FORMANT SHIFTING FOR SPEECH INTELLIGIBILITY IMPROVEMENT IN CAR NOISE ENVIRONMENT

Karan Nathwani[1], *Morgane Daniel*[2], *Gaël Richard*[1], *Bertrand David*[1] *and Vincent Roussarie*[2]

1: LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France
2: PSA Peugeot Citroën, Chemin de Gisy, 78943 Vélizy-Villacoublay

ABSTRACT

In this paper, we propose a novel approach aiming at improving the intelligibility of speech in the context of in-car applications. Speech produced in noisy environments is subject to the Lombard effect which gathers a number of voice transformation effects compared to the speech produced in calm environments. To improve intelligibility of in car speech (radio, message alerts,...), we propose to modify the original speech signal by incorporating one of the important Lombard effect, namely the shift of the lower formant center frequencies away from the competing noise regions. The proposed approach exploits traditional Linear Prediction analysis and overlap and add synthesis. We explore several modification strategies and the merit of each modification is evaluated using both objective and subjective tests. It is in particular shown that the improvement of speech intelligibility in car noise is significantly improved for a majority of listeners.

***Index Terms*—** Speech Intelligibility, Lombard Effect, Formant Shifting, Filter Masking, HINT and Car Noise Environment.

1. INTRODUCTION

While a large amount of literature is dedicated to improve the speech quality by reducing the level of surrounding noise, the effort to design algorithms for improving speech intelligibility is comparably much less. Furthermore, it has also been observed in [1], [2] that improving speech quality does not necessarily improves speech intelligibility. This non-correlation between speech quality and speech intelligibility has been analyzed in more depth in [3]. In this paper, we thus focus on improving the intelligibility of speech and we more specifically target the context of in-car applications, where the issue is particularly relevant considering the noise background produced by the engine, the air turbulence and the road/tire contact. Speech intelligibility improvement could also have significant applications in mobile communications, hearing aids and cochlear implant devices.

In the context of speech production, the source/filter model is a wide spread paradigm for describing the speaker or the phoneme characteristics [4]. Some proposals to enhance intelligibility focus more on modifying the source characteristics such as the speech rate (by increasing the duration of the signal) and a pitch contour [5]. On the other hand, vocal tract modification based methods are primarily used in order to transform a voice into another. These approaches are based on a conventionally learning phase for finding the statistical differences between the parameters of two voices using Hidden Markov Models, Gaussian Mixture Model, neural networks [6], [7] or learn the transformation function (weighted linear interpolations and bilinear models [8]).

This work is supported by the French National Research Agency (ANR) as a part of AIDA Project.

For in-car environment, noise robust speech recognition systems have been developed by normalizing the cepstral domain segmental features [9], [10] or with the help of Linear Prediction (LP) applied to one sided autocorrelation sequence together with a robust similarity measure [11], [12], [13]. However, much less work is done in the context of improving speech intelligibility car noise environment at various speed. Hereafter, we describe a novel system which implements an interesting feature of the so-called Lombard effect [14–17], naturally produced by speakers in noisy or competing environments. It relies on modifying the vocal tract characteristics and particularly the central frequency of the formants. We are thus targeting a modification that could be assessed as "natural" by the listeners. Moreover, it has been observed that the Lombard speech has been scored as more intelligible than speech produced in quiet when both are mixed with noise at the same SNR [18], [19].

To synthesize the effect, the Linear Prediction Coding (LPC) framework is utilized for computing poles and formants frequencies of speech signal which are then shifted to a better hearing region. This work is a part of a wider project with prime objective to provide means for improving the intelligibility of vocal sound messages coming from the device (like radio, telephone or alert message) in a car noise environment. The performance of the proposed method is evaluated by conducting subjective and objective tests. The subjective analysis is obtained using the Hearing In Noise Test (HINT) protocol [20], [21]. The objective analysis is conducted using perceptual evaluation of speech quality (PESQ), weighted spectral slope (WSS), speech intelligibility index (SII) and log likelihood ratio (LLR) measures [22], [23]. Additionally in this work, the listeners are screened out into partial or no hearing loss based on their speech reception threshold obtained through subjective analysis and prescribed different treatments or modifications.

The remainder of the paper is organized as follows. Section 2 describes the proposed formant location modification framework. This is followed by the performance evaluation which includes the development of the database and its protocol used for subjective analysis in section 3. Section 4 presents a brief conclusion and future scope.

2. PROPOSED FORMANT LOCATION MODIFICATION IN LPC FRAMEWORK

2.1. Principle of Formant Shifting

The contribution of this work is to develop a speech intelligibility system for normal hearing listeners in car noise environment by converting normal speech to Lombard speech. Instead of using statistical model or transformation function, this conversion is achieved by synthesizing one of the properties of the Lombard effect namely shifting upward the location of the formant center frequencies. The amount of shift is controlled by a specific delta function.

The principle of formant shifting is depicted in Figure 1. The detailed explanation of each step is as follows.

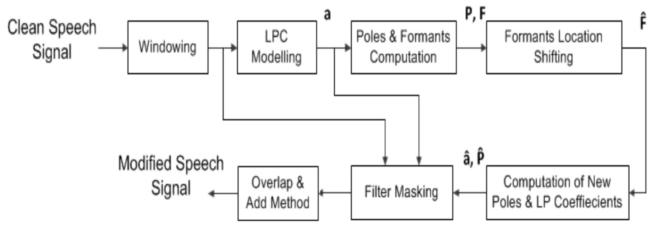

Fig. 1. Flow diagram of the proposed formant shifting method.

1. **Windowing** : The clean speech signal is windowed before applying linear prediction coding (LPC) modelling. The window length used herein is of 20 ms duration.
2. **LPC Modeling** : A LPC model is a powerful front end tool to process speech signal. In LPC model, a windowed speech signal $s(n,m)$ can be expressed in terms of a p^{th} order linear predictor [24] as
$$s(n,m) = \sum_{j=1}^{p} a_j s(n-j,m) + e(n,m) \tag{1}$$
Here, a_j are the LP coefficients, $e(n,m)$ is the residual error and p is equal to 12. n and m are speech sample index and short time window index respectively. The LP coefficients a_j are obtained by minimizing C, the mean square residual error in each analysis window.
3. **Poles and Formants Location Computation**: The LP filter $A(z)$ is then computed from LP coefficients as
$$A(z) = 1 + \sum_{j=1}^{p} a_j z^j \tag{2}$$
The poles **P** and formant frequencies location **F** are then estimated as the roots of the LP coefficients and angles of the estimated poles respectively.
4. **Formant Shifting** : The formants obtained in previous steps are then shifted upwards by an amount specified by the delta function ($\Delta(f)$). Once, the delta function is generated as explained in section 2.2, the formant location shifting $\hat{F}(f)$ is performed as
$$\hat{F}(f) = \begin{cases} F(f) + \Delta(f) & if \quad f_1 < f < f_3 \\ F(f) & \text{otherwise} \end{cases} \tag{3}$$
Here, f_1 and f_3 are the first and third formant frequency respectively, empirically selected to achieve different delta function shapes. Similarly, the shifting of formants location are performed for the negative frequencies of f_1 and f_3. Thus, combining both positive and negative shift of formants location, we finally obtain $\hat{\mathbf{F}}$ as the new location of the formants.
5. **Computation of New Poles and LP Coefficients** : The new poles $\hat{\mathbf{P}}$ are then computed from the estimated new formants location $\hat{\mathbf{F}}$ as
$$\hat{\mathbf{P}} = B * (cos(\hat{\mathbf{F}}) + i \;\; sin(\hat{\mathbf{F}})) \tag{4}$$
Here, B is the amplitude of the original poles. The roots of $\hat{\mathbf{P}}$ are then converted into polynomial and real portion of polynomial corresponds to modified LP coefficients.
6. **Filter Masking** : Synthesizing the modified speech using the modified LP coefficients results in a signal with some rare but annoying localised artifacts. To reduce these artifacts which are mostly due to phase incoherence, the original speech amplitude spectrum is modified in order to match the modified LP spectrum while the original phase spectrum is kept. The modified speech frame is then obtained by inverse Fourier transform and its overall energy is equalized to the energy of the unprocessed frame.
7. **Overlap and Add (OLA) Method** : Finally, the modified signal is obtained using classic Overlap and Add synthesis.

2.2. Various Shapes of Delta Function

In this section, the four different modifications on clean speech signal are described by varying the shapes of the delta function. These modifications will produce different effects on clean speech.

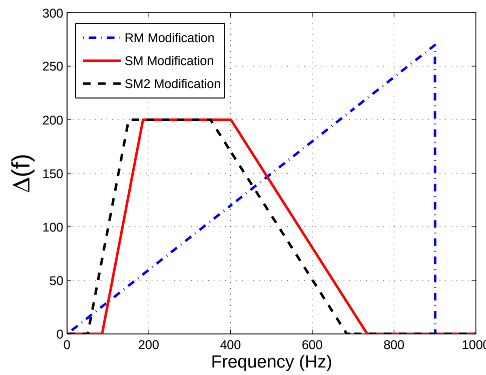

Fig. 2. Delta function used for modification RM (dash-dot line), SM (solid line) and SM2 (dash line) respectively.

- *Ramp Shape Modification (RM)* : The delta function in this case is obtained as a percentage of shift of center frequency of formants. The delta function used herein is illustrated by dash-dot line in Figure 2. It can be seen from RM in Figure 2 that most of the modification is done around 900 Hz.
- *Smooth Shape Modification (SM)* : In this condition, the smooth modification is achieved by selecting a delta function with shape shown in solid line in Figure 2. In this case, no modification is done on formants frequencies lying below 85 Hz and above 735 Hz. The formants frequencies lying between 85 Hz and 735 Hz are modified according to the slow increasing ramp up function followed by flat and ramp decreasing functions respectively.
- *Smooth Shape Modification 2 (SM2)* : In this case, the shape of the delta function is similar to SM but the delta function is shifted leftwards to make an smoother modification on clean speech signal as shown (dash line) in Figure 2. Thus, the formants located below 50 Hz are not modified in SM2 compared to RM. Also, the formants lying above 685 Hz are not modified in SM2 in contrast to SM and RM.
- *SNR Dependent Smooth Shape Modification (SNRSM)* : In this case, the different shapes of smooth delta function is obtained based on different SNR. This means that at low and high SNR, aggressive and no modification is required respectively. For instance, when the signal is mixed with noise at moderate SNR, a delta function based on this SNR is used to achieve moderate modifications on clean speech to improve

Table 1. Mean objective scores for all 5 conditions using SII, PESQ, LLR and WSS measures at different SNRs.

	SNR=-26				**SNR=-14**				**SNR=-8**				**SNR=0**				**SNR=10**			
Method	SII	PESQ	LLR	WSS	SII	PESQ	LLR	WSS	SII	PESQ	LLR	WSS	SII	PESQ	LLR	WSS	SII	PESQ	LLR	WSS
NM	0.39	1.26	1.37	70.43	0.63	1.91	0.61	64.18	0.76	2.31	0.42	55.48	0.89	2.94	0.26	40.96	**0.98**	3.73	0.14	23.29
RM	0.41	1.30	1.39	68.68	0.63	1.90	0.59	63.27	0.76	2.29	0.41	54.91	0.89	2.90	0.26	40.63	0.97	3.71	0.14	23.48
SM	0.41	1.42	1.34	67.78	0.62	2.04	0.57	62.22	0.75	2.42	0.39	54.01	0.89	3.03	0.25	39.85	**0.98**	3.81	0.13	22.54
SM2	**0.44**	**1.46**	**1.25**	**64.95**	**0.65**	**2.11**	**0.53**	**59.05**	**0.77**	**2.52**	**0.37**	**51.00**	**0.90**	**3.14**	**0.23**	**37.26**	**0.98**	**3.87**	**0.12**	**21.24**
SNRSM	0.41	1.40	1.32	66.10	0.63	2.06	0.57	61.61	0.76	2.45	0.40	53.50	0.89	3.06	0.25	39.60	**0.98**	**3.87**	0.13	21.70

the intelligibility of signal in noise. The difference between SNRSM and SM is in the different shapes of delta function used by SNRSM which are varied with SNR.

The general strategy followed to select different shapes of delta function is based on best PESQ scores obtained for all modifications. PESQ measure has been widely used in literature as an objective measure for speech enhancement and has shown relatively strong correlation for speech intelligibility evaluation [25]. In this work, PESQ scores are obtained between the synthesized signal (obtained from modification such as RM, SM, SNRSM and SM2) and the synthesized signal added with noise at different SNR. These PESQ scores are also compared with the PESQ scores computed from the clean signal and the clean signal plus noise. Additionally, we also make sure that the signal should not be modified so much that signal naturalness is lost. In order to retain the naturalness, PESQ scores between the clean speech and synthesized speech (obtained from different modification) is not allowed to go less than 3. It may be noted that the delta function defined for modifications RM, SM and SM2 are fixed for all SNR in contrast to SNRSM.

2.3. Filter Masking

As briefly discussed above, the direct synthesis from the modified LP coefficients leads to unsatisfactory results. Thus, in order to limit the artifacts in the modified speech signal, a different strategy is preferred. Given $S(f) = |S(f)|e^{j\phi(f)}$ the Short Time Fourier transform of an original speech frame $s(n,m)$ and $|A(f)|$ the module of the LP spectrum of $s(n)$, the modified speech frame spectrum $S'(f)$ is obtained as $S'(f) = |S(f)| * |A'(f)|/|A(f)|e^{j\phi(f)}$ where $A'(f)$ is the Fourier transform of the modified LP coefficients. The modified speech frame, obtained by inverse Fourier transform, is then normalized to match the energy of the original speech frame. The complete modified speech signal is then obtained using a classic overlap and add procedure.

3. PERFORMANCE EVALUATION

In this section, performance evaluation of the proposed method is conducted for all four proposed modifications and are compared to the intelligbility of the original signal with no modification (NM) at different SNRs. The database and material used in the evaluation is also presented in this section. This section deals with subjective and objective evaluation measures to observe the correlation between them. The subjective test is performed based on the HINT protocol [21]. The objective measures used for evaluating intelligibility are speech intelligibility index (SII) [26], PESQ, LLR and WSS [22].

3.1. Development of the Database and Material

The French lists for HINT were adapted from the English version in [20]. Hence, 5 lists of 20 sentences used for the test were taken from an audiometry CD recording [27]. The 4 modifications mentioned in section 2.2 along with the reference clean signal (NM) were applied to those lists. The car noise was recorded by a Head Acoustics recording head dummy in a Peugeot 308 at 130km/h steady speed. The power density spectrum of the noise is illustrated in Figure 3. It can be seen that it contains few energies/frequencies above 1000 Hz. Thus based on noise characteristics, various shapes of delta function have been designed in Section 2.2 resulting formant modification only below 1000 Hz. Finally the sentences obtained from 4 modifications along with the clean reference signal are mixed with the car noise recording at various SNR (selected empirically). The mix was presented under Sennheiser HD650 headphones and played from a Head Acoustics Digital Equalizer (PEQ V). The level of the noise is set at 67dB speech perception level (SPL) as it was the level in the car during the recording and only the level of the speech is varying.

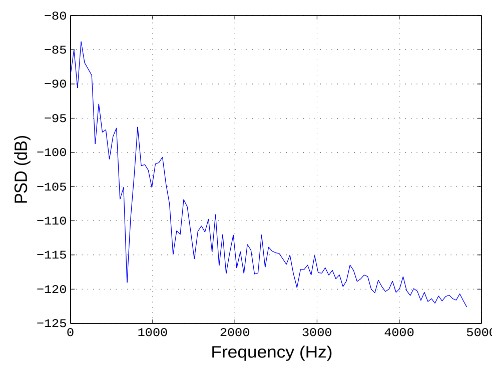

Fig. 3. Power Spectrum Density (PSD) of the car noise recorded at 130km/h

3.2. Objective Evaluation

The objective evaluation is conducted using PESQ, LLR, WSS [25] and SII [26], [28] measures. PESQ analyzes the speech signal sample-by-sample after a temporal alignment of corresponding excerpts of the synthesized signal (SS) and the synthesized signal added with noise (SSN). PESQ principally models mean opinion score (MOS) results that cover a scale from 1 (bad) to 5 (excellent). WSS is a distance measure which computes the weighted difference between the spectral slopes of SS and SSN in each frequency band. LLR is a LPC-based measure which finds the spectral envelope difference between the SS and SSN [23]. SII model [26] basically calculates the average amount of speech information available to a listener. The value of the SII varies from 0 (completely unintelligible) to 1 (perfect intelligibility).

Table 1 shows the mean scores for all the objective measures at various SNRs for different modifications. These mean objective scores are computed on all the sentences of the databases. It can be noted from the Table 1 that at very low SNR, higher SII and PESQ scores along with lower LLR and WSS scores are observed for different modifications compared to no modification case (clean signal). In general, a method having higher SII, PESQ scores and lower LLR, WSS scores is supposed to have high intelligibil-

Table 2. Number of participants having positive and negative relative thresholds for different modifications.

Methods	Total Sum of RT ($TSRT$)		Different Combinations of RT from -3 to 3 dB						
	$TSRT \leq -1$	$TSRT \geq 1$	$RT \leq -3$	$-3 < RT \leq -2$	$-2 < RT \leq -1$	$-1 < RT < 1$	$1 \leq RT < 2$	$2 \leq RT < 3$	$RT \geq 3$
RM	**10**	5	0	**4**	**6**	14	3	1	1
SM	8	**14**	**1**	2	5	7	6	**6**	2
SM2	4	**14**	**1**	1	2	11	5	4	**4**
SNRSM	5	**14**	0	1	4	10	**8**	4	2

ity [22] for modified speech when played in noise compared to original speech. However, the RM shows marginal improvement in intelligibility measure scores over NM compared to SM, SNRSM and SM2 at low SNRs. When SNR increases, the difference in the mean objective scores between the NM and modification based on RM, SM and SNRSM is decreased gradually. On the other hand, SM2 shows significant improvement for all objective scores at various SNRs over NM and other modifications. SNRSM shows second best performance after SM2. However on increasing the SNR to zero or further positive, it is also observed that the difference in intelligibility between no modification and all modifications including SM2 is vanished.

3.3. Subjective Evaluation

3.3.1. HINT Protocol & Participants for Subjective Evaluation

A slightly modified Hearing In Noise Test (HINT) [21] is used for measuring the enhancement of intelligibility provided by the different speech treatments. A speech reception threshold (SRT) is obtained for each 5 conditions. A total of 29 native French speaking male and female subjects participated to the test. All subjects were screened normal hearing (mean besides 20dB HL over 0.5 to 6 kHz). Their ages ranged from 21 to 50 years with a mean age of 31.

The subjects are asked to listen to a sentence and to repeat aloud what they hear. The first sentence is presented at a level below the SRT, usually at -30 dB. Then the SNR is increased by 2 dB steps until it is repeated correctly. The subsequent sentences are presented once each (inorder to avoid traning effect) at a level depending on the correct repetition of the preceding sentence. If it is repeated correctly, presentation level is attenuated by 2 dB, otherwise it is increased by 2 dB. For each condition, a 20-sentences list is presented in a random order. There is one list per condition. The presentation order of the 5 conditions is balanced over the participants as well as the presentation of the lists. The experimenter compares the listener's response to a text version to determine whether it is correct or not. Small variations in the sentences are allowed as specified in [21].

3.3.2. Experimental Results

In this section, speech reception thresholds for each condition is compared to the reference no modification (NM) condition. SRTs have been averaged over the last fifteen SNR obtained with HINT. The relative threshold (RT) (in dB) is obtained by taking the difference between the SRT of NM and SRT of each modification. A positive relative threshold indicates an improvement in intelligibility for the particular modification.

Table 2 illustrates the number of participants showing positive and negative RT (who gets benefit or not) from each modification. We consider that results with RT between -1 and 1 dB are not significant and are not further discussed below. It can be seen from Table 2 that only 17.2% of participants have a significant positive RT for RM indicating negligible improvement over NM. However, the other modifications SM, SNRSM and SM2 shows 48.3% of participants each with RT significantly exceeding 1 dB. This illustrates that these modifications improve intelligibility in the presence of high speed car noise compared to NM. It may be noted that the SM2 and SM show more participants with RT above 2 dB. This may be explained by the fixed delta function parameters at all SNRs used in SM and SM2 modifications compared to SNRSM. Hence there is least modifications by SNRSM than SM and SM2 resulting less improvement at high SNR and hence less audible artifacts for SNRSM modification at high SNR. SM2 gets the highest numbers of participants with RT above 2 dB but also the least participants with negative RT. Only 27.6%, 17.2% and 13.8% of participants show worst performances with SM, SNRSM and SM2 respectively with respect to NM. These percentages could be decreased even further if we reduce the artifacts remaining over some sentences.

Table 3. Statistical analysis of the RT obtained from subjective evaluation using 29 listeners.

	Mean	Std Deviation	Minimum	Maximum
RM	-0.286	**1.517**	**-2.533**	4.400
SM	0.471	2.106	-4.533	4.267
SM2	**0.948**	1.980	-4.800	**4.533**
SNRSM	0.748	1.567	-4.533	3.867

Table 3 gives a statistical analysis of the subjective evaluation obtained from RT as explained previously. The negative mean for RM modification confirms there is no improvement in intelligibility. The best RT mean is obtained for SM2 with 0.94 dB. However, the standard deviation of SM2 is 1.98 indicating that SM2 can also decrease the intelligibility for a few participants due to audible artifacts at high SNRs. On the other hand, SNRSM shows second highest mean with lowest standard deviation of 1.567 but also has a lowest maximum RT of 3.867 dB over the positive RT. A way to improve those methods would be to find a compromise between SM2 and SNRSM by reducing audible artifacts at high SNR, resulting in less negative RT but still high positive RT. The artifacts mostly correspond to musical noise caused by the modification process.

4. CONCLUSION & FUTURE SCOPE

In this work, we have proposed a novel method for improving speech intelligibility in the car environment for normal hearing listeners. The subjective and objective evaluations have shown improvement in intelligibility for the proposed modifications when compared to clean speech played in noise. Smooth shape modification 2 (SM2) shows the best results for objective measures at various SNRs. However in the subjective evaluation, fixed delta function parameters used in SM2 results in audible artifacts at high SNR. This decreases the percentage of participants having positive relative threshold (RT).

Future scope would be to further improve the proposed modifications which will find a better compromise between SM2 and SNR dependent smooth shape modification (SNRSM) by reducing audible artifacts at high SNR. Additionally, future work would also investigate new methods which will consider hearing abilities of impaired persons in shifting the formants away from the region of their loss.

5. REFERENCES

[1] G. Kim and P. C. Loizou, "Improving speech intelligibility in noise using environment-optimized algorithms," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 2080–2090, 2010.

[2] Y. Hu and P. C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1777–1786, 2007.

[3] P. C. Loizou and G. Kim, "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 47–56, 2011.

[4] K. S. Rao and B. Yegnanarayana, "Voice conversion by prosody and vocal tract modification," in *9th IEEE International Conference on Information Technology (ICIT)*, 2006, pp. 111–116.

[5] J. S. Laures and K. Bunton, "Perceptual effects of a flattened fundamental frequency at the sentence level under different listening conditions," *Journal of communication disorders*, vol. 36, no. 6, pp. 449–464, 2003.

[6] J. Nurminen, V. Popa, J. Tian, Y. Tang, and I. Kiss, "A parametric approach for voice conversion," *TCSTAR WSST*, pp. 225–229, 2006.

[7] S. Desai, E. V. Raghavendra, B. Yegnanarayana, A. W. Black, and K. Prahallad, "Voice conversion using artificial neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 3893–3896.

[8] M. Zhang, J. Tao, J. Tian, and X. Wang, "Text-independent voice conversion based on state mapped codebook," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 4605–4608.

[9] O. Viikki and K. Laurila, "Cepstral domain segmental feature vector normalization for noise robust speech recognition," *Speech Communication*, vol. 25, no. 1, pp. 133–147, 1998.

[10] M. Shozakai, S. Nakamura, and K. Shikano, "Robust speech recognition in car environments," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1998, pp. 269–272.

[11] J. Hernando, C. Nadeu, and J. Mariño, "Speech recognition in a noisy car environment based on lp of the one-sided autocorrelation sequence and robust similarity measuring techniques," *Speech communication*, vol. 21, no. 1, pp. 17–31, 1997.

[12] S. Beack, H. Minsoo *et al.*, "A new speech enhancement algorithm for car environment noise cancellation with mbd and kalman filtering," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 88, no. 3, pp. 685–689, 2005.

[13] R. Chen, C.-F. Chan, H.-C. So, J. S. Lee, and C. Leung, "Speech enhancement in car noise envoronment based on an analysis-synthesis approach using harmonic noise model," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 4413–4416.

[14] E. Lombard, "Le signe de l'elevation de la voix," *Ann. Maladies Oreille, Larynx, Nez, Pharynx*, vol. 37, no. 101-119, p. 25, 1911.

[15] Y. Lu and M. Cooke, "Speech production modifications produced by competing talkers, babble, and stationary noise," *The Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 3261–3275, 2008.

[16] W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses," *The Journal of the Acoustical Society of America*, vol. 84, no. 3, pp. 917–928, 1988.

[17] M. Cooke, C. Mayo, and C. Valentini-Botinhao, "Intelligibility-enhancing speech modifications: the hurricane challenge." in *Interspeech*, 2013, pp. 3552–3556.

[18] J. J. Dreher and J. O'Neill, "Effects of ambient noise on speaker intelligibility for words and phrases," *The Journal of the Acoustical Society of America*, vol. 29, no. 12, pp. 1320–1323, 1957.

[19] Y. Lu and M. Cooke, "The contribution of changes in f0 and spectral tilt to increased intelligibility of speech produced in noise," *Speech Communication*, vol. 51, no. 12, pp. 1253–1262, 2009.

[20] V. Vaillancourt, C. Laroche, C. Mayer, C. Basque, M. Nali, A. Eriks-Brophy, S. D. Soli, and C. Giguère, "Adaptation of the hint (hearing in noise test) for adult canadian francophone populations: Adaptación del hint (prueba de audición en ruido) para poblaciones de adultos canadienses francófonos," *International Journal of Audiology*, vol. 44, no. 6, pp. 358–361, 2005.

[21] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *The Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, 1994.

[22] P. C. Loizou, *Speech enhancement: theory and practice*. CRC press, 2013.

[23] J. Ma, Y. Hu, and P. C. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," *The Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3387–3405, 2009.

[24] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*. Prentice Hall, 1978.

[25] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2001, pp. 749–752.

[26] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2010, pp. 4214–4217.

[27] CND. (2015) National college of audioprothesist - speech audiometry compact disk (text in french). [Online]. Available: http://www.college-nat-audio.fr/listes-cd-audiometrie-vocale.html

[28] E. Jokinen, U. Remes, M. Takanen, K. Palomaki, M. Kurimo, and P. Alku, "Spectral tilt modelling with extrapolated gmms for intelligibility enhancement of narrowband telephone speech," in *14th IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014, pp. 164–168.