



Original

Improving speech intelligibility for binaural voice transmission under disturbing noise and reverberation using virtual speaker lateralization

A.L. Padilla Ortiz, F. Orduña-Bustamante*

Grupo de Acústica y Vibraciones, Centro de Ciencias Aplicadas y Desarrollo Tecnológico (CCADET), Universidad Nacional Autónoma de México (UNAM), México D.F., México

Received 9 August 2014; accepted 27 November 2014

Abstract

Subjective speech intelligibility tests were carried out in order to investigate strategies to improve speech intelligibility in binaural voice transmission when listening from different azimuth angles under adverse listening conditions. Phonetically balanced bi-syllable meaningful words in Spanish were used as speech material. The speech signal was played back through headphones, undisturbed, and also with the addition of high levels of disturbing noise or reverberation, with a signal to noise ratio of $SNR = -10$ dB and a reverberation time of $T_{60} = 10$ s. Speech samples were contaminated with interaurally uncorrelated noise and interaurally correlated reverberation, which previous studies have shown the more adverse. Results show that, for speech contaminated with interaurally uncorrelated noise, intelligibility scores improve for azimuth angles around $\pm 30^\circ$ over speech intelligibility at 0° . On the other hand, for interaurally correlated reverberation, binaural speech intelligibility reduces when listening at azimuth angles around $\pm 30^\circ$, in comparison with listening at 0° or azimuth angles around $\pm 60^\circ$.

All Rights Reserved © 2015 Universidad Nacional Autónoma de México, Centro de Ciencias Aplicadas y Desarrollo Tecnológico. This is an open access item distributed under the Creative Commons CC License BY-NC-ND 4.0.

Keywords: Speech intelligibility; Speaker angles; Uncorrelated noise; Correlated reverberation

1. Introduction

Speech intelligibility measures the accuracy with which a normal listener can understand a spoken word or phrase. Two factors that reduce speech intelligibility are background noise and reverberation (Houtgast & Steeneken, 1985a, 1985b; Steeneken & Houtgast, 1980). However, there are other factors involved, such as the orientation of the speaker relative to the listener; the azimuth angle has a significant influence on the perception of sound that reaches our ears (Nordlund & Fritzell, 1963). Both the acoustic environment in which the listener is immersed and the azimuth angle play a significant role in speech perception.

There are some studies of the effect of azimuth angle on the speech pattern or in speech intelligibility. Flanagan (1960) measured the directionality of the sound pattern radiated from the mouth, although he did not carry out measurements of speech intelligibility. Some studies investigate how speech signals are influenced by the angle of incidence of sound towards the listener, in monaural and binaural rendering; the results show a clear influence of azimuth on intelligibility (Nordlund, 1962; Nordlund & Lidén, 1963; Plomp & Mimpén, 1981). Studies also

demonstrate that speech intelligibility for binaural listening is significantly better than for monaural listening at all azimuth angles tested (Nordlund, 1962; Nordlund & Lidén, 1963; Plomp & Mimpén, 1981).

Additional research in this direction was carried out by Plomp and Mimpén (1981), who measured the speech-reception threshold (SRT) for sentences, as a function of the orientation of the speaker's head and the azimuth of a noise source. Their results show that the maximum effect of direction of speech radiation from the mouth of the speaker is about 6 dB, in terms of signal-to-noise ratio (SNR), and the maximum effect of the azimuth of the noise source is about 10 dB in SNR.

Objective conceptual models for binaural speech intelligibility have been proposed (Beutelmänn & Brand, 2006; Beutelmänn et al., 2010; Durlach, 1963; Jelfs et al., 2011; Lavandier & Culling, 2010; Van Wijngaarden & Drullman, 2008), trying to instrumentally reproduce the results of neural processing of binaural speech signals by the human brain. A commonly accepted model is the binaural equalization-cancellation (EC) auditory mechanism proposed by Durlach (1963), which takes advantage of the fact that acoustic signals coming from different directions cause different interaural time and level differences. The equalization step is supposed to neurally attenuate and delay the signal from one ear with respect to the other, so that left- and right-ear signals are first matched in amplitude

*Corresponding author.

E-mail address: felipe.orduna@ccadet.unam.mx (F. Orduña-Bustamante).

and phase, while in the cancellation step, the signal in one ear is then neurally subtracted from the other ear, aiming at maximizing the SNR.

There has been a growing interest in artificially controlling the apparent spatial localization of sound sources; this has important implications in some applications, such as in air traffic control displays (Begault & Wenzel, 1992), devices for the blind (Loomis et al., 1990; Loomis et al., 1994; Loomis et al., 1998), conference systems (Kanada, 2004, 2005) and others. Figure 1 illustrates an application which is proposed in this article, and that motivates the speech intelligibility tests that are presented. In the proposed application, voice transmission is rendered binaurally through headphones, this requires processing a monophonic speech signal through Head Related Transfer Functions (HRTF) to produce binaural signals (left and right ear), consistent with the perception of a virtual speaker, possibly lateralized. The azimuth angle is to be chosen so as to improve speech intelligibility, especially under acoustically adverse listening conditions in the presence of high levels of noise and reverberation.

In this article, we present experiments that extend previous and current research on speech intelligibility, more especially with respect to binaural rendering of voice transmissions, considering aspects such as: influence of noise and reverberation and angle of presentation of the speech signal. Interaurally uncorrelated noise and interaurally correlated reverberation are used, as these conditions have been shown more detrimental to speech intelligibility in a previous study (Padilla-Ortiz & Orduña-Bustamante, 2012), from which Table 1 shows a summary of results.

As it can be seen in Table 1, especially for binaural presentation at $\pm 30^\circ$, speech intelligibility under interaurally correlated noise is larger than for uncorrelated noise, and larger under interaurally uncorrelated, rather than correlated, reverberation. These seemingly paradoxical results are in fact consistent, and can be explained on the basis of the EC model (Durlach, 1963), according to which, interaurally correlated noise enables the cancellation step, while interaurally uncorrelated noise makes cancellation difficult or impossible. In that case, lateralization of the speech signal generates interaural differences that the binaural hearing mechanism can still exploit in order to emphasize the

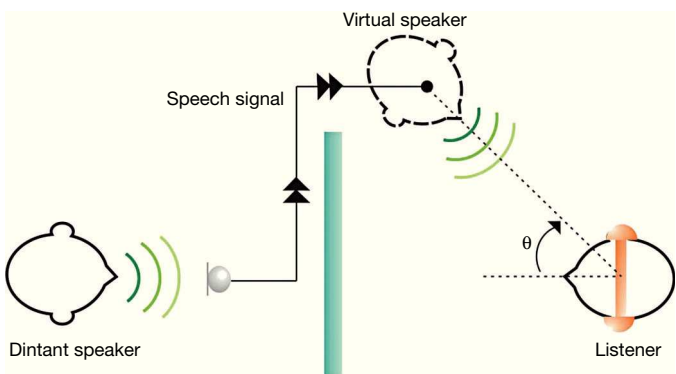


Fig. 1. Proposed voice transmission system, with binaural rendering through headphones. The virtual speaker can be lateralized to improve speech intelligibility under acoustically adverse listening conditions, with high levels of noise and reverberation.

Table 1
Speech Intelligibility Scores (Percent) Under Interaurally Correlated and Uncorrelated Noise and Reverberation, From Padilla-Ortiz and Orduña-Bustamante (2012).

		Binaural 0°	Binaural $\pm 30^\circ$
Noise	Correlated	91.1 \pm 2.56	97.5 \pm 1.98
	Uncorrelated	89.7 \pm 4.45	90.8 \pm 4.10
Reverberation	Correlated	78.1 \pm 9.80	73.0 \pm 6.87
	Uncorrelated	79.8 \pm 8.80	83.4 \pm 8.04

speech-to-noise ratio, slightly improving speech intelligibility. On the other hand, under interaurally uncorrelated reverberation, the reverberant impulse responses are different in both ears, but maintain a time invariant interaural phase relationship, which enables the binaural hearing mechanism to equalize and cancel the unwanted disturbance in order to improve speech intelligibility (Padilla-Ortiz & Orduña-Bustamante, 2012). In view of these previous results, and for the purposes of further investigation of the opposite, more acoustically adverse conditions, interaurally uncorrelated noise and interaurally correlated reverberation are used in the present study.

2. Speech Intelligibility Tests With Different Speaker Angles

The objective of these tests is to assess the intelligibility at different angles on the horizontal plane in order to determine possibly preferred positions at which the listener better perceives the speech signal under adverse listening conditions. The study was performed in three stages: first, binaural recordings of speech material were made, then these recordings were processed in order to contaminate them with interaurally uncorrelated noise, or interaurally correlated reverberation, finally, subjective speech intelligibility tests were carried out.

2.1. Subjects

Fifteen subjects (7 female, 8 male) took part in the listening tests. Their age range was 19 to 34 years, with an average of 24.5 years. All listeners were audiometrically screened (Audiometer Brüel & Kjaer type 1800), and all of them had pure-tone thresholds lower than 15 dB Hearing Level (HL) at all audiometric (octave) frequencies between 500 Hz and 8000 Hz. All of them were university students and Mexican Spanish native speakers. None of the participants were previously familiar with the lists of words used in the study. Listeners had no prior experience in any psychoacoustic experiments.

2.2. Speech Stimuli

Bi-syllable words with meaning in Spanish were used in this study. Prosodically, all of the words are of paroxytone type (with an accent on the penultimate syllable of the word), representing the most common type of bi-syllable words in Spanish (Tato, 1949; Zubick et al., 1983). Although standard speech test material is widely available in English (ANSI/ASA S3.2, 2009;

ISO/TR 4870, 1991), this is not the case in Spanish, for which some research has been published already in that respect (Benitez & Speaks, 1968; Berruecos & Rodriguez 1967; Cancel, 1965; Ferrer, 1960; Tato, 1949; Zubick et al., 1983). Some of this research has pointed out the difficulty of making lists of monosyllabic words in Spanish, because of the lack of a sufficient number of meaningful monosyllables (Tato, 1949; Zubick et al., 1983).

In our tests, speech material consisted of four different lists of words, with 50 bi-syllable phonetically balanced (PB) Spanish words each (Castañeda & Pérez, 1991) (see Appendix). In the test recording, words were preceded by different carrier sentences in Spanish, which in translation are similar to: "The next word is..." Speech was produced in Spanish by a female speaker born in Mexico City.

3. Recording of Speech Material at Different Azimuth Angles

Binaural recordings were carried out in an anechoic chamber at different angles from -60° to $+60^\circ$ in steps of 15° (Fig. 2); the elevation angle was always 0° . The speech signals were binaurally recorded with a Head and Torso Simulator, Brüel & Kjaer type 4128 (Fig. 2). Binaural recording began at $+60^\circ$, which in this study is the angle at which the loudspeaker is closer and more directly oriented towards the ears of the mani-

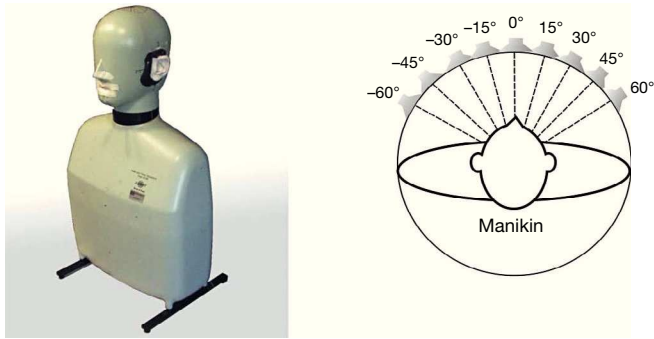


Fig. 2. Acoustic manikin, Brüel & Kjaer Head and Torso Simulator type 4128, used in the binaural recordings (left). Position of the loudspeaker for binaural recordings (right).

kin. This position had the highest recording level, and this allowed us to set the same microphone gain for this and all other angles recorded, avoiding digital overload in recording the microphone signal. The original monophonically recorded source speech signal was digitally stored in a laptop and was played back through an Event Electronics ALP5 loudspeaker. Both playback and recording were done through an M-Audio Pro Fire 610 digital audio interface. The distance between the head and torso simulator and the loudspeaker was 1 m. In order to obtain binaural recordings at different azimuth angles, the loudspeaker was moved to the corresponding position. The binaural speech stimuli were recorded and stored as 16-bit PCM digital audio at 44.1 kHz sampling rate.

The experimental setup for binaural recording is shown in Figure 3.

3.1. Processing of Speech Material

Binaural recordings were digitally processed in order to contaminate them with noise and reverberation. The concept of amplitude modulation reduction, proposed by Houtgast and Steeneken (1985a, 1985b), was used as a guide. In order to obtain a severe degradation of the signal, considering separate noise and reverberation disturbances, a modulation reduction factor of $m = 0.1$ was proposed, corresponding to a Speech Transmission Index (STI) = 10%, which is qualitatively associated with bad speech intelligibility (Houtgast & Steeneken, 1985a, 1985b). The modulation reduction factor in the presence of noise and reverberation can be estimated as follows:

$$m(F) = \frac{1}{\sqrt{1 + \left[2\pi F \frac{T}{13.8} \right]^2}} \times \frac{1}{1 + 10^{\frac{-SNR}{10}}} \quad (1)$$

where F is the modulation frequency in hertz, T the reverberation time in seconds, and SNR the signal-to-noise ratio in decibels. Assuming a modulation frequency of $F = 2$ Hz, consistent with the typical pace of speech production in our recordings (of about two syllables per second), the proposed modulation reduction factor of $m = 0.1$ is then obtained approximately, according to Equation (1), either with a reverberation time of $T = 10$ s, or with a $SNR = -10$ dB.

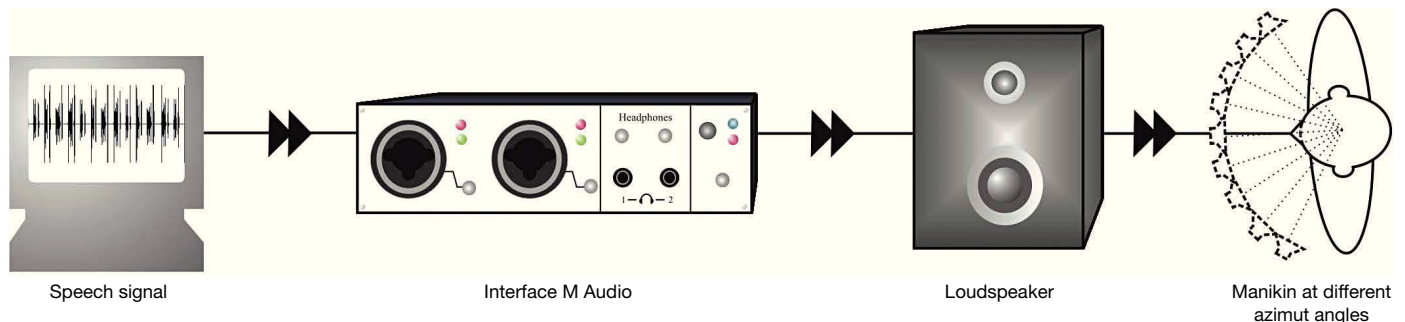


Fig. 3. Experimental setup for binaural recording at different azimuth angles from -60° to $+60^\circ$ in steps of 15° .

3.2. Speech Disturbed With Reverberation

In order to obtain a reverberated speech signal, recorded speech was convolved with an artificial reverberant impulse response, generated as follows:

$$h(t) = h_0 u(t) \exp\left(-\frac{6.9t}{T}\right); \quad (2)$$

where T is the reverberation time, $u(t)$ is one instance of a random signal with a uniform distribution, zero mean, and unit variance, and h_0 is a scale factor. The reverberation time was set to $T = 10$ s. The reverberated speech signals were obtained by convolution, as follows:

$$S_{LT} = S_L(t) * h_L(t) \quad (3)$$

$$S_{RT} = S_R(t) * h_R(t) \quad (4)$$

where $S_L(t)$, $S_R(t)$ are the original (clean) binaural signals recorded at the left and right ears of the acoustic manikin; $h_L(t)$, $h_R(t)$ are the reverberant impulse responses, which are equal in this case (correlated reverberation), but could be different (uncorrelated reverberation), at the left and right ears.

The use of Equation (2) to generate artificial room impulse responses can be justified as follows. There are a number of computational acoustics methods that can generate room impulse responses (Kuttruff, 2009) which are, still artificial, but more realistic in their basic properties. However, we are interested in impulse responses with acoustics characteristics which are not particular to any given room, but which can be generalized to a broad class of rooms, whose main characteristic is reverberation time only. In this sense, Equation (2) randomly generates room impulse responses with no particular early reflection pattern, while always ensuring an exponential amplitude decay with the prescribed reverberation time.

3.3. Speech Disturbed With Noise

In order to contaminate the speech signals, white noise was added in the two channels of the clean speech recording. The signal-to-noise ratio was set to $SNR = -10$ dB.

$$S_{LN} = S_L(t) + N_L(t) \quad (5)$$

$$S_{RN} = S_R(t) + N_R(t) \quad (6)$$

where $N_L(t)$, $N_R(t)$ are random noise signals with uniform distribution, zero mean, and scaled to the specified SNR.

4. Procedure

Intelligibility tests were carried out inside an anechoic chamber. Speech material was played back via circum-aural headphones (SONY MDR-SA1000) in order to convey proper binaural spatialization cues, and to avoid acoustic channel

crosstalk which is common in rendering two-channel sound through loudspeakers. The subject's task was to write down the words they listened, onto a paper sheet; possible spelling mistakes were ignored (e.g. "jestor"/"gestor") and in the case of homophones words in Spanish (e.g. "hacia"/"Asia") both options were taken as correct. In order to allow subjects enough time to write the words, a silent pause of 3 seconds was included after each word. Tests were split in three sessions of 50 minutes each, in order to avoid fatigue and a possible word memorization effect (in spite of the steps taken to avoid immediate repetition of the word lists, as mentioned below).

The sound pressure level of clean speech samples at 0° presented to the subjects was also measured by playback through the same headphones on the head and torso simulator, obtaining a measured presentation level of $L_{eq} = 70.0$ dB. Without modifying the presentation gain, speech samples contaminated with noise were measured in the same way, obtaining an increased presentation level of $L_{eq} = 80.0$ dB, consistent with the SNR of -10 dB that was used. Speech samples contaminated with reverberation had a measured presentation level of $L_{eq} = 69.7$ dB, showing a negligibly small reduction in the presentation level (possibly due to numerical convolution with the artificial room impulse responses) in comparison with the clean speech samples.

Table 2 shows a summary of angles, and different acoustic conditions (clean, interaurally uncorrelated noise, or interaurally correlated reverberation) used in each of the three sessions of the test. Also shown in the Table are values of interaural cross-correlation (IACC) measured after binaural playback, for each combination of the word list, azimuth angle, and acoustic condition presented. The labels L1, L2, L3, and L4 refer to the different word lists used in each case.

Care was taken to avoid repeating the same word list or the same azimuth angle immediately in sequence, and to vary as much as possible the different acoustic listening conditions (clean, noise or reverberation) after each presentation.

After a first analysis of the results, it was observed that some words from the lists were consistently very difficult to understand by the majority of subjects; especially under severe reverberation disturbance. The number of these words was different in each list, leading to some inconsistent results in the comparative evaluation of speech intelligibility under different conditions which were tested with different lists of words. For that reason, some words were eliminated in some of the tests, considering for further analysis only those words that were understood by at least 33% of the subjects in each of the tests; in some cases, these lead to reduced lists of less than 50 words, as shown in Table 3.

The 33% intelligibility threshold was selected empirically, with the objective of including more than 50%, 25 words, in any of the tests. The use of slightly different lists of words in different tests is a probable source of bias; however, we assume that lists with at least 25 words (50%) are still sufficiently varied and diverse so as to yield reasonably unbiased results. This is consistent with our underlying assumption that results are independent on the use of different word lists in different tests. Table 3 provides some interesting results in that the mean number of words and standard deviation used with added reverberation,

Table 2

Presentation of 50-Word Lists in Each Session, With an Indication of the Angle Used, the Acoustic Condition (Clean, Noise, or Reverberation), and IACC Values. No Tests Were Conducted at 0° Azimuth, but Results From Our Previous Study Are Available in Table 1.

Session 1			Session 2			Session 3		
Angle (°)	Acoustic Condition	IACC	Angle (°)	Acoustic Condition	IACC	Angle (°)	Acoustic Condition	IACC
15	L4/Noise	0.2345	30	L3/Reverb	0.9085	-30	L1/Noise	0.1975
-30	L3/Reverb	0.9123	-15	L4/Noise	0.2279	45	L4/Reverb	0.8718
-45	L1/Clean	0.8301	45	L1/Clean	0.8237	30	L2/Clean	0.9052
45	L2/Noise	0.1646	60	L2/Reverb	0.7805	60	L3/Noise	0.1260
60	L4/Clean	0.7448	-60	L3/Noise	0.1228	-15	L1/Reverb	0.9380
-15	L3/Clean	0.9155	-45	L4/Noise	0.1571	-60	L4/Clean	0.7436
-60	L2/Reverb	0.7850	-30	L2/Clean	0.9050	15	L3/Clean	0.9263
30	L1/Noise	0.2034	15	L1/Reverb	0.8621	-45	L2/Reverb	0.8550

Table 3

Number of Words Used in the Analysis of the Results (Which Were Intelligible for at Least 33% of the Subjects), for Tests at Different Angles With Added Noise and Reverberation.

Angle Presentation (°)	Noise	Reverberation
	Number of words	Number of words
0	44	40
+15	48	44
+30	50	35
+45	48	45
+60	49	40
-15	50	46
-30	50	30
-45	50	46
-60	50	43

41 ± 5 words, is smaller, and shows a greater deviation, than in the case of added noise 49 ± 2 words. A possible explanation is that bi-syllable words are more affected by reverberation because of masking of the second syllable by the first one. However, these issues are not pursued further in this article, while we maintain our assumption of results being independent of the list of words, provided sufficient words are included in each.

4.1. Results

Figure 4 shows intelligibility scores obtained at different azimuth angles, intelligibility scores show a clear influence of azimuth, especially when the speech signal is presented under adverse listening conditions. Under clean speech conditions, higher intelligibility scores were obtained for all azimuth angles, as it can be expected.

Binaural speech intelligibility scores under interaurally uncorrelated noise conditions are highest (96.6%) when listening at angles of +30° and +45°, with an approximately symmetrical behavior for angles on the left hand side; while the lowest scores are obtained at 0°, with 89.7% intelligibility.

Binaural speech intelligibility scores under reverberant conditions are higher when listening at angles of -60° and -45°, 81.6% and 78.2%, respectively, also with an approximately symmetrical behavior for angles on the opposite (right hand) side; while the lowest scores were obtained at angles of -30° and +30°, 70% and 73.3% respectively.

5. Localization Experiment

Together with the speech intelligibility test, a simple localization test was also performed, this was done in order to verify that the intended lateralization of the recordings had been perceived by the subjects. After each word list presentation at a given angle, subjects were asked to approximately identify the apparent location of the speech signal when listening through the headphones. Subjects were shown a diagram as shown in Figure 5; they were given 13 different optional answers (angles) from -90° to +90° in steps of 15°. They had to identify the apparent position from which the speech signal seemed to originate.

5.1. Results

The localization test results are shown in Figure 6. There is a clear and correct distinction between right and left directions. However, on each side, answers are largely scattered, and there is a tendency to overly lateralize the perception of angle. Angles slightly off-center at the right or left tend to be extremely later-

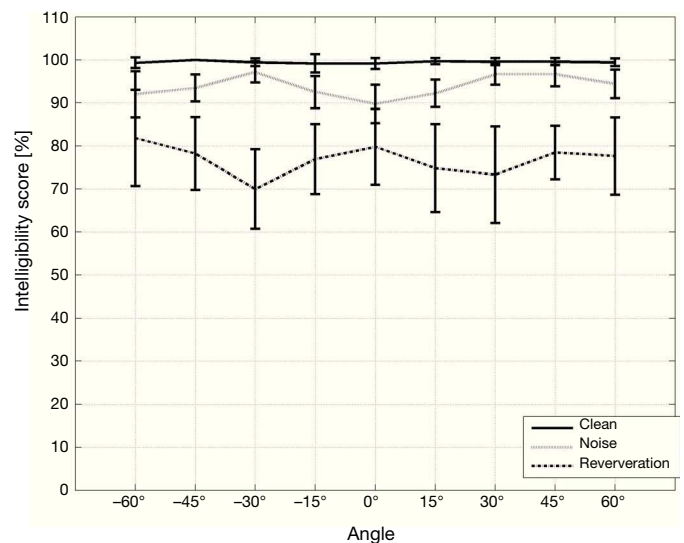


Fig. 4. Speech intelligibility score for different azimuth angles under different listening conditions, average for 15 subjects. The error bars indicate standard deviations; bars above 100% should be ignored, but are shown as a graphical representation of errors.

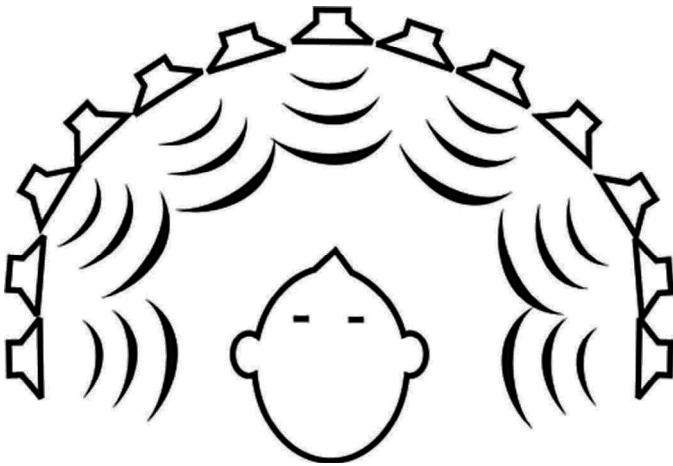


Fig. 5. Diagram presented to the subjects for them to indicate the perceived localization of the speech source during the tests, from -90° to $+90^\circ$ in steps of 15° .

alized up to the extremes $\pm 90^\circ$ of the angular scale. The angles with most correct localization responses were at $\pm 60^\circ$, while at $\pm 15^\circ$ were the least number of correct responses.

6. Discussion

These results are consistent with, and can be explained in terms of, the well established EC model of binaural speech intelligibility. For interaurally uncorrelated noise, the binaural cancellation of the noise disturbance is difficult or impossible to achieve, so that the binaural hearing mechanism is able to extract the maximum benefit for intermediate lateral angles ($\pm 30^\circ$) at which the speech signals are spectrally different (unlike at 0°), but also not extremely different in level (as they are at $\pm 60^\circ$). For correlated reverberation, a seemingly paradoxically different result holds. Under this condition, binaural hearing actually reduces speech intelligibility for intermediate lateral angles ($\pm 30^\circ$), at which reverberated speech signals are different in level, but hold the same interaural reverberant phase relationship (in our experiments), so that the binaural cancellation step would eliminate the speech signal altogether. In this case of interaurally correlated reverberation, the highest intelligibility scores are obtained in fact when no binaural advantage occurs, that is: at angles around 0° , and also around $\pm 60^\circ$, for which interaural level differences are larger, dominated by the ear more directly facing the talker, and the binaural benefit reduces to an almost monaural case.

The above results show that lateralization of the talker reduces binaural speech intelligibility under interaurally correlated reverberation. However, this condition can be expected to be rarely found in most practical situations, because based on well known statistical room acoustics principles (Kuttruff, 2009; Pierce, 1981): a strong reverberation also implies a strongly diffuse sound field in which the spatial cross-correlation of the sound pressure field, and also, in particular, interaural cross-correlation, are expected to be small at middle to high frequencies (of around and above 1 kHz). Under other condi-

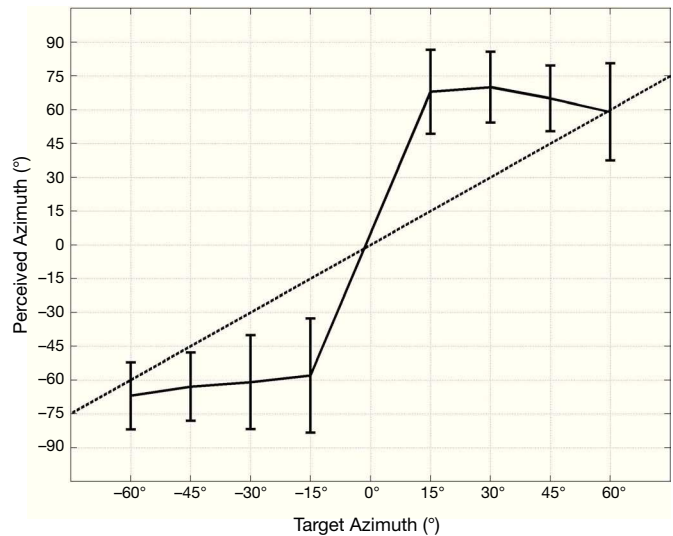


Fig. 6. Perceived localization with clean speech, from -60° to $+60^\circ$. The diagonal dotted line corresponds to a perceived angle equal to the recording angle.

tions, including: interaurally uncorrelated reverberation, and interaurally correlated or uncorrelated noise, which have been considered in this and a previous study (Padilla-Ortiz & Orduña-Bustamante, 2012), it has been shown that lateralization of the talker generally does in fact help to improve binaural intelligibility. These conditions can be expected to be more commonly found in many practical situations, so that *lateralization of the talker will generally improve binaural speech intelligibility under many acoustically disturbing conditions*.

7. Conclusions

Subjective intelligibility tests were carried out in order to investigate speech intelligibility for different azimuth angles under different adverse listening conditions (noise and reverberation), with interaurally correlated and uncorrelated disturbances. Under clean listening conditions, excellent speech intelligibility was obtained (almost 100% of correct words) for all azimuth angles; however, under noisy or reverberant listening conditions, the position of the speaker relative to the listener has a notable influence on speech intelligibility. Under interaurally uncorrelated noise, speech intelligibility slightly improves when the listener is not directly facing the talker, it shows a maximum for azimuth angles around $\pm 30^\circ$; while in contrast, under interaurally correlated reverberant conditions, binaural speech intelligibility is slightly worst around these same angles ($\pm 30^\circ$), and best at angles around 0° and $\pm 60^\circ$, at which little or no binaural advantage occurs.

Acknowledgements

The valuable suggestions and contributions of the reviewers are very kindly appreciated. Partial funding for this work has been granted by Intel Corporation, in support of research in the

Appendix: Lists of Words Used in the Binaural Speech Intelligibility Tests.

List 1:

nube	clavo	grasa	nina	ellos	suela	selva	Gestor	simple	dime
dejo	surco	neta	botes	sigo	duque	Caro	listo	persa	lenta
Críos	pleno	medios	perros	piano	mimo	Cierta	pera	toro	celda
cuales	flaca	miden	calor	choca	cita	crean	cifra	deme	tiendas
beca	torno	reno	duna	llenos	diosa	Une	prima	veinte	nada

List 2:

cedros	pluma	nombre	manto	perla	jalan	miope	laca	talco	dique
gose	fierro	cumbres	cebra	cena	nuca	Puse	niña	seda	lista
meta	lina	disco	timbre	celo	noche	Viena	himno	conde	seso
sella	pista	ciega	dije	tira	codo	cero	cera	tiro	sigo
duda	duelo	trenza	norte	lince	nená	pacto	halla	saco	cura

List 3:

Corea	pura	canción	busto	hacia	damas	ciega	tina	fino	padre
dulce	vino	reto	turco	cama	pajes	libre	vienen	cielo	pardo
siglo	níquel	tecleo	viernes	lloro	cabe	deme	regla	necio	onda
yeso	monte	tarde	quepa	luces	Carmen	mismo	suela	dota	nadie
taches	Nilo	flanes	celtas	premios	corta	surco	dardo	trance	pica

List 4:

radio	algún	unos	Diego	pili	brazo	senda	control	dante	buque
mide	cano	reto	seda	calle	sebo	clame	tape	dones	queso
sello	dieta	freno	jaque	lila	lacre	liso	Carmen	seta	esos
mini	tigre	cerca	pisen	cardo	plato	curas	lista	nave	siete
Irma	hombro	sudo	prensa	hambre	tapia	sones	feria	nulo	asno

area of “Binaural sound technologies for mobile communication devices” at Universidad Nacional Autónoma de México (UNAM). Participation of author Ana Laura Padilla Ortiz has been additionally supported by a student grant from CEP-UNAM. The authors wish to acknowledge to everyone who participated in the tests.

References

- ANSI/ASA S3.2-2009 Method for Measuring the Intelligibility of Speech over Communication System.
- Begault, D.R., & Wenzel, E.M. (1992). Techniques and applications for binaural sound manipulation. *The International Journal of Aviation Psychology*, 2, 1-22.
- Benitez, L., & Speaks, C. (1968). A test of speech intelligibility in the Spanish language. *International Journal of Audiology*, 7, 16-22.
- Berrueros, T., & Rodriguez, J. (1967). Determination of the phonetic percent in the Spanish language spoken in Mexico City, and the formation of PB lists of trochaic words. *International Journal of Audiology*, 6, 211-216.
- Beutelmann, R., & Brand, T. (2006). Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 120, 331-342.
- Beutelmann, R., Brand, T., Kollmeier, B. (2010). Revision, extension, and evaluation of a binaural speech intelligibility model. *The Journal of the Acoustical Society of America*, 127, 2479-2497.
- Cancel, C.A. (1965). Multiple-choice intelligibility lists for Spanish speech audiometry. *International Journal of Audiology*, 4, 91-93.
- Castañeda, G.R., & Pérez, R.S. (1991). Análisis fonético de las listas de palabras de uso más extendido en logaudiometría. *Anales de la Sociedad Mexicana de Otorrinolaringología*, 1, 23-30.
- Durlach, N.I. (1963). Equalization and cancellation theory of binaural masking-level differences. *The Journal of the Acoustical Society of America*, 35, 1206-1218.
- Ferrer, O. (1960). Speech audiometry: a discrimination test for Spanish language. *The Laryngoscope*, 70, 1541-1551.
- Flanagan, J.L. (1960). Analog measurements of sound radiation from the mouth. *The Journal of the Acoustical Society of America*, 32, 1613-1620.
- Houtgast, T., & Steeneken, H.J.M. (1985a). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, 77, 1069-1077.
- Houtgast, T., & Steeneken, H.J.M. (1985b). *Technical Review No. 3, The Modulation Transfer Function in room acoustics* (pp. 1-44). Marlborough, MA: Brüel & Kjaer Instruments.
- ISO/TR 4870:1991 Acoustics — The construction and calibration of speech intelligibility tests.
- Jelfs, S., Culling, J.F., & Lavandier, M. (2011). Revision and validation of a binaural model for speech intelligibility in noise. *Hearing Research*, 275, 96-104.
- Kanada, Y. (2004). *Multi-context voice communication controlled by using an auditory virtual space* (pp. 1-6). 2nd IASTED International Conference on Communication and Computer Networks (CCN), Nov. 8-10.
- Kanada, Y. (2005). Simulated virtual market place using voicemap communication medium. In *Proceedings of the 13th annual ACM international conference on Multimedia*, 794-795.
- Kuttruff, H. (2009). *Room Acoustics* (5th Ed.). London: Spon Press.
- Lavandier, M., & Culling, J.F. (2010). Prediction of binaural speech intelligibility against noise in rooms. *The Journal of the Acoustical Society of America*, 127, 387-399.
- Loomis, J., Hebert, C., & Cicinelli, J. (1990). Active localization of virtual sounds. *The Journal of the Acoustical Society of America*, 88, 1757-1764.
- Loomis, J., Golledge, R., Klatzky, R., Speigle, J., & Tietz, J. (1994). Personal guidance system for the visually impaired. In *Proceedings of the first annual ACM conference on Assistive technologies*, 85-90.
- Loomis, J., Golledge, R., & Klatzky, R. (1998). Navigation system for the blind: Auditory display modes and guidance. *Presence: Teleoperators and Virtual Environments*, 7, 193-203.
- Nordlund, B. (1962). Physical factors in angular localization. *Acta Otolaryngologica*, 54, 75-93.

- Nordlund, B., & Fritzell, B. (1963). The influence of azimuth on speech signals. *Acta Otolaryngologica*, 56, 632–642.
- Nordlund, B., & Lidén, G. (1963). An artificial head. *Acta Otolaryngologica*, 56, 493–499.
- Padilla-Ortiz, L., & Orduña-Bustamante, F. (2012). Binaural speech intelligibility and interaural cross-correlation under disturbing noise and reverberation. *Journal of Applied Research and Technology*, 10, 347–360.
- Plomp, R., & Mimpen, M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the Speech-Reception Threshold for sentences. *Acta Acustica United with Acustica*, 48, 325–328.
- Pierce, A.D. (1981). *Acoustics: An Introduction to Its Physical Principles and Applications*. New York: McGraw-Hill Book Company.
- Steeneken, H.J.M., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America*, 67, 318–326.
- Tato, J.M. (1949). *Lecciones de Audiometría*. Buenos Aires, Argentina: El Ateneo.
- Van Wijngaarden, S.J., & Drullman, S. (2008). Binaural intelligibility prediction based in the speech transmission index. *The Journal of the Acoustical Society of America*, 123, 4514–4523.
- Zubick, H., Irizarry, L.M., Rosen, L., Feudo, P., Kelly, J.H., & Strome, M. (1983). Development of speech-audiometric materials for native Spanish-speaking adults. *International Journal of Audiology*, 22, 88–102.