IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)
e-ISSN: 2278-1676,p-ISSN: 2320-3331, Volume 7, Issue 3 (Sep. - Oct. 2013), PP 48-54
www.iosrjournals.org

To Improve Speech Recognition in Noise for Cochlear Implant Users

Goutam Goyal, Dr. K.K. Dhawan, Dr. S.S. Tiwari
M.E., Biomedical Engineering Research Group , Singhania University Rajasthan , India,
Ph.D. Director, Shekhawati Group Of College Shekhawati, Rajasthan,India,
Ph.D. , Managing Director , Sensors Technology Pvt. Ltd. , Gwalior , India

Abstract: *The hypothesis that when listening to speech in fluctuating maskers, CI users can not fuse the pieces of the message over temporal gaps because they are not able to perceive reliably the acoustic landmarks introduced by obstruent consonants (e.g., stops). To test this hypothesis, CI users were first presented with sentences containing clean obstruent segments, but corrupted sonorant segments (e.g., vowels). The other experiment investigated the hypothesis that envelope compression smears acoustic landmarks which signify syllable/word boundaries. To test this, CI users were presented with noise-corrupted stimuli processed using logarithmic compression during voiced segments and a weakly-compressive mapping function during unvoiced segments. All patients were profound-totally deaf, adults with a post lingual onset of impairment. The data support the efficacy of a feature extraction coding system where specific formant and amplitude information are transmitted via direct electrical stimulation to the cochlea. To examine the hypothesis that the newer generations of cochlear implants could provide considerable speech understanding to late-implanted, prelingually deaf adult patients.*

I. The Clinical Program

1.1Experiment 1: Masking Releases for Cochlear Implant Users

Methods Subjects A total of seven postlingually deafened Clarion CII implant users participated in this experiment. All subjects had at least 3 years of experience with their implant devices. Most subjects visited our lab two times. The biographical data for each subject are given in Table 1.1. **Stimuli**

The speech material consisted of sentences taken from the IEEE database . All sentences were produced by a male speaker. The sentences were recorded in a sound-proof booth in our lab at a 25 kHz sampling rate. Two types of maskers were used. The first was speech-shaped noise, which is continuous (steady-state) and had the same long-term spectrum as the test sentences in the IEEE corpus.

Table 1.1. Biographical data of the CI users

Subject	Gender	Age (yr)	Duration of deafness prior to implantation(yr)	CI use (yr)	Number of active electrodes	Stimulation rate (pulses/s)	Etiology
S1	Female	60	2	4	15	2841	Medication
S2	Male	42	2	4	15	1420	Hydrops/ Menier's syndrome
S3	Female	47	>10	5	16	2841	Unknown
S4	Male	70	3	5	16	2841	Unknown
S5	Female	62	<1	4	16	1420	Medication
S6	Female	53	2	4	16	2841	Unknown
S7	Female	40	5	8	14	1420	Genetic

The second masker was a two-talker competing speech (female) recorded in our lab. Two long sentences, produced by a female talker, were used from the IEEE database. This was done to ensure that the target signal was always shorter (in duration) than the masker. The IEEE sentences were manually segmented into two broad phonetic classes:

(a) the obstruent consonants which included the stops, fricatives and affricates, and
(b) The sonorant sounds which included vowels, semivowels and nasals. More detailed description can be found in [12].

(C) Signal Processing Signals were first processed through a pre-emphasis filter (2000 Hz cutoff), with a 3 dB/octave roll off, and then band pass filtered into 16 channels (according to subjects' condition) using sixth-order Butterworth filters. Logarithmic filter spacing was used to allocate the channels across a 300-5500 Hz bandwidth. The envelope of the signal was extracted by full-wave rectification and low-pass filtering (second-order Butterworth) with a 400 Hz cutoff frequency. The envelopes in each channel were log compressed to the subject's electrical dynamic range. The speech stimuli were generated using the above algorithm in two different conditions. In the first control condition, the corrupted speech stimuli were left unaltered. That is, the obstruent consonants remained corrupted by the maskers (baseline). In the second

condition, the speech stimuli contained clean obstruent segments but corrupted sonorant segments (unv clean).

(D) Produce The above stimuli were generated off-line in MATLAB and presented directly (via the auxiliary input jack) to CI users via the Clarion Research Interface platform. Prior to the test, subjects listened to some sentences to become familiar with the processed stimuli. The training session lasted for about 20-30 minutes. During the test, the subjects were asked to write down the words they heard. Subjects participated in a total of 8 conditions (= 2 SNR levels × 2 algorithms × 2 maskers). Two lists of IEEE sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions. Sentences were presented to the listeners in blocks, with 20 sentences/block for each condition. The different conditions were run in random order for each listener.

II. Results and Discussion

The mean scores for all conditions are shown in Figure 1.1 Performance was measured in terms of the percentage of words identified correctly (all words were scored). Two- way ANOVA with repeated measures was used to assess the effect of masker type. The control noisy stimuli (shown in Figure 5.1 with blue bars) showed no significant effect of masker type ($F[1, 6] = 0.036$, $p = 0.89$). No significant interaction ($F[1, 6] = 2.1$, $p = 0.176$) was found between SNR level and masker type. Performance with steady noise was comparable to that attained by the two-talker masker, consistent with findings reported in other cochlear implant studies

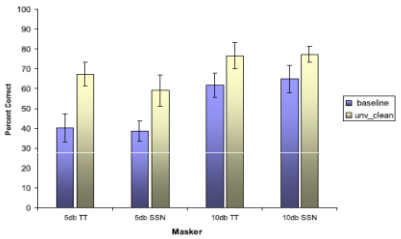


Figure 1.1 Mean speech recognition scores as a function of SNR level for the var- ious masker (TT=two-talker and SSN=steady noise) and channel conditions. The blue bars denote scores obtained with stimuli containing clean obstruent consonants, and the yellow bars denote scores obtained with the control stimuli containing corrupted obstruent consonants. The performance obtained in quiet (Q) is also shown for comparative purposes. Error bars denote standard errors of the mean.

A different pattern in performance emerged in the conditions in which the obstruent consonants were clean and the remaining sonorant sounds were left corrupted shown in Figure 1.1 with yellow bars). Performance obtained with the two-talker masker was higher than performance obtained with the steady noise masker, at least at 5 dB SNR. Two-way ANOVA showed a significant effect ($F[1, 6] = 21.5$, $p = 0.005$) of SNR level, a non-significant effect ($F[1, 6] = 2.8$, $p = 0.167$) of masker type and a significant interaction ($F[1, 6] = 9.61$, $p = 0.03$). The interaction between SNR level and masker type was due to the fact that a larger improvement was observed for the low SNR level (5 dB) condition compared to the higher SNR (10 dB) condition. More specifically, performance improved (relative to the corrupted stimuli) by roughly 20-30 percentage points at 5 dB SNR, and by 10-15 percentage points at 10 dB SNR. Post-hoc tests revealed that performance in the two-talker masker conditions (with clean obstruent consonants) was significantly higher ($p = 0.04$) than the corresponding performance in SSN conditions at 5-dB SNR, but the difference was not statistically significant ($p = 0.81$) at 10 dB SNR. This outcome suggests that the CI users received masking release, at least in the low-SNR condition, when they had access to the clean obstruent consonants.

Introducing the clean obstruent consonants in the corrupted vocoded stimuli produced a substantial improvement in performance in both SNR conditions (Figure 1.1). The magnitude of the improvement obtained by the CI users when they had access to information provided by the clean obstruent consonants seemed to depend on the SNR level. At 5 dB SNR, the improvement ranged from a low of 20 percentage points (in the SSN masker condition) to nearly 30 percentage points (in the two-talker masker condition). The improvement was smaller at 10 dB SNR and ranged from 12-15 percentage points. This SNR dependency is probably due to the different set of acoustic cues (and reliability of those cues) available to the listeners when presented with spectrally degraded speech. The fact that masking release was observed only at low SNR levels is consistent with the outcomes of prior studies with normal-hearing listeners . In our study, masking release was assessed in conditions wherein spectral information was degraded and speech redundancy was reduced. In our study, when CI users were presented with a severely degraded (spectrally) signal at low SNR levels and were provided with access to the obstruent consonants and associated landmarks, they were able to exploit the dips in the fluctuating masker, much like normal-hearing listeners are able to do so. At higher SNR levels, however, there is little room

for confusion between the target and masker, and CI users likely utilize other cues available to them. Consequently, they rely less on exploiting the masker dips in the envelopes.

There are several underlying mechanisms responsible for the above improvement in performance . For one, listeners had access to multiple spectral/temporal cues when the clean obstruent consonants were introduced, although the saliency of ,those cues was highly dependent on the SNR level. Additionally, CI users had better access to F1/F2 transitions to/from the vowel and sonorant sounds, more accurate voicing information and consequently better access to acoustic landmarks which per- haps aided the listeners in identifying more easily word boundaries in the noisy speech stream.

It was not clear from the above discussion as to why CI users did not receive release of masking when presented with corrupted (unprocessed) stimuli. One possible reason is that they were not able to perceive clear acoustic landmark information such as the presence of vowel/consonant boundaries, particularly at the low SNR levels. We hypothesize that the envelope compression, which is commonly implemented in CI devices, is responsible for that, as it smears the acoustic landmarks present in the signal, particularly at low SNR levels. This hypothesis is tested next.

1.2 Experiment 2: Impact of Envelope Compression on Speech Recognition in Noise

In this experiment, we test the hypothesis that envelope compression (which is of- ten logarithmic) smears the acoustic landmarks present in the signal and limits the ability of CI users to receive masking release. The smearing is caused by the fact that the use of log envelope compression tends to amplify the low-energy weak consonants (e.g., fricatives and stops), thus distorting the inherent vowel-to-consonant energy ratio in natural speech. To test the above hypothesis, we present to CI users noise-corrupted sentences processed via an algorithm that compresses the envelopes using a logarithmic acoustic-to-electric mapping function during voiced segments (e.g., vowels) and a weakly-compressive mapping function during unvoiced segments (e.g., stops). The underlying motivation for the use of the weakly-compressive mapping function applied during unvoiced segments is to maintain a more natural vowel-to-consonant ratio, which in turn will make the acoustic landmarks more evident.

Methods

A. Subjects and Stimuli

The same seven subjects who participated in Experiment 1 also participated in this experiment. The same speech materials (IEEE sentences) were used as in Experiment 1. Two types of maskers were used for Experiment 2. The first was speech-shaped noise. The second masker was a 20-talker babble (Auditec CD, St. Louis, MO). The average long-term spectrum of the multitalker babble is shown in .

B. Signal Processing

The signal processing strategy used by the CI users is the same as in Experiment 1. The main difference lies in the use of two different acoustic-to-electric mappings, which are applied to the corrupted envelopes depending on the phonetic segment present in the sentences. For voiced segments (e.g., vowels) a logarithmic mapping is used (same as used in the CI user's daily strategy), while for unvoiced segments a less compressive mapping function is utilized. The acoustic-to-electric mapping is implemented as follows: Y (n) = A · [X(n)]q + B (5.1) where Y (n) indicates the electrical amplitude output (measured in clinical units or micrometers) at time n, X(n) denotes the acoustic envelope amplitude, and the con stants A and B are used to ensure that the acoustic amplitudes are mapped within the electrical dynamic range.

The power exponent q is used to control the steepness of the compression function. And the value of q = −0.0001 was used for log compression and the value of q = 0.33 was used for weak (less) compression.

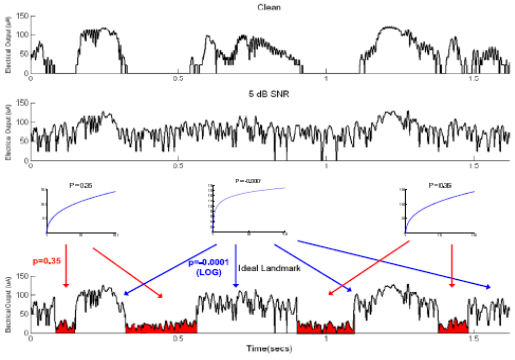


Figure 1.2 Envelope (4^{th} channel with center frequency of 600 Hz) of clean (upper panel) signal, envelope in 5 dB SSN (middle panel) with log compression function and envelope processed (bottom panel) with varied compression steepness (q = −0.0001 for unvoiced segment and q = 0.35 for voiced segments).

The speech stimuli are processed in two different conditions. In the first control condition, the corrupted speech stimuli are processed using the log compression, as used in their daily strategy. In the second condition, the corrupted speech envelopes are compressed using a logarithmic-shaped function (q = −0.0001) during voiced segments (e.g., vowels) and a less-compressive mapping function (q = 0.33) during unvoiced segments (e.g., stops). The weakly compressive function is only applied to the low frequency channels, and more precisely, the seven most apical channels spanning the bandwidth of 200-1000 Hz. The remaining nine higher-frequency channels are processed using the log-mapping function. For subjects with only 14-15 active electrodes (see Table 1.1), the remaining 7-8 higher-frequency channels are processed using the log mapping function. The motivation for applying a different mapping function in the low frequencies is to make the low-frequency phonetic boundaries more evident without suppressing the high-frequency cues present in most obstruent consonants (e.g., /t/). The details of the system can be found in Figure 1.2. It should be pointed out that the selective compression is applied to the corrupted speech envelopes in both voiced and unvoiced segments of the sentences. That is, unlike the conditions in Experiment 1, in this experiment both voiced and unvoiced segments in the sentence materials remained noise corrupted. Similar to Experiment 1, it is assumed that we have access to the true voiced/unvoiced acoustic boundaries.

Figure 1.2 shows as an example a noise-corrupted envelope (with center frequency = 600 Hz) processed using selective compression. The sentence was corrupted in SSN at 5 dB SNR. By comparing the bottom two panels, it is clear that the use of selective compression renders the consonant boundaries more evident and perhaps perceptually more salient. The effect of applying a weakly compressive mapping function to the low-frequency region of the unvoiced segments is evident in the bottom

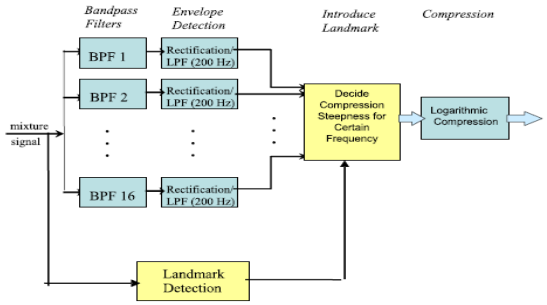


Figure 1.3 System diagram of introducing the landmark cues to the CI system panel. As can be seen from this panel, the envelopes are attenuated relative to the envelopes in the voiced segments, thereby rendering the vowel/consonant boundaries (present in the low frequencies, i.e., < 1000 Hz) more clear.

Procedure

The above stimuli were generated off-line in MATLAB presented directly (via the auxiliary input jack) to CI users via the Clarion Research Interface platform. Prior to the test, training session was applied to the subjects as experiment 1. Subjects participated in a total of 16 conditions (= 2 SNR levels × 3 algorithms × 2 maskers 2 SNR levels × 2 algorithms × 1 masker). Two lists of IEEE sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions. Sentences were presented to the listeners in blocks, with 20sentences/block for each condition. The different conditions were run in random order for each listener.

Results and Discussion

The mean scores for all conditions are shown in Figure 1.3 Performance was measured in terms of the percentage of words identified correctly (all words were scored). Two- way ANOVA, with repeated measures, was used to assess the effect of masker type. ANOVA indicated no significant ($F[1, 6] = 0.103$, $p = 0.782$) effect of masker type and no significant interaction ($F[1, 6] = 1.25$, $p = 0.329$) between SNR level and masker type. This suggests that the selective compression did not provide masking release Additional analysis was conducted to assess whether the use of selective compression improved performance relative to that obtained with the un-processed stimuli, which were processed (for all phonetic segments) using the log-mapping function. Two-way ANOVA, run on the scores obtained in the SSN conditions, indicated significant effect ($F[1, 6] = 161.7$, $p < 0.0005$) of SNR level, significant effect ($F[1, 6] = 36.8$, $p = 0.001$) of the method of compression and non-significant interaction ($F[1, 6] = 1.2$, $p = 0.307$). Similarly, two-way ANOVA, run on the scores obtained in the

two-talker masker conditions, indicated significant effect ($F[1, 6] = 10.24$, $p = 0.024$) of SNR level, significant effect ($F[1, 6] = 23.4$, $p = 0.005$) of the method of compression and non-significant interaction ($F[1, 6] = 1.8$, $p = 0.235$).The above analysis clearly indicates that the use of selective compression can improve significantly speech intelligibility at both SNR levels, relative to the baseline condition.

As shown in Figure 1.3, selective compression improved performance by nearly 15 percentage points at 5 dB SNR and by approximately 10-15 percentage points at 10 dB SNR. The improvement in performance was found to be consistent for both types of maskers. As mentioned earlier, both voiced and unvoiced segments in the sentences were left noise corrupted. Yet, a consistent, and statistically significant, improvement was noted when selective compression was applied to the low-frequency channels. Results indicated that access to the low-frequency (0-1000 Hz) region of the clean obstruent-consonant spectra was sufficient to realize significant improvements (about 10-15 percentage points) in performance and that was attributed to improvement in transmission of voicing information. Hence, we deduce that by applying selective compression in the low frequencies (as done in the experiment) we can improve significantly the transmission of voicing information.

The improvement obtained with selective envelope compression was not as large as that obtained in Experiment 1 (about 20-30 percentage points) when the listeners had access to the clean obstruent consonant spectra. No masking release was found either in the experiment. We attribute this to the following reasons. First, un- like the conditions tested in Experiment 1, the obstruent consonants in Experiment 2 were either left corrupted or were suppressed (at least in the low frequencies). Consequently, listeners did not have access to clean obstruent consonant information. That, in turn, impaired their ability to fuse (or ”glimpse”) pieces of the target signal across temporal gaps (e.g., [9][12]). Second, the low-frequency envelope suppression was done without taking into account the spectral content or spectral energy distribution of the obstruent consonants at hand. The labial stop consonants (e.g., /p/, /b/), for instance, are characterized by low-frequency energy concentration; hence suppressing the low-frequency region might introduce conflicting (burst) cues to the listeners.

The presence of conflicting burst cues will in turn force listeners to rely on formant transitions , which we know that CI listeners are not able to perceive reliably . On the other hand, the alveolar stop consonants and fricatives (e.g., /t/, 86 /s/) have high-frequency energy concentration and the applied envelope suppression would be more appropriate and more likely to be beneficial. In all, selective envelope compression can not reach its full potential (as demonstrated in Experiment 1) given that it is applied to all obstruent consonants.

The impact of envelope compression on speech recognition (in quiet and in noise) was also examined in other studies . However, as the SNR level decreased, the effect of nonlinear mapping became dramatic and asymmetric: performance with weakly compressive mappings declined mildly in noise, but performance declined dramatically in noise with a strongly compressive amplitude mapping. This outcome is partially consistent with the findings of the experiment. Performance with the strongly compressive mapping was significantly worse (see Figure2.3) than performance with the (selective) weak mapping ($q = 0.35$). Hence, in agreement with prior studies , we can conclude that the use of a strongly compressive mapping function that is applied to all phonetic segments is not appropriate, or beneficial, for CI users when listening to speech in noisy environments.

A selective compression function was proposed in this experiment for enhancing access to the acoustic landmarks in noisy conditions. An alternative approach was proposed in based on the use of s-shaped input-output functions which are expansive for low input levels, up to a knee point level, and compressive thereafter.

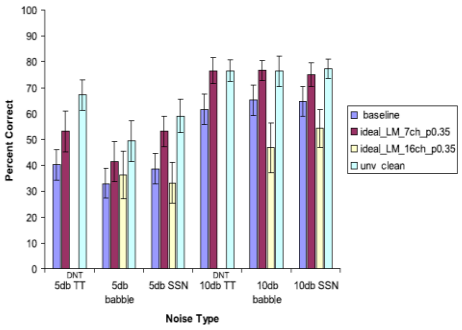
The knee points of the s-shaped functions changed dynamically and were set proportional to the estimated noise floor level. For the most part, the expansive (i.e., less compressive) part of the s-shaped functions operated on obstruent segments, which generally have lower intensity and energy compared to that of sonorant segments. The main advantage of using s-shaped functions for mapping the acoustic signal to electrical output is that these functions do not require landmark detection algorithms as they are applied to all phonetic segments. Replacing the conventional log mapping functions with the s-shaped functions yielded significant improvements in speech intelligibility in noise by nine cochlear implant users .

III. Conclusion

Combining the findings for Experiments 1 and 2, we can reach the conclusion that in order for CI users to receive masking release it is necessary for them to perceive reliably the low-energy weak consonants (e.g., stops). The perception of weak consonants can be facilitated or mediated, at least to some extent, by reliable detection of the vowel/consonant boundaries. As demonstrated in the experiment, providing access to the vowel/consonant boundaries to CI users produced significant improvement in performance, but that was not sufficient to observe masking release owing to the fact that the weak consonants were noise corrupted. Put differently, the use of selective compression enabled the listeners to identify the location (the where about) of the weak consonants in the corrupted speech stream, but did not allow the listeners to perceive reliably their identity (the what). Further (selective) enhancement, perhaps by a noise-reduction algorithm, of the weak

consonants might be needed to obtain both (location and identity of weak consonants), and subsequently observe masking release. In a realistic scenario, one can envision an algorithm that first detects the vowel/consonant boundaries and then removes or suppresses the noise from the corrupted weak consonants. Such an algorithm would require automatic detection of vowel/consonant boundaries in noise, and the performance of such an algorithm is examined next.

The study examined the longstanding question as to why CI users are not able to receive masking release. The hypothesis posed and tested was that CI users are not able to receive masking release because they are not able to fuse the Pieces of the message over temporal gaps as they are unable to perceive reliably the acoustic landmarks introduced by obstruent consonants (e.g., stops). This hypothesis was tested in Experiment 1 by presenting to listeners stimuli that contained corrupted sonorant segments (e.g., vowels) and clean obstruent consonants (e.g., stops).



[36]

Figure 2.1 Mean speech recognition scores as a function of SNR level for the various maskers (TT=two-talker, SSN=steady noise and babble noise) when introducing the ideal landmark cues. The blue bars denote scores obtained with the control stimuli containing corrupted obstruent consonants (baseline), the red bars denote scores obtained from the condition ideal LM 7ch p0.35, the yellow bars denote scores obtained from the condition ideal LM 16ch p0.35 and the green bars denote score from condition unclean. Error bars denote standard errors of the mean.

Results indicated substantial improvement (20-30 percentage points) in intelligibility, particularly at low SNR levels (5 dB). Performance in the 2-talker masker conditions (5 dB SNR) was found to be significantly higher than performance in the SSN conditions, thus demonstrating that CI users can receive masking release. Experiment 2 focused on answering the question: What is the contributing factor(s) for the absence of masking release observed in cochlear implants? Envelope compression was posited to be one of the contributing factors, and this was based on the hypothesis that applying envelope compression in noisy environments tends to amplify the weak consonants and smear the acoustic landmarks, such as those signaled by the spectral discontinuities associated with the onsets/offsets of consonant releases. This hypothesis was tested by using selective envelope compression wherein a weakly compressive function was applied during the obstruent consonants (in the low frequencies, within 1 kHz) and a relatively strong (log) compressive function was applied during sonorant segments. This had the effect of suppressing the envelopes of the obstruent consonants in the low frequencies, thereby making the vowel/consonant boundaries more evident. Results revealed a significant improvement in intelligibility when selective envelope compression was used, but no evidence of masking release. This was attributed to the fact that the CI users had a clear access to the vowel/consonant boundaries, but per- haps perceived conflicting spectral information since the high-frequency region (> 1 kHz) of the obstruent consonants was left corrupted. The significant improvement in performance obtained with selective compression applied to the low frequencies, was attributed to the better transmission of voicing information.

Considering together the outcomes from Experiments 1 and 2, we can conclude that in order for CI users to receive masking release, it is necessary for them to perceive reliably not only the presence and location of the vowel/consonant boundaries (as tested in Experiment 2) but also the information contained in the low-energy obstruent consonants. Put simply, CI users are not able to receive masking release because they do not perceive reliably the obstruent consonants in noise. These consonants have low energy and they are easily masked in noise, more so than vowels . The situation is further exacerbated by the fact that a rather strongly compressive mapping is typically used in cochlear implants, which in turn smears the vowel/consonant boundaries (thus making it difficult to detect the presence/absence of consonants) and amplifies the already noise-masked weak consonants. The use of selective envelope compression (as done in Experiment 2) seems to be more appropriate for processing speech in noise.

Bibliography

[1] J. G. Bernstein and K. W. Grant. Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners.J. Acoust. Soc. Am., 125:3358–3372, 2009.

[2] M. Dorman and P. Loizou. Relative spectral change and formant transitions as cues to labial and alveolar place of articulation. J. Acoust. Soc. Am., 100:3825–3830, 1996.

[3] M. Dorman, M. Studdert-Kennedy, and L. Ralphael. Stop consonant recognition: Release bursts and formant transitions as functionally equivalent context-dependent cues. Percept. Psychophys., 22:109–122, 1977.

[4] L. M. Friesen, R. V. Shannon, D. Baskent, and X. Wang. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. J. Acoust. Soc. Am., 110:1150–1163, 2001.

[5] Q. J. Fu and R. V. Shannon.Effects of amplitude nonlinearity on speech recognition by cochlear implant users and normal-hearing listeners. J. Acoust. Soc.Am., 104:2571–2577, 1998.

[6] Q.J. Fu and R. V Shannon. Effect of acoustic dynamic range on phoneme recognition in quiet and noise by cochlear implant users.Percept.Psychophys. 06:L65–L70, 1999.

[7] K. Kasturi and P. Loizou. Use of s-shaped input-output functions for noise suppression in cochlear implants. Ear Hear., 28(3):402–411, 2007.

[9] N. Li and P. Loizou. The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise. J. Acoust. Soc. Am., 124(6):3947–3958, 2008.

[10] N. Li and P. Loizou. Effect of spectral resolution on the intelligibility of ideal binary masked speech. J. Acoust. Soc. Am., 123(4):59–64, 2008.

[11] N. Li and P. Loizou. Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. J. Acoust. Soc. Am., 123(3):1673–1682,2008.

[12] N. Li and P. Loizou. Factors affecting masking release in cochlear implant vocoded speech. J. Acoust. Soc. Am., 126:338–348, 2009.

[13] B. Munson and P.B. Nelson. Phonetic identification in quiet and in noise by listeners with cochlear implants. J. Acoust. Soc. Am., 118(4):2607–2617, 2005.

[14] P. Nelson and S. Jin. Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am.,115(5):2286–2294, 2004.

[15] P. Nelson, S. Jin, A. Carney, and D. Nelson. Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. J. Acoust.Soc. Am., 113(2):961–968, 2003.

[16] M. Nilsson, S. Soli, and J. Sullivan. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am., 95(2):1085–1099, 1994.

[17] M.Owren and G. Cardillo. The relative role of vowels and consonants in discriminating talker identity versus word meaning. J. Acoust. Soc. Am., 119(3):1727–1739, 2006.131

[18] A. J. Oxenham and A. M. Simonson. Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. J. Acoust.Soc. Am., 125:457–468, 2009, 2009.

[19] P.Paalanen, J. K. Kamarainen, J. Ilonen, and H. Kalviainen. Feature representation and discrimination based on gaussian mixture model probability densities, practices and algorithms. Pattern Recognition, 39:1346–1358, 2006.

[20] G. Parikh and P. Loizou. The influence of noise on vowel and consonant cues.J. Acoust. Soc. Am., 118(6):3874–3888, 2005.

[21] S. Phatak and J. Allen. Consonants and vowel confusions in speech-weighted noise. J. Acoust. Soc. Am., 121(4):2312–2326, 2007.

[22] M. Qin and A. Oxenham. Effects of envelope-vocoder processing on f0 discrimination and concurrent-vowel identification. Ear Hear., 26:451–460, 2005.

[23] N. Radford and H. Geoffrey. A view of the EM algorithm that justifies incremental, sparse, and other variants. Cambridge, MA: MIT Press., 1999.

[24] N. Roman and D. Wang. Pitch-based monaural segregation of reverberant speech. J. Acoust. Soc. Am., 120:458–469, 2006.

[25] N. Roman, D. Wang, and G. Brown. Speech segregation based on sound localization. J. Acoust. Soc. Am., 114:2236–2252, 2003.132

[26] S. Seneff and V. Zue. Transcription and alignment of the timit database. In Proc. Second Symposium on Advanced Man-Machine Interface through Spoken Language, 1988.

[27] R. Shannon, A. Jensvold, M. Padilla, M. Robert, and X. Wang. Consonant recordings for speech testing. J. Acoust. Soc. Am., 106:71–74, 1999.

[28] R. V. Shannon, F-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid. Speech recognition with primarily temporal cues. Science, 270:303–304, 1995.

[29] R. Smith. Adaptation, saturation and physiological masking singe auditory-nerve fibers. J. Acoust. Soc. Am., 65(1):166–178, 1979.

[30] J. Sohn, N. S. Kim, and W. Sung. A statistical model-based voice activity detection. IEEE Signal Process. Lett, 6:1–3, 1999.

[31] J. Sohn and W. Sung. A voice activity detector employing soft decision based. Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, pages 365–368, 1998.

[32] J. Sohn and W. Sung. A voice activity detector employing soft decision based noise spectrum adaptation. In Proc. IEEE Int. Conf.Acoust., Speech, Signal Processing, pages 365–368, 1998.

[33] H. Steeneken and T. Houtgast. A physical method for measuring speech transmission quality. J. Acoust. Soc. Am., 67(1):318–326, 1980.

[34] P. Stelmachowicz, A. Pittman, B. Hoover, and D. Lewis. Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults. J. Acoust. Soc. Am., 110(4):2183–2190, 2001.

[35] K. Stevens. Toward a model for lexical access based on acoustic landmarks and distinctive features. J. Acoust. Soc. Am., 111(4):1872–1891, 2002.

[36] K. N. Stevens, S. Y. Manuel, S. Shattuck-Hufnagel, and S. Liu. Implementation of a model for lexical access based on features. In Proc. ICSLP, pages 499–502, 1992.

[37] G. Stickney, P. Assmann, J. Chang, and F-G. Zeng. Effects of implant processing and fundamental frequency on the intelligibility of competing sentences. J. Acoust. Soc. Am., 122(2):1069–1078, 2007.

[38] G. Stickney, F-G. Zeng, R. Litovsky, and P. Assmann. Cochlear implant speech recognition with speech maskers. J. Acoust. Soc. Am., 116(2):1081–1091, 2004.