Let's all speak together! Exploring the impact of various languages on the comprehension of speech in multi-linguistic babble

Aurore Gautreau [1,2,3], *Michel Hoen* [1,2,3], *Fanny Meunier* [1,2,3]

[1] INSERM U1028, Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, Lyon, F-69000, France
[2] CNRS UMR5292, Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, Lyon, F-69000, France
[3] University Lyon 1, Lyon, F-69000, France

aurore.gautreau@gmail.com, fanny.meunier@ish-lyon.cnrs.fr

10.21437/Interspeech.2011-11

Abstract

Our research aims at exploring psycholinguistic processes implicated in the speech-in-speech situation. Our studies focused on the interferences observed during speech-in-speech comprehension. Our goal is to clarify if interferences exist only on an acoustical level or if there are clear psycholinguistic interferences. In 3 experiments, we used 4 talkers cocktail-party signals using different world languages: French, Breton, Irish and Italian. Participants had to identify French words inserted in a babble. Results first confirmed that it is more difficult to understand a French word in a French background than in a babble composed of unknown languages. This result demonstrates that the interference effect is not purely acoustic but rather linguistic. Results also showed differences in the observed performances depending on the unknown language spoken in the background and demonstrated that some languages interfered more with French than some others.

Index Terms: babble noise, speech comprehension, interlanguages

1. Introduction

Although we can very easily understand speech delivered through headphones in a sound-proof laboratory environment, we are more usually confronted with speech occurring in the acoustic chaos of traffic-jam or within the noisy chatter of multiple talkers in crammed meeting rooms. While for many years language comprehension has been studied by psycho- and neurolinguists in quiet experimental booths, a recently grown area of research started to investigate what the situation of speech-in-noise comprehension can teach us on the specific cognitive processes involved. Indeed, since the seminal work by Cherry [1], who first coined the terms "cocktail-party phenomenon", an impressive number of experiments have focused on the psychoacoustic aspects of this effect and explored the processes involved in auditory stream segregation ([2]; [3]). However, one constant limitation of these studies is that they tend to consider speech as just 'another type of sound', while speech in fact is a very particular type of sound for human ears and brains and therefore the particular question of speech comprehension in concurrent noise cannot be addressed exclusively from a psychoacoustic perspective. Another domain that has been interested by the speech-in-noise comprehension challenge is audiology, but again the main interest is not language comprehension but rather hearing and hearing deficits. In our research, we focused on psycholinguistic processes implicated during speech-in-noise comprehension.

Previous studies on speech-in-noise comprehension have been done comparing different types of noise and of particular interest non-linguistic noise vs. speech. It has already been shown that it is easier to understand a word when it is presented in non-linguistic noise rather than in concurrent speech. For example Simpson and Cooke [4] found lower speech intelligibility scores in natural babble than in babble-modulated noise when there were more than two talkers in the background [5]. Recently our group examined the performance on word comprehension depending on the background used: babble or babble-like noise (more precisely time-reversed babble). Our results revealed that at a low talker number (N = 4), performances decreased in the natural babble compared to the reversed babble condition, suggesting increased linguistic interference [6]. These results clearly suggest that linguistic interference plays a role in the effects of natural multi-talker babble on target speech perception.

The present studies aimed at exploring psycholinguistic processes implicated in the particular situation of speech-in-speech comprehension. This situation allows tackling the issue of informational masking. Informational masking has been formerly characterized in the field of psychoacoustics as a competition between information originating from a target sound and comparable information present in background noise that interfere during their processing ([7]; [8]; [9]; [10]). In the particular context of speech-in-speech comprehension, this dimension becomes extremely relevant. Indeed, speech-in-speech situations are of primary interest for psycholinguists in order to investigate the competitions occurring between potential candidates within the various processing levels (acoustic, phonetic, phonologic, lexical and semantic) during word identification, as it offers a natural example where competitions may be directly tested through behavioural measures in controlled experimental settings.

To look at the nature of the interferences that come into play during speech-in-speech comprehension, a number of recent studies have manipulated the language used as background noise. They demonstrated that the language of the masker can affect the intelligibility of a target sentence ([11]; [12]; [13]). Rhebergen et al. [11] found worse speech reception thresholds for Dutch sentences presented in competing Dutch speech than when the background speech was in Swedish. Using a consonant in vowel-context identification task, Garcia Lecumberri and Cooke [12] showed that English stimuli were better recognized when the language in the background was Spanish rather than English. Van Engen and Bradlow [13] demonstrated that English sentence intelligibility was better when the noise consisted of two-talker Mandarin Chinese babble than when it was composed of two-talker English babble. These results suggest that a masking talker using the same language as the target stimuli increases

Copyright © 2011 ISCA

informational masking relative to one using a different unknown language.

1.1. The present study

In our studies we manipulated different languages used in the background as competitors in order to examine differences in masking effects whose origin could be from different linguistic characteristics between these languages. While the targets were always French words, four languages, with different linguistic characteristics have been tested as competitors: French, Breton, Irish and Italian. These languages were chosen for their various linguistic characteristics such as rhythmic class and phonological inventory. In the series of experiments we created cocktail-party signals using these four different languages.

As Brungart et al. [7] showed that with up to 3-talkers, masking is mostly due to the processing of pitch information that possibly interacts with or even covers masking effects due to other psycholinguistic dimensions of speech, we decided to study cases in which individual voice characteristics are less predominant, i.e. situations with 4 talkers.

We contrasted situations where the babble was made-up of French speech and therefore contained understandable words vs. situations in which babble was made-up of languages unknown to our participants. In the first experiment the foreign language used was Breton, in the second Irish and in the third Italian.

Non understandable languages were thus considered in the experiment as speech sounds containing different type of phonetic and phonologic information but no (or very little) lexical information. These different background noises (French1, Breton1, French2, Irish2, French3 and Italian3) were all tested at 2 different SNRs of 0 and -5 dB. Participants had to identify a French word inserted in the babble.

2. Experiments

The three experiments were conducted with very similar methods and designs.

2.1. Materials and methods

2.1.1. Participants

Forty volunteers participated in the first experiment, forty in the second one and thirty-six in the third. Volunteers did not participate to more than one experiment of these series. They were all students, aged 18–32 years and native French speakers with no known hearing or language disorders. They were paid for their participation.

2.1.2. Multitalker babble sounds

Two female voices and 2 male voices were mixed to create the babble signals. Each voice was first recorded separately in a sound-proof room, reading extracts from books (the same book translate in different languages). Individual recordings were modified according to the following protocol: (i) removal of silences and pauses of more than 1 s, (ii) suppression of sentences containing pronunciation errors, exaggerated prosody or proper nouns, (iii) noise reduction optimized for speech signals, (iv) intensity calibration in dB-A and normalization of each source at 80 dB-A and (v) final mixing of individual sources into cocktail party sound tracks.

2.1.3. Target words

One hundred thirty-six French disyllabic words were recorded in a sound-proof booth by a female native French speaker aged of 28 years old. Words were selected in a middle range of frequency of occurrence (ranging from 0.2 to 566.5 per million; mean = 30.58, SD = 74.87), according to the French database Lexique2 [14], in order to avoid extremely high- or low-frequency items that volunteers typically either overuse or ignore.

2.1.4. Stimuli and word lists

Stimuli consisted of the 136 single target-words mixed together with 4 s of background noise. Target-words were always inserted 2.5 s from the start of the stimulus, so that participants always had the same exposure to the background noise before the target-word was presented. Stimuli were composed by mixing one chunk of background noise, randomly selected from 40 chunks extracted from the original noise files, with one target word. Individual intensity levels for background noise and target words were adjusted according to the global rms power of the original sounds to be mixed. As this resulted in some intensity modulation of the final stimuli and in order to avoid global intensity to become predictive of the SNR, a final randomized intensity roving over a ±3 dB range in 1 dB steps was applied to each created stimulus. We created 4 experimental lists per experiment. Each list contained each target word only once to avoid priming effects. Across lists and experiments, all target words were presented against the 4 languages and at the 2 SNRs (structure: within items and within subjects). Within lists, target words were balanced for frequency and phonological neighbourhood across conditions and order of presentation was randomized.

2.1.5. Procedure

Participants sat in a quiet room, facing a computer monitor. Stimuli were delivered diotically via headphones (Beyerdynamic DT 48, 200 Ω) at 60 dB SPL (comfortable sound level). The task for participants consisted in a single-word transcription, participants being asked to write down the sounds they heard. Before the testing phase, participants were given 12 practice items to accommodate themselves to the stimulus presentation mode and target's voice. The experiment lasted an average of 45 min.

2.2. Results

Answers from participants were analyzed in terms of correct word identification rates by calculating the proportion of transcribed words that corresponded to target words. Spelling errors were not taken into account. These individual word identification rates were used as dependant variables in the following analyses. Raw scores were converted to percent correct and then to rationalized arcsine units (RAU). This transformation allows for valid comparisons of differences across the entire range of the scale (Studebaker, 1985). Scores on this scale range from 13.5 RAU corresponding to 0% correct to 86.5 RAU corresponding to 100% correct. A repeated-measures analysis of variance (ANOVA) was conducted on each experiment considering RAU scores as the dependent variable and Language (Known, i.e. French vs. Unknown) and SNR (0 vs. -5dB) as within-subjects factors.

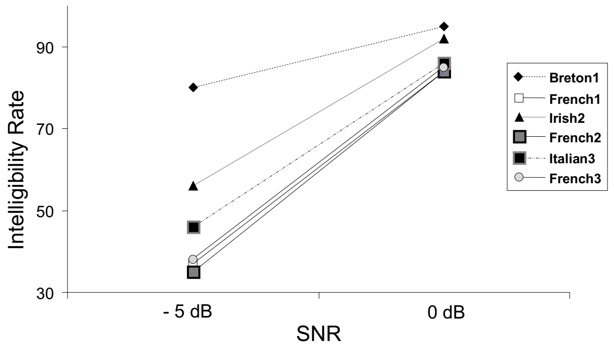

Figure 1: *Results for the 3 experiments, showing the effects of languages used in the background depending of the SNR (-5dB and 0dB). French1 corresponded to the French condition in Experiment 1 (with Breton), French2 to the French condition in Experiment 2 (with Irish), French3 to the French condition in Experiment 3 (with Italian). Results for these identical conditions are very similar over the 3 experiments.*

2.2.1. *Experiment 1 – Breton*

Results showed a significant effect of the language in the babble: French words were recognized 27% better when babble was made of Breton compared to French ($F(1,39) = 408{,}31$; $p < .001$). With Breton in the babble, scores reached 87% (SD = 10) while the score was of 60% (SD = 13) with French babble. Results also showed a significant 31% effect of the SNR ($F(1,39) = 514{,}85$; $p < .001$). In average, performances were lower at -5dB (58%; SD = 13) compared to 0dB (89%; SD = 9). We also observed a significant interaction between the two factors ($F(1,39) = 193{,}98$; $p < .001$), with a bigger language effect at -5dB SNR (42%) compared to the one observed at 0dB SNR (12%). Post-hoc comparisons done with the HSD Tukey test revealed that the language effect was still significant at both SNRs (-5dB - $p = .0001$ and 0dB - $p = .0001$).

We therefore observed the expected effect of SNR, an effect of the language in the background and an interaction of the two, showing that it is even more difficult to understand French word in a French babble compared to Breton babble when the SNR is of -5dB compared to 0dB.

2.2.2. *Experiment 2 – Irish*

Results showed a clear effect of the language used in the background of 16% that was significant ($F(1,39) = 202{,}59$; $p < .001$). On average participants were better when the babble was made of Irish (75%; SD = 10) than made of French (59%; SD = 12). We also found a significant effect of SNR of 39% ($F(1, 39) = 875{,}98$; $p < .001$). In average scores were lower for a SNR of -5dB (48%; SD = 13) than for one of 0dB (87%; SD = 8). We also observed a significant interaction between the languages used in the babble and the SNR ($F(1,39) = 22{,}23$; $p < .001$). As showed on Figure 1, the effect of Language was more important at -5 dB (21%) than at 0 dB (10%). Post-hoc comparisons done with the Tukey HSD test showed that the language effect was significantly different for both SNR (-5dB - $p = .0001$; 0dB - $p = .0001$).

As in the previous experiment with Breton, in this experiment using Irish in the background, we observed the expected effect of SNR, an effect of the language in the background and an interaction of the two, showing that it is even more difficult to understand French word in a French babble compared to Irish babble when the SNR is of -5dB compared to 0dB.

2.2.3. *Experiment 3 – Italian*

Results showed a significant effect of the language in the babble: French words were recognized 4% better when babble was made of Italian compared to French ($F(1,35) = 8{,}993$; $p < .005$). With Italian in the babble, scores reached 66% (SD = 10) and 62% (SD= 13) in French babble. Results also showed a significant 43% effect of the SNR ($F(1,35) = 964.911$; $p < .001$). On average performances were lower at -5dB (42%) compared to 0dB (85%). We also observed a significant interaction between the two factors ($F(1,35) = 5{,}753$; $p < .05$), with a larger language effect at -5dB SNR (8%) compared to a 0dB SNR (1%). Post-hoc comparisons performed with the HSD Tukey test revealed that the language effect was only significant at a SNR of -5dB ($p = .0001$).

As for the previous experiments we observed the expected effect of SNR, an effect of the language in the background and an interaction of the two, showing that it is more difficult to understand French words in a French- compared to Italian babble when the SNR is of -5dB; however when the SNR is of 0dB, no difference is observed between the French and the Italian babble conditions.

2.2.4. *Comparison of the 3 experiments*

In order to compare the scores obtained for the 4 languages used in babble noise (French, Breton, Italian and Irish), a repeated-measures analysis of variance (ANOVA) was conducted on all experiments considering RAU scores as dependent variable. Language (Known, i.e. French vs. Unknown) and SNR (0 vs. -5dB) were declared as within-subjects factors and Identity of the language in the babble (Breton-exp1, Irish-exp2 and Italian-exp3) as between-subjects factor.

Results showed an effect of Language ($F(1,113) = 251.415$, $p<.0001$), overall it was easier to understand a French word in a babble made of unknown languages than in a French babble. There was a main effect of SNR ($F(1,113) = 1358{,}396$, $p<.0001$) and a main effect of the Identity of the unknown language ($F(2,113) = 23.802$, $p<.0001$). Also all interactions were significant: Language*Identity ($F(2,113) = 47.164$, $p<.0001$), SNR*Identity ($F(2,113) = 13.695$, $p<.0001$), SNR*Language ($F(1,113) = 100.324$, $p<.0001$), SNR*Identity ($F(2,113) = 13.695$, $p<.0001$) and the 3rd-level interaction, Language*SNR*Identity ($F(2,113) = 18.879$, $p<.0001$).

To go further we first checked that the French babble conditions were not significantly different over the 3 experiments and then we looked at the differences between Breton, Italian and Irish. We first ran a repeated-measures analysis of variance (ANOVA) including only the scores obtained with French babble with SNR (0 vs. -5dB) as within-subjects factor and Identity of the language (Breton-exp1, Irish-exp2 and Italian-exp3) as between-subjects factor. Only SNR showed a significant effect ($F(1,113) = 1220.968$, $p<.0001$; the other two Fs<1, n.s.). French conditions indeed gave very similar scores over the 3 experiments as can be seen on Figure 1. Another analysis including the same factors have been run on the unknown languages scores. It showed an effect of SNR ($F(1,113) = 508.031$, $p<.0001$), an effect of the Identity of the language in the babble ($F(2,113) = 83.019$, $p<.0001$) and an interaction between the 2 factors ($F(2,113) = 31.236$, $p<.0001$). It appears that the effect of the Identity of the language is significant at -5dB ($F(2,113) = 60.098$, $p<.0001$) and at 0dB ($F(2,113) = 19.928$, $p<.0001$). Specific comparisons showed that all comparisons between Italian, Irish and Breton are significant at each SNR ($p<.05$). So it is easier to understand French words in Breton, followed by Irish and then Italian.

This means that overall the 3 unknown languages led to different performances, which suggests that the different characteristics of these languages are playing a role during word-comprehension in this challenging situation.

3. Discussion and Conclusion

In this paper we were interested in speech-in-speech comprehension and more particularly in determining if linguistic characteristics of the language presented in the background interfere with word identification.

First we wanted to see if, as it was shown in English and in Dutch, it is easier to identify a word when the language in the background is different from the language of the target. Clearly our results showed that it is harder to understand a French word in a French babble than in any other language babbles. This result demonstrates that the interference effect is not purely acoustic but rather linguistic. This effect could reflect the fact that in the case of 4-talkers babble some words in the background are still activated and therefore compete with the identification of the target word. To further investigate this effect, we tested if we found differences on French word identification while manipulating the language in the background. We choose three languages that were unknown to our participants: Breton, Irish and Italian. Our results showed very clearly that the three unknown languages interfere differently with the French target-words. The less masking language is Breton, followed by Irish, then Italian (with the harder language being French).

To understand the differential effects depending on the language of the background we looked at the characteristics of each language. One characteristic that could be of interest is phonologic inventory, as competition could happen between sounds of languages. It may happen that the overlap between phonemes of the languages used in the background and French could explain our results. Another characteristic that could come into play is speech-rate as it could be that some languages are spoken slower, leaving more gaps to listen into them. The last characteristic we looked at is fundamental frequency (F0) values, as it has been shown that this parameter helps to segregate speech-in-speech.

French is a Romance language spoken mainly in Europe; it has 16 vowels and 21 consonants. The female voices in the babble had an average F0 of 227 Hz and the speech-rate was on average of 5.11 syllables per second. Breton is an insular Celtic language (as Welsh and Cornish) spoken in French Brittany, with 13 vowels and 25 consonants and 7 semi-consonants. More importantly it has 7 vowels and 19 consonants in common with French. The female voices in the babble had an average F0 of 226 Hz and the speech-rate was on average of 4.54 syllables per second. Irish is a Goidelic language of the Indo-European language family, historically spoken by the Irish people. It has 20 vowels, 45 consonants and 4 semi-consonants. It has 4 vowels and 9 consonants in common with French. The female voices in the babble had an average F0 of 212 Hz and the speech-rate was on average of 4.60 syllables per second. Italian is a Romance language derives diachronically from Latin. It has 7 vowels, 23 consonants and 7 vowels and 11 consonants in common with French. The 2 female voices in the babble had an average F0 of 206 Hz and the speech-rate was on average of 5.29 syllables per second. Globally our results don't appear to be explained by phonological distances between French and the other languages. However, for French, Italian and Irish, performances followed the phonological proximity of the languages with French (number of phonemes in common). In this framework, it is puzzling that Breton gave the best performances. This could be explained by the fact that speakers of Breton, even if highly bilinguals, are native of French. A hyperarticulation may result in their productions and facilitates the segregation between target word and babble. Further analyses need to be done to clarify this point.

However our results clearly demonstrated that the interference observed in a speech-in-speech situation is not purely acoustic but rather linguistic and moreover language-specific. The speech-in-speech situation seems ideal to look at interferences occurring during spoken word comprehension. More work is however needed to explain these differences and identify the characteristics that come into play in this situation.

4. Acknowledgement

This research was supported by a European Research Council grant to the SpiN project N°209234, and by the French “Direction Générale de l'Armement”.

5. References

[1] Cherry, E. C., "Some experiments on the recognition of speech with one, and with two ears", *JASA,* 25, 975-979, 1953.

[2] Bregman, A. S., "L'analyse de la scène auditive: l'audition dans des environnements complexes" in *S.E. McAdams & E. Bigand, (Eds.) Penser les sons: Psychologie cognitive de l'audition*. Paris: Presses Universitaires de France, 10-39, 1994.

[3] Bronkhorst, A., "The cocktail party phenomenon: A review of research on, speech intelligibility in multiple-talker conditions" *Acustica*, 86, 117-128, 2000.

[4] Simpson, S. A. & Cooke, M., "Consonant identification in N-talker babble is nonmonotonic function of N. ", *JASA*, 118, 2775-2779, 2005.

[5] Sperry, J. L., Wiley, T. L. & Chial, M. R, "Word recognition performance in various background competitors", *J. Am. Acad. Audiol.*, 8, 71-80, 1997.

[6] Hoen, M. & al., "Phonetic and lexical interferences in informational masking during speech-in-speech comprehension", *Speech Comm.,* 49, 905–916, 2007.

[7] Brungart, D. S., "Informational and energetic masking effects in the perception of two simultaneous talkers", *JASA*, 109, 1101-1109, 2001.

[8] Brungart, D.S & al., "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation", *JASA*, 120, 4007-4018, 2006.

[9] Dirks, D. D., Bower, D., "Masking effects of speech competing messages", *J. Speech Hear. Res*., 12, 229-245, 1969.

[10] Festen, J., Plomp, R., "Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing", *JASA*, 88, 1725-1736, 1990.

[11] Rhebergen, K. S., & al., "Release from informational masking by time reversal of native and non-native interfering speech (L)", *JASA*, 118, 1274-1277, 2005.

[12] Garcia Lecumberri, M. L., Cooke, M., "Effect of masker type on native and non-native consonant perception in noise", *JASA*, 119, 2445-2454, 2006.

[13] Van Engen, K. J., Bradlow, A. R., "Sentence recognition in native and foreign-language multi-talker background noise", JASA, 121, 519-526, 2007.

[14] New B. & al., "Lexique 2 : A new French lexical database", Behavior Research Methods, Instruments, & Computers, 36, 516-524, 2004.