Re-Synthesis," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, München, Germany, Apr. 1997.
- [10] C. Shahnaz, W.-P. Zhu, and M.O. Ahmad, "A Robust Pitch Estimation Algorithm in Noise," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, HI, U.S.A., Apr. 2007, vol. IV, pp. 1073–1076.
- [11] P. Jax and P. Vary, "On Artificial Bandwidth Extension of Telephone Speech," *Signal Processing*, vol. 83, no. 8, pp. 1707–1719, Aug. 2003.
- [12] G.W. Hughes and M. Halle, "Spectral Properties of Fricative Consonants," *Journal of the Acoustical Society of America*, vol. 28, no. 2, pp. 303–310, Mar. 1956.
- [13] F. Li and J.B. Allen, "Manipulation of Consonants in Natural Speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 3, pp. 496–504, Mar. 2011.
- [14] H. Pulakka, L. Laaksonen, M. Vainio, J. Pohjalainen, and P. Alku, "Evaluation of an Artificial Speech Bandwidth Extension Method in Three Languages," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 6, pp. 1124–1137, Aug. 2008.
- [15] P. Bauer, T. Fingscheidt, and M. Lieb, "Phonetic Analysis and Redesign Perspectives of Artificial Speech Bandwidth Extension," in *Proc. of Conference on Electronic Speech Signal Processing*, Frankfurt a.M., Germany, Sept. 2008.
- [16] P. Bauer and T. Fingscheidt, "A Statistical Framework for Artificial Bandwidth Extension Exploiting Speech Waveform and Phonetic Transcription," in *Proc. of European Signal Processing Conference*, Glasgow, Scotland, Aug. 2009, pp. 1839–1843.
- [17] P. Bauer, M.-A. Jung, J. Qi, and T. Fingscheidt, "On Improving Speech Intelligibility in Automotive Hands-Free Systems," in *Proc. of IEEE International Symposium on Consumer Electronics*, Braunschweig, Germany, June 2010.
- [18] A. Palva and K. Jokinen, "Presbycusis," *Acta Otolaryngologica*, vol. 70, no. 4, pp. 232–241, 1970.
- [19] A. Palva and K. Jokinen, "The Role of the Binaural Test in Filtered Speech Audiometry," *Acta Otolaryngologica*, vol. 79, no. 3-6, pp. 310–314, 1975.
- [20] M.W. Skinner and J.D. Miller, "Amplification Bandwidth and Intelligibility of Speech in Quiet and Noise for Listeners with Sensorineural Hearing Loss," *International Journal of Audiology*, vol. 22, no. 3, pp. 253–279, Jan. 1983.
- [21] H. Lazarus, C.A. Sust, R. Steckel, M. Kulka, and P. Kurtz, *Akustische Grundlagen sprachlicher Kommunikation*, Springer-Verlag, 2007.
- [22] P. Bauer, R.-L. Fischer, M. Bellanova, H. Puder, and T. Fingscheidt, "On Improving Telephone Speech Intelligibility for Hearing Impaired Persons," in *Proc. of ITG Conference on Speech Communication*, Braunschweig, Germany, Sept. 2012, pp. 275–278.
- [23] M. Bellanova, M. Serman, M. Latzel, and U. Hoppe, "Entwicklung eines Logatomtests zur Mikroskopischen Differenzierung Unterschiedlicher Hörgerätealgorithmen am Beispiel eines Kompressionsalgorithmus für Hörgeräte," in *Proc. of DGA's Annual Conference*, Frankfurt a.M., Germany, Mar. 2010.
- [24] M. Bellanova, M. Serman, and M. Latzel, "Non-adaptive Logatome Testing," German Patent # IPCOM000205581D, Mar. 2011.
- [25] M. Serman and M. Bellanova, "Method for Training Speech Recognition, and Training Device," World Patent # WO/2011/103934, Sept. 2011.
- [26] "ITU-T Recommendation P.830, Subjective Performance Assessment of Telephone-Band and Wideband Digital Codecs," ITU, Feb. 1996.
- [27] "ITU-T Recommendation P.341, Transmission Characteristics for Wideband Digital Loudspeaking and Hands-Free Telephony Terminals," ITU, Mar. 2011.
- [28] "ITU-T Recommendation G.191, Software Tool Library 2009 User's manual," ITU, Nov. 2009.
- [29] "Speech Codec Speech Processing Functions: AMR Wideband Speech Codec; Transcoding Functions (3GPP TS 26.190, Rel. 6)," 3GPP; TSG SA, Dec. 2004.
- [30] "Mandatory Speech Codec Speech Processing Functions: AMR Speech Codec; Transcoding Functions (3GPP TS 26.090, Rel. 6)," 3GPP; TSG SA, Dec. 2004.
- [31] "ITU-T Recommendation P.56, Objective Measurement of Active Speech Level," ITU, Dec. 2011.
- [32] R.B. Frary, "Formula Scoring of Multiple-Choice Tests (Correction for Guessing)," *Educational Measurement: Issues and Practice*, vol. 7, no. 2, pp. 33–38, 1988.