THE CONTRIBUTION OF THE VISUAL MODALITY TO SPEECH INTELLIGIBILITY IN NATIVE AND NON-NATIVE SPEAKERS OF ENGLISH

MA ADVANCED ENGLISH STUDIES
LINGUISTICS TRACK

SINÉAD MARIE CECILIA RANKIN

SUPERVISOR: PROF. MARIA-JOSEP SOLÉ SABATER

UAB
Universitat Autònoma de Barcelona

Acknowledgements

I am deeply grateful to my supervisor, Professor Maria-Josep Solé, for her invaluable guidance and support throughout the process of writing this Master's thesis. Her expertise, dedication to the field and enthusiasm for my ideas have been instrumental in shaping my research and academic development. My appreciation also extends to the participants who generously volunteered their time and efforts to take part in the research study. Without their active involvement, this thesis would not have been possible.

I also want to thank my teachers and colleagues on the linguistics track who provided me with valuable feedback and suggestions during the various stages of this research. Their insights and discussions have greatly enriched my interest and understanding of the subject matter.

Finally, I am profoundly grateful to my family and Len for their unwavering support and their belief in my abilities that made it possible for me to pursue this course of study. I am indebted to them for their love, sacrifice, and willingness to lend an ear.

Table of contents

Index of Figures

Abstract

Since the pandemic, masks have been known to reduce speech intelligibility due to obscuring visual cues. This can present a communication problem, especially when involving foreign-accented speech. Research dating as far back as Sumby and Pollack (1953) has analysed the visual contribution to speech intelligibility, finding that listeners rely more on the visual cue under conditions of noise disturbance. Past research has predominantly focused on the role of visual cues in consonant perception. This study assesses the contribution of the visual modality to speech intelligibility, with a specific focus on vowels. It compares the effectiveness of visual cues provided by a native English speaker and a French L1 non-native English speaker. The study uses audio and audio-visual stimuli, involves native English perceivers (n = 24), and employs an orthographic vowel intelligibility test. The results demonstrate a significant audio-visual benefit, with improvements observed across both speaker groups. However, the degree of visual modality effectiveness varies across different vowel features and between speaker groups, with the central vowel /ʌ/ showing a negative visual impact when provided by the French speaker group, as well as /ə/, with characteristic French lip-rounding. This highlights the influence of language-specific gestures on L2 production. These findings provide insight into the various phonological challenges faced by non-native English speakers when it comes to producing sounds accurately and highlights the importance of the visual cue for speech perception. The results of this research have implications for any instructional strategies that may be used to assist non-natives in pronunciation and for the development of more effective speech perception strategies, as well as for research in the diagnosis and treatment of speech and language disorders.

Keywords: speech intelligibility, visual modality, audio-visual speech perception, visual cues, vowel recognition, foreign-accented speech

1. Introduction

Speech intelligibility is defined as the ability to recognise and understand spoken language. While holding other considerations equal, this is impacted by a variety of factors, such as the acoustics of the environment, the listener's hearing ability, and the presence or absence of visual cues (Munro, 1995). During the COVID-19 pandemic, communication came into the spotlight for a number of reasons. The widespread adoption of face masks that cover the mouth and jaw region made some interactions difficult. In addition to muffling sounds, speech intelligibility was reduced due to the obscuring of visual cues (Giuliani, 2020). So, while it is widely accepted that hearing impaired individuals rely on visual cues to aid in speech comprehension, the question arises of whether normal-hearing individuals may also unconsciously utilise visual information more than previously recognised (Rosenblum, 2005). The use of these cues is often referred to as lip-reading, and although these lip and mouth movements may not always be crucial, they become increasingly valuable in situations where communication is hindered, such as in the presence of loud background noise and when conversing in a foreign language.

Foreign accented speech can make communication between native and non-native speakers challenging. The degree of the foreign accent in non-native speakers can vary, with some accents being easier to understand than others (Flege et al., 1995; Rogers, 2004). In quiet environments, these accents do not usually present an issue (Bent & Bradlow, 2003). However, in a noisy environment, foreign accented speech may be difficult to perceive, with the visual modality becoming more relied upon as background noise increases (Hazan et al., 2006). Along similar lines, it is not uncommon to hear of the discomfort

experienced during phone conversations between native and non-native speakers, which is understandable given the lack of visual information and the possibility of noise interference. So, in challenging environments, both native and non-native English listeners with normal hearing may benefit from visual cues. The present study is interested in exploring the effectiveness of non-native visual cues in noisy environments and the unique visual characteristics that distinguish native and non-native English speakers. Specifically, the study focuses on examining whether listeners show improved accuracy in recognition of non-native speech when visual cues are presented, and comparing the effectiveness of these cues to those provided by native English speakers.

There are many language backgrounds that could provide interesting results from a study about the effectiveness of visual cues, particularly when considering previous studies on the notable differences between English and Asian languages like Mandarin. However, European languages are rarely studied, in spite of the fact that English is the common language of the European Union, prompting the current study to concentrate on this context. Specifically, it investigates L1 French speakers of English. Additionally, whilst most related studies on audio-visual L2 perception focus on consonant visemes (Hazan et al., 2002; Kawase & Wang, 2014; Sennema et al., 2003), the present study focuses on English vowels. Non-natives may produce different visemes when speaking a second language due to the influence of their native language's phonetic and articulatory features. This can result in variations in mouth shapes and movements when compared to native speakers of the target language. For example, in French, lip-rounding is the main feature distinguishing phonological contrasts in front vowels, whereas, in English, lip-rounding is a secondary feature

of back vowels, but not contrastive. Also, French speakers tend to exhibit lip protrusion accompanying French rounded vowels, but this feature is less pronounced in English rounded vowels (Zerling, 1992). Therefore, the present study investigates the potential influence of these unique, language-specific features on intelligibility, especially in contexts where the visual modality is most utilised. Since visual cues are known to be useful in L2 learning (McGuire & Babel, 2012; Sekiyama et al., 1996), this study aims to provide insight into the various phonological challenges faced by non-native English speakers when it comes to producing sounds accurately and to highlight the importance of the visual cue for speech perception.

The following section will address the research aims and hypotheses, followed by a chapter that reviews the theoretical background related to the current study. This will examine existing literature on speech intelligibility in noise, the visual modality and its contribution to speech intelligibility in noise, as well as foreign-accented speech intelligibility, with a specific focus on French-accented speech and mouth movements. After this, the experiment and method will be outlined in detail. Finally, the results will be presented, accompanied by a discussion and the relevant conclusions.

1.1 Research overview and aims

This study aims to assess the contribution of the visual modality to speech intelligibility, specifically vowels, comparing the usefulness of cues provided by a Native English speaker and a French L1 non-native English speaker. This will be achieved through the use of audio and audio-visual stimuli, native English perceivers, and an orthographic vowel intelligibility test. The study posits the following hypotheses:

1. The visual modality will enhance speech intelligibility for both native and non-native English speakers, with several possible sub hypotheses:

   1.a The degree of effectiveness of the visual modality will be *the same* for both language backgrounds (native and non-native).

   1.b The contribution of the visual modality will be *greater for non-native* than for native speakers. That is, since L1 French speakers of English are expected to score lower in intelligibility than native speakers in the auditory presentation, the contribution of the visual modality may be expected to show a larger effect on intelligibility in the former (i.e., they have larger room for improvement).

   1.c The contribution of the visual modality will be *greater for native* than non-native speakers due to language-specific speech gestures.
   In order to further explore the last scenario, a second hypothesis was stated.

2. The distinctive lip movements and protrusions characteristic of French speakers may confuse perceivers in the audio-visual condition, potentially compromising speech intelligibility.

2. Theoretical Background

2.1 Speech intelligibility in noise

In 1953, Cherry discovered The Cocktail Party Effect, whereby a person is able to filter out background noises in order to focus on what is necessary, such as a single conversation. This effect enables people to communicate in a noisy environment, like a crowded room or a party. It also highlights the brain's ability to selectively process auditory information (Cherry, 1953). Since this was discovered, more studies have investigated speech intelligibility, looking at different variables. Some have experimented by manipulating the speech-to-noise ratio (SNR) (Brungart et al., 2020), whilst others have focused on factors such as the type of background noise (Assmann & Summerfield, 2004; Rogers et al., 2006), participants' hearing ability (George et al., 2006), stimulus types, and various other aspects (Bronkhorst, 2000; Summerfield, 1992). Many studies have investigated the effects of noise on the intelligibility of foreign accented speech (Melguy & Johnson, 2021; Rogers, 2004), which will be discussed further in Section 2.3.

Listeners exhibit varying degrees of proficiency in comprehending speech under adverse circumstances, with the tolerable speech-to-noise ratio being contingent upon the nature of the background noise (McLaughlin et al., 2018). Furthermore, it is widely accepted that as the ratio of signal to noise (SNR) reduces, speech intelligibility decreases (Munro, 1998). Zhao (2022) states that there is a lot of disagreement regarding the optimal SNR for speech intelligibility, drawing on Robinson & Casali's (2003) finding of approximately 12 dB, but

mentioning that other studies estimate significantly lower ratios (e.g., Shadle, 2007; as cited in Zhao, 2022). It is important to remember that type of background noise and fluctuation are also factors. For a review of speech in noise and the effects of different types of background noise, see Bronkhorst (2000).

In response to the numerous factors that contribute to creating unfavourable listening conditions, individuals utilise additional cognitive, perceptual, and linguistic abilities. Many studies have argued that the addition of the visual modality aids in speech perception. This will be discussed in the next section.

2.2 The visual contribution to speech perception

As aforementioned, previous studies have indicated that the visual modality plays a significant role in speech comprehension. This has been highlighted by the McGurk effect, a cognitive phenomenon in which conflicting audio and visual cues can cause someone to perceive a completely different sound (McGurk & MacDonald, 1976, as cited in Bicevskis et al., 2016). The McGurk effect thus showed that speech perception can be biased by altering the visual information accompanying the acoustic signal, providing support to the view that speech perception involves the use of visual as well as acoustic information.

Visual cues in speech perception can refer to the visual information obtained from observing a speaker's facial movements, lip movements, gestures, and other visible articulatory features, which assist in understanding and interpreting spoken language. In particular, facial cues such as eyebrow flashes, head nods, and beat gestures have been shown to be the visual correlates of prominence and sentence focus (see Borràs-Comes & Prieto (2011) for a review). It is believed that visual information can be used to supplement, or even replace,

auditory information in order to understand speech (Gabbay et al., 2017; Hardison, 2003; Munro, 1998). Von Raffler-Engel (1980) even argued that "eliminating the visual modality creates an unnatural condition, which strains the auditory receptors to capacity" (Von Raffler-Engel, 1980, p. 235, as cited in Sueyoshi & Hardison, 2005), with weight being added to this claim in a study examining neuromagnetic responses, which discovered that activity in the human auditory cortex is altered by visual cues (Sams et al., 1991). Rosenblum (2005) proposes that multimodal speech is not merely an additional feature reliant on auditory speech, but is indeed the primary mode of speech perception.

Although visual cues are undoubtedly valuable, it should be noted that the visual modality alone has limited use for speech perception (Stacey et al., 2016). In a study conducted by Grant et al. (1998), they found that in a visual-only condition, sentence recognition scores varied from 0% to 20%. However, audio-visual scores ranged from 23% to 94%, and from 5% to 70% for an audio-only condition. Therefore, to fully leverage the benefits of the visual modality, a significant amount of auditory information is necessary to enhance the interpretation of visual cues in speech (Grant et al., 1998).

In their classic study, Sumby and Pollack (1954) found that the visual modality is relied upon more under conditions of noise disturbance, an argument that continues to be substantiated (Girin et al., 2001; Stacey et al., 2016; Sueyoshi, & Hardison, 2005). In fact, Summerfield (1992) argued that lip-reading facilitates the individual to withstand an additional 4-6 dB of background noise while preserving the performance level attained through auditory perception alone.

2.3 The visual contribution to the perception of non-native speech in noise

In challenging auditory environments, all speech comprehension can be demanding, however, it is generally acknowledged that this difficulty is exacerbated when encountering foreign accented speech (Munro, 1998; Rogers et al., 2006; Van Dommelen & Hazan, 2010). Rogers et al. (2004) theorised that even fluent non-native speakers may not be as easily understood as native speech when listening conditions are poor. Nevertheless, the majority of studies on L2 intelligibility have used audio-only stimuli, so there is a lack of literature on the visual role of speech perception, as noted by different researchers (e.g., Bicevskis et al., 2016; Wheeler & Saito, 2022).

Those studies that have looked at the contribution of the visual modality in L2 speech have found that speech intelligibility involves a number of interacting factors. For example, Wheeler and Saito (2022) found that iconic gestures significantly enhanced intelligibility when the auditory signal was challenging to decipher, such as when speech included vowel mistakes or when the listener was a second language user. Additionally, Xie et al. (2014) discovered that the native language of both the *speaker* and the *listener* impacted how much people benefited from the audio-visual modality. They argued that visemes from non-native speakers may differ from those of native speakers, which could make the visual speech cues of non-native speakers less effective for native listeners. This, they say, could be associated with factors related to both the speaker and the listener, i.e., native listeners having a tendency to view non-native audio-visual speech as less trustworthy, possibly overemphasising the perceived foreignness of the production, which could lead to the disregarding of non-native visual cues. This finding was mirrored in Kawase and Wang's (2014)

study, which suggested that although visual cues are typically helpful for natives perceiving non-native speech, there can be an inhibitory effect. This means that incorrect articulatory movements may actually reduce the intelligibility of visual speech. Nevertheless, it appears that visual cues overall offer some assistance in enhancing speech intelligibility, consistent with Barros (2010), who found that visual cues aided native English students to understand Brazilian-accented English better. Along the same lines, Banks et al. (2015) found that identification accuracy of Japanese-accented speech in noise by native English listeners was significantly better in an audio-visual condition than in an audio-only condition.

Other studies examining the role of the visual modality have studied the perspective of a non-native listener, i.e., an L2 learner. Not all studies have included background noise as a factor, but they have found evidence of individual differences in perceptions of foreign-accented speech among L2 listeners. Factors such as metacognition, proficiency level, and L1 -L2 distance were identified as influencing the value of visible components in language perception, with the authors highlighting the need for further research (e.g., Cebrian et al., 2012; Saito et al., 2019; Sueyoshi & Hardison, 2005, Van Dommelen & Hazan, 2010). Also, in a study by Ortega-Llebaria et al. (2001), they found that the addition of visual information significantly improved English consonant identification for both native and L2 speakers, regardless of their language background, by 3.7% for Spanish speakers and 5.7% for English speakers. However, vowel identification did not show a significant improvement in the AV presentation, by only 1.7%.

Many of the Studies that have investigated visual cues from non-native English speakers have focused mainly on Asian-accented speech (Hardison, 2003; Rogers et al., 2004; Sekiyama & Tohkura, 1993; Wheeler & Saito, 2022).

These studies have been extremely informative in showing the influence of speaker background on the utility of visual cues. In one such study, Yi et al. (2013) showed audio-visual stimuli to native English perceivers. The stimuli consisted of English sentences provided by both native English and Korean speakers. They discovered that although visual cues helped improve the understanding of the English sentences produced by Korean speakers, the visual information was not as pronounced as for the sentences produced by native English speakers (Yi et al., 2013, as cited in Kawase & Wang, 2014).

2.4 French-accented speech

As stated earlier in the paper (section 1.1), the current study hypothesises that visual cues vary from language to language and, therefore, that the language background of the speaker will influence the visual cues in the production of English vowels (McGuire & Babel, 2012). As most studies have investigated visual cues provided by non-native English speakers from Asian language backgrounds, this led the current study to focus on European French.

In addressing the significance of lip and mouth movements in speech perception, it is essential to narrow the scope of investigation. Visemes, the visual representations of speech sounds, can represent both vowels and consonants. Visemes are often categorised based on the phonemes they represent. For example, the viseme for the vowel sound /a/ would involve an open mouth with the tongue low and flat, while the viseme for the consonant sound /m/ would involve closed lips. In a study about word recognition and visemes by Pattamadilok and Sato (2022), they offer the following image depicting some visemes and their corresponding phoneme.

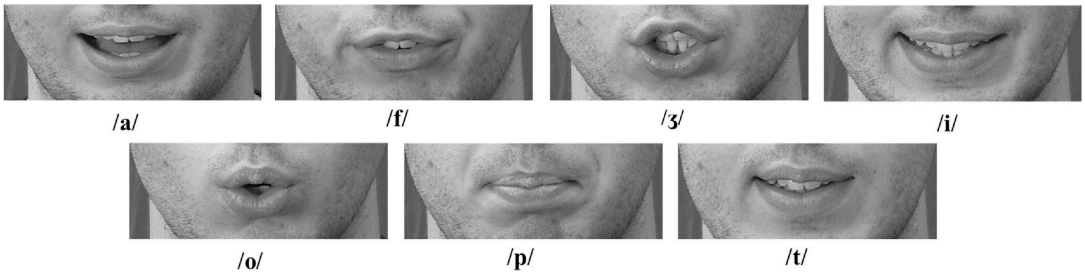

Figure 1: Visemes and their phonemes (Pattamadilok & Sato, 2022)

As the present study is only focusing on vowels, French was chosen due to its known propensity for lip-rounding and larger number of rounded vowels than English (Tranel, 1987; Zerling, 1992). Lip-rounding occurs not only for vowels, but also in anticipation of vowels and the consonants that precede and follow them. In French, this lip-rounding is contrastive in front vowels, while in English, lip-rounding is a secondary feature of high and mid back vowels. Figure 2 illustrates the vowel system of French, indicating degree of aperture, backness, and lip-rounding (Léon, 1992).

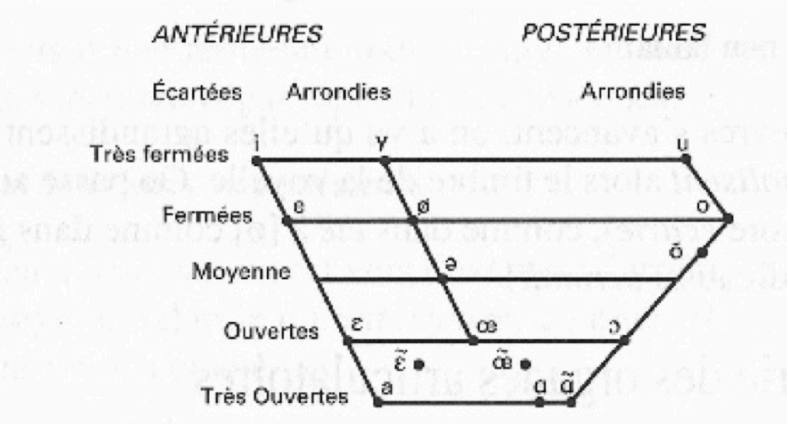

Figure 2: Diagram of French oral vowels (Léon 1992)

French speakers also have a tendency for lip protrusion, which appears to accompany French rounded vowels but not so much English rounded vowels (Zerling, 1992). In a comparative analysis of French and English phonetics, French exhibits more anterior vowels, with corresponding vowels in both languages demonstrating a more anterior resonance in French. Additionally, half of the sixteen French vowels require rounded lips and a convex tongue position during articulation. According to Monod (1971), this front characteristic, along with lip rounding and protrusion, is rarely seen in English.

Léon (1992), along with Fougeron and Smith (1993) describe the schwa /ə/ as a central vowel with rounding in French (see Figure 2). However, Tranel (1987) argues that vowel reduction occurs in English as a compensation for the strength of stress, yet the relatively weaker stress in French permits all syllables to maintain the complete quality of their vowels. In other words, the guiding principle in French is to refrain from reducing vowels to a schwa sound (Tranel, 1987). Consequently, French L1 speakers might unintentionally use more rounded lip gestures when attempting to produce /ə/ in English, which could lead to perceptible differences in pronunciation compared to native English speakers.

The difference in amount of rounding and lip protrusion between the two languages is known to cause problems for students (Delattre, 1951, as cited in Monod, 1971). Tranel (1987) reports that although transitioning from French to English and vice versa involves removing specific vowels and learning new ones, in fact the phonetic differences between the two languages are much greater than what the two inventories indicate. Certain vowels, which initially appear transferable between the languages, actually necessitate significant adjustments in their articulation. So, while phonetic symbols used for transcribing French and

English vowels may overlap, this does not mean that the sounds are exactly the same. Important phonetic differences exist between the two languages. The shared symbols indicate a relative phonetic similarity, but primarily mark the parallel articulatory characteristics within each system. For example, [i] represents both the French word 'lit' and the English word 'beat', but despite being the front unrounded vowel with the smallest degree of aperture, French [i] and English [i] are not phonetically identical, despite their similarities (Tranel, 1987).

These language-specific differences can prove interesting in a study about the visual contribution to speech intelligibility, especially given that different features are more useful for different modalities, e.g., tongue height being more robust for an audio-only mode and rounding being prominent on the audio-visual level (Robert-Ribes et al., 1998). Huang and Erickson (2019) investigated the tongue movements and jaw articulation of L2 French speakers of English in a study that focused on prominence in English sentences. They found that mouth movement patterns differed considerably from those of the native speaker. Other studies have looked at French vowels in noise (Benoit et al., 1994), although the author is not aware of any that have investigated the intelligibility of French productions of English.

3. Method

A small pilot study involving both a native and non-native English speaker was carried out before the main study. This helped determine several crucial elements for the main experiment, including the most suitable location for speaker recordings, how to present the stimuli, and the appropriate level of background

noise. These factors, alongside the remaining methodological framework, will be outlined in this section.

3.1 Speakers

Two French speakers of L2 English (one female, one male) and one native English speaker (male) recorded the test items. The two non-native speakers from France were highly proficient in English and had previously lived in England, where they studied post-graduate qualifications in English. The native English speaker from London presented a standard RP British English accent, whilst the L2 speakers maintained a moderate degree of French accent, as identified by the author. All speakers were between 35 and 45 years old with no speech impairment. They all participated on a voluntary basis.

3.2 Stimuli

The stimuli used consisted of 32 consonant-vowel-consonant (CVC) minimal pairs, specifically selected to assess the intelligibility of English vowels: (/ɔː/, /ɑː/, /iː/, /ɪ/, /æ/, /ʌ/, /e/, /æ/). Additionally, 8 tokens were included to test /ə/ and /ɜː/, with some of these being two-word phrases containing 'a'. Each experimental condition comprised of a list of 25 tokens which all used real English words (see Appendix A). Tokens were categorised by vowel features plus 5 distractor words. Tokens were presented to examine lip rounding versus neutral lips (/ɔː/ vs /ɑː/), lip spreading versus neutral lips (/iː/ vs /ɪ/), front versus non-front vowels (/æ/ vs /ʌ/), degree of jaw opening (/e/ vs /æ/) and French production of English (/ə/ and /ɜː/). Each feature was tested by means of two sets of minimal pairs per condition e.g. *Cart* vs *Court*, *Tart* vs *Taught*. The test words were read

in the carrier phrase ‘I say ___’, and the phrases including /ə/ and /ɜ:/ in the carrier phrase ‘Now I say___’. The test vowels were presented in CVC words with neutral lip consonants: /t d s z l n/. Distractors included these and also consonants with known visual cues, such as /p b m r w θ/.

3.3 Recording

Recordings were made in a silent room with a white background. Speakers were instructed to maintain a natural speaking rate. They were given time to familiarise themselves with the stimuli before recording. Flashcards of the stimuli were used as prompts. Each speaker recorded 25 tokens presented in the carrier phrase. Audio recordings were made using Praat software on a Macbook Pro using a Shure SM58 microphone through a Focusrite Scarlett Audio Interface at a 48 000 Hz sampling rate (Kawase & Wang, 2014). Video recordings were made using a Samsung Galaxy A12 Front HD camera attached to a CXYP 18-inch LED Ring Light on a Tripod Stand. Speakers sat a foot (30.5cm) from the camera and recording device (see Figures 3 & 4).

Figure 3. French speaker audio-visual condition

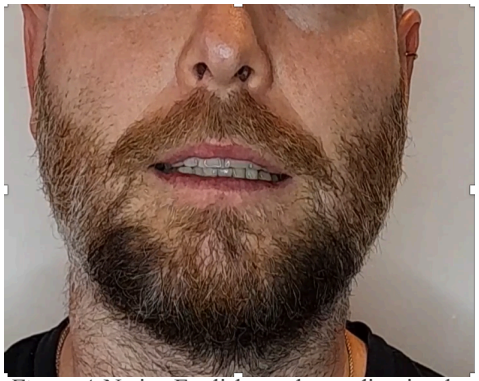

Figure 4. Native English speaker audio-visual condition

3.4 Editing

The raw audio files were imported into Logic Pro X for post-processing. In order to help stabilise the volume and dynamic range of each track, they were normalised at 0.1 dB, following the method proposed by Kawase and Wang (2014). Additionally, compression was applied using the 'natural vocal' pre-set in Logic Pro X. The mean intensity of each audio track was determined using Praat software, and subsequently the cafeteria noise was added. Cafeteria noise was selected as the background noise due to its representation of typical everyday situations (Howard-Jones, 1993; Munro, 1998).

Determining the optimal signal-to-noise ratio (SNR) for the speech signal remains a topic of debate, depending on the characteristics of the masking stimulus. Several studies have employed an SNR of -15 dB as it has been described as “quite challenging” in terms of background noise (Jin & Liu, 2014, p. 6). This SNR value also represents a midpoint between -30 dB, where speech is virtually unintelligibly, and SNR 0, which signifies equal levels of noise and speech (Sumby & Pollack, 1953). At -15dB, consistent with the pilot study, the masking background noise renders speech perception difficult, yet not impossible (Jin & Liu, 2014; Summers et al., 1988). Therefore, all conditions were subjected to cafeteria noise added at a signal-to-noise ratio of -15dB.

The combination of the audio tracks and cafeteria noise was performed using Logic Pro X software. Notably, cafeteria noise was added during the post-processing stage rather than during the recording process to prevent the Lombard effect, wherein individuals involuntarily adjust their vocal output in noisy environments, such as increasing loudness or making other modifications to make their speech audible (Assmann & Summerfield, 2004).

For video presentations, Adobe Premiere Pro was used to edit the videos as necessary. To minimise visual distractions, the videos were cropped to show only the lower part of the speaker's face, following the method employed by Dubois et al. (2012). Finally, PowerPoint was used for the actual presentations.

3.5 Listening participants

Listening participants consisted of 24 listeners (10 female, 14 male) recruited using snowball sampling. Listeners were all native monolingual English speakers from England.[1] They filled in a sociodemographic questionnaire to control for social variables (see Appendix B). None of the listening participants had lived in France or had any experience of French beyond high school learning. They all reported normal hearing and normal or corrected vision. They were aged between 19 and 64 years old (Mean age = 39.9).

3.6 Presentation and Test

The test was conducted using a MacBook Pro laptop and Sony MDR-ZX310 headphones. In order to maintain consistency, all participants were provided with identical equipment and were seated within a dimly lit room. Prior to the test, a short practice session was carried out and instructions were given that each token would be presented twice in the carrier phrase 'I say...' with a beep in between each one. Presentations were counterbalanced in terms of both speaker (NE vs FR) and modality order (A vs AV) (Hazan et al., 2002; Sennema et al., 2003). All participants observed all the auditory (A) and audio-visual (AV) conditions presented by the native speaker (NS) as well as one of the two French

[1] Three participants were living in Spain at the time of the experiment and had learnt some basic Spanish.

speakers. Listeners were asked to write down the word they heard in the two test conditions, A and AV. They were given the option to pause the audio/video if they needed time to write.

4. Results

Scoring criteria were based on the correct identification of vowel phoneme regardless of the perceived word. That is, because the study only assessed vowel recognition, participants' responses were evaluated based on their accurate identification of the vowel phoneme, without requiring precise word recognition. For example, a point would be awarded for 'sag' if the target word was 'sack'. The data were analysed to obtain percentages of correct vowel identification per feature, modality and speaker group for each of the native English listeners.

In a first analysis, listeners' mean correct identification data were examined to assess the contribution of the visual information for the French and native English speaker. In this analysis, the data for the presence and absence of each vowel feature have been pooled. A two-way repeated measure analysis of variance (ANOVA), with speaker group (French, English), and modality type (A and AV) as factors on mean correct identification was performed.

In a second analysis, directed to assess the effect of modality type on each feature individually for the native English (NE) and French (FR) speakers, listeners' mean correct identification data were analysed. Two-way ANOVAs, with feature (e.g. presence or absence of lip-rounding) and modality type (A and AV) as factors were performed separately for each speaker group (FR, NE). The vowel features analysed were lip rounding versus neutral lips (/ɔː/ vs /ɑː/), lip

spreading versus neutral lips (/i:/ vs /I/), front versus non-front vowels (/æ/ vs /ʌ/), degree of jaw opening (/e/ vs /æ/) and French production of English (/ə/ and /ɜ:/). If interaction effects are significant, it indicates that the feature effect on correct identification depends on modality (AV or A), e.g., that the vowel /i:/ with lip-spreading is better identified in the AV than /ɪ/ without lip-spreading, but not in the A-only condition.

In a third analysis directed to assess the effect of the feature (e.g. presence or absence of lip-rounding) in the two modalities for the French and English speakers, the gain of the visual modality (i.e., the 'visual benefit') was computed as the difference in correct identification between the audio-visual and the audio-only modality (that is, AV-A). Further two-way ANOVAS were performed with gain or visual benefit (AV-A) as the dependent variable and feature (e.g. presence or absence of lip-rounding) and speaker (NE vs FR) as factors. This analysis allowed for examination of whether the difference between the AV and the A modality is greater for visible than non-visible features (e.g. presence vs absence of lip-rounding), and whether it differs for FR and NE speakers.

4.1 Effect of visual information across all features

A two-way ANOVA was carried out to assess the effect of the visual information across all vowel features. The results are presented in Figure 5.

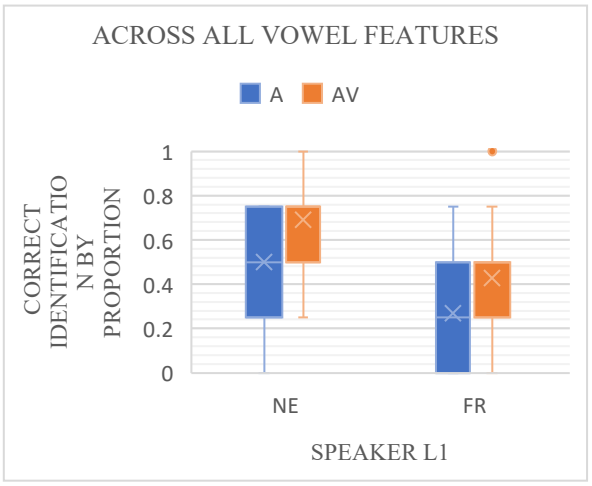

Figure 5: Mean proportion of vowel intelligibility scores (on the y-axis) across all features for native English and French speakers (on the x -axis) in the Audio-only and Audio-visual condition[2]

Mean vowel identification scores for the native English speaker were 50% in the audio-only condition and 69% in the audio-visual condition. Mean vowel identification scores for the French speaker were 27% for audio-only condition and 43% for the audio-visual condition, see Fig. 5. Significant main effects were observed for modality (A vs AV) [$F(1.476) = 70.57$, $p < 0.001$], with higher identification for the AV presentation, and for speaker [$F(1.476) = 141.64$, $p < 0.001$], with higher intelligibility for the English speaker. The lack of significant interaction effects, indicate that the increase in intelligibility in the AV condition was overall similar in both speaker groups, as illustrated in Figure 5.

In the next sections an analysis of the effect contributed by vowel features studied: lip rounding versus neutral lips (/ɔː/ vs /ɑː/), lip spreading versus neutral lips (/iː/ vs /I/), front versus non-front vowels (frontness) (/æ/ vs /ʌ/), degree of jaw opening (/e/ vs /æ/) and French production of English (/ə/ and /ɜː/)

[2] These results include all stimuli tested i.e., both vowels presenting the tested feature and neutral vowels.

will be presented.

4.2 Lip Rounding

Words containing /ɔː/ vs /ɑː/ were tested to assess the efficacy of lip rounding cues /ɔː/ vs neutral lips /ɑː/. Two-way ANOVAs for correct vowel identification in the AV and A conditions (modality factor) and presence or absence of the feature (lip-rounding vs neutral) were carried out for each speaker (NE and FR) separately.

Results of the two-way ANOVAs are shown in the boxplots below, see Fig. 6.

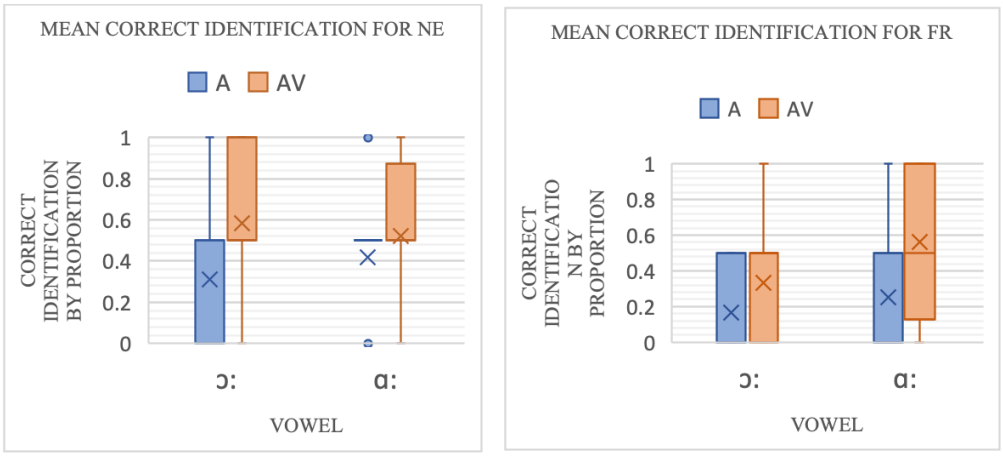

Figure 6: Intelligibility scores for the lip rounding feature in /ɔː/ vs /ɑː/ across both modalities (A and AV) for the native speaker (left panel) and the French speaker (right panel).

For the native English (NE) speaker, results show a significant effect of modality [$F_{(1, 92)}$= 7.53, p< 0.01] with participants performing better in the audio-visual condition (*M*=0.55) compared to the audio-only condition (*M*=0.36). No significant effect of vowel feature, i.e., of lip-rounding (with values pooled across

the two modalities), or significant interaction between the two factors were observed (see Fig 6, left panel).

For the French speaker (FR) the results show a significant effect of modality [$F_{(1, 92)}$= 14.36, p< 0.01] with participants gaining higher scores for both /ɔ:/ in the audio-visual condition (*M*=0.33) compared to the audio-only condition (*M*=0.16), and for /ɑ:/ in the audio-visual condition (*M*=0.56) compared to the audio condition (*M*=0.25). A significant effect of vowel feature [$F_{(1, 92)}$= 6.11, p< 0.05] was also found, with /ɑ:/ showing higher correct recognition than /ɔ:/ across both modalities (see Fig 6, right panel). Interaction effects did not reach significance.

A second two-way ANOVA was carried out, this time to examine whether the difference between the AV and the A modality (now the dependent variable) is greater for visible than non-visible features (e.g., presence vs absence of lip-rounding), and whether it differs for FR and NE speakers. To achieve this, the gain of the visual modality (i.e., the 'visual benefit') was computed as the difference in correct identification between the audio-visual and the audio-only modality (that is, AV-A) for each speaker, see Fig 7.

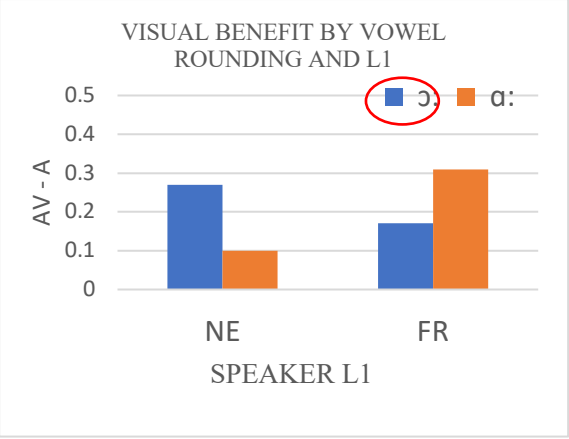

Figure 7: Visual benefit (y axis) for rounded and unrounded vowels for native English and French speakers (x axis). The red circle indicates the vowel containing the visual feature.

The results of the two-way ANOVA show no significant effect of lip-rounding (/ɔ:/ vs /ɑ:/) or L1 (English vs French speaker) on the improvement brought about by the visual modality ('visual benefit' from now on), but a significant interaction between the lip-rounding feature and L1 ($F_{(1, 92)}$= 3.79, p< 0.05). The interaction between the two main factors indicates that the visual benefit for the feature lip-rounding varies with L1, such that for NE speaker the visual benefit is larger for the rounded (/ɔ:/, M = 0.27) than the unrounded vowel (/ɑ:/, M= 0.10), as expected, but the opposite is the case for the French speaker. This is illustrated in Figure 7.

4.3 Lip Spreading

Words containing /i:/ vs /I/ were used to test the efficacy of the visual lip spreading cues. Two-way ANOVAs for correct vowel identification in the AV and A conditions (modality factor) for the presence or absence of the feature (e.g. lip-spreading vs neutral) were carried out for each speaker (NE and FR) separately, see Fig 8.

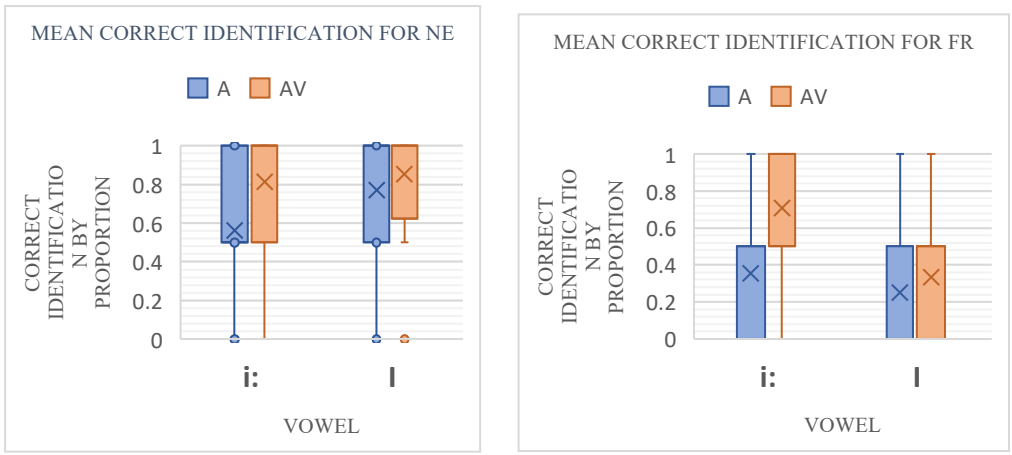

Figure 8: Intelligibility scores for lip spreading feature /ɪ:/ vs /I/ *across both modalities (A and AV) for Native English speaker (left panel) and the French speakers (right panel).*

Results of the two-way ANOVA for the NE speaker show a significant effect of modality [$F_{(1, 92)}$= 6.26, p< 0.05], with participants performing better for both /i:/ in the audio-visual condition (*M*=0.81) compared to the audio-only condition (*M*=0.56), and also for /I/ in the audio-visual condition (*M*=0.85) compared to the audio-only condition (*M*=77). Presence or absence of the feature lip-spreading and interaction effects did not reach significance (Fig 8, left panel).

Results of the two-way ANOVA for the FR speaker show a significant effect of modality [$F_{(1, 92)}$= 10.03, p< 0.01], and also for vowel feature [$F_{(1, 92)}$= 12.03, p< 0.01]. Participants performed better in the AV (*M*=0.52) compared to the Audio-only condition (*M*=0.25), and also for /i:/ (M=0.53) than /ɪ/ (M=0.29). An interaction almost reaches significance (p=0.052) i.e., presence of lip-spreading in /i:/ in the AV condition results in higher rate of correct identification than in /ɪ/ signifying that the beneficial effect of AV depends on the lip-spreading feature (Fig 8, right panel).

A second two-way ANOVA was carried out, this time to compare the effect of the lip-spreading feature in the two speaker groups (NE and FR) on the 'visual benefit' (AV-A). The results are presented in Figure 9.

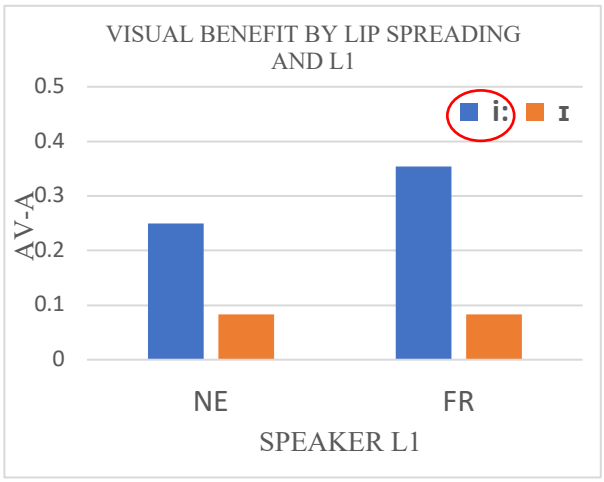

Figure 9: Visual benefit (y axis) for lip-spreading in vowels for native English and French speakers (x axis).

The results of the two-way ANOVA show a significant main effect of the lip-spreading feature (/i:/ vs /ɪ/) [($F_{(1, 92)}$= 7.09, p< 0.01], but not for L1 (English vs French speaker) on the visual benefit. This indicates that the AV benefit is greater for lip-spreading (/i:/, *M*=0.30) than for (/ɪ/, *M*=0.08) for both NE and FR speakers. This suggests that the effect of the lip-spreading cue on AV benefit does not depend on the speaker's L1. This is illustrated in Figure 9, showing a roughly comparable 'visual benefit' in both languages.

4.4 Front and non-front vowels (frontness)

To assess productions of front and non-front (central) vowels, /æ/ vs /ʌ/ were tested. Two-way ANOVAs for correct vowel identification in the AV and A conditions (modality factor) and presence or absence of the feature (front vs non-front) were carried out for each speaker (NE and FR) separately, see Fig 10.

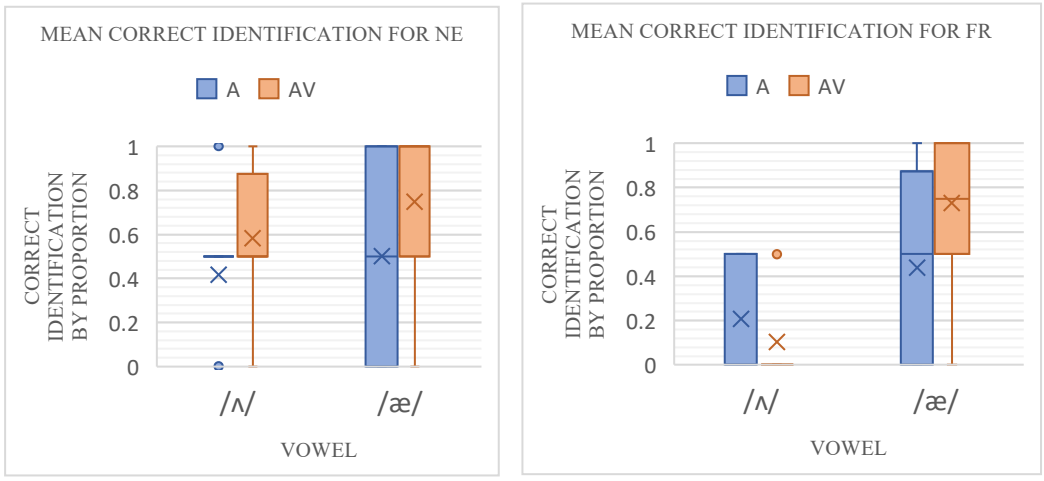

Figure 10: Intelligibility scores for front vs non-front (frontness) feature /æ/ vs /ʌ/ across both modalities (A and AV) for native English speaker (left panel) and French speaker (right panel).

For the NE speaker show a significant effect of modality [$F_{(1, 92)}$= 10.45, p< 0.01], with participants performing better the audio-visual condition (*M*=0.67) compared to the audio-only condition (*M*=0.45). Participants generally gained higher intelligibility scores for /æ/ than /ʌ/ overall, however, and the effect of the frontness feature approached significance [F(1, 92)= 3.76, p= 0.0554]. No interaction was observed (Fig 10, left panel).

For the French speaker, the results show a significant effect of vowel feature [$F_{(1,92)}$= 49.76, P=0.001], with /æ/ (*M*=0.58) being more intelligible than /ʌ/ *(M=0.16)*. However, this feature effect is moderated by a significant interaction effect [$F_{(1,92)}$= 10.69, P=0.01], which finds that /æ/ is more intelligible in the AV condition (*M*=0.73) than in the A condition (*M*=0.44), but that /ʌ/ is less intelligible in the AV condition (*M*=0.10) compared to the A condition (*M*=0.21) (Fig 10, right panel).

A second two-way ANOVA was carried out, this time to compare the effect of vowel frontness in the two speaker groups (NE and FR) on the visual benefit (AV-A), see Fig 11.

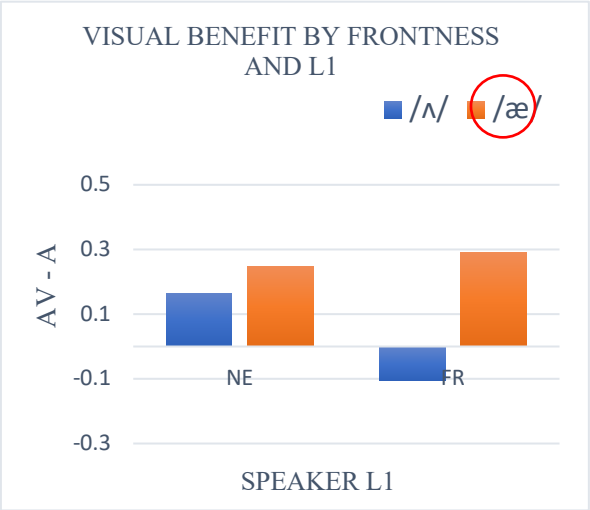

Figure 11: Visual benefit (y axis) for frontness in vowels for native English and French speakers (x axis)

The results of the two-way ANOVA show a significant main effect of the feature (frontness) [($F_{(1, 92)}$= 11.76, p< 0.001] with participants performing better for /æ/ (*M*=27) than for /ʌ/ (*M*=0.03) and no significant effect of speaker L1. However, a signification interaction was observed [($F_{(1, 92)}$= 0.59, p< 0.05] showing that the AV benefit was dependent on speaker L1 background. As depicted in Figure 11, improvements in intelligibility for /æ/ were similar for both L1 backgrounds. However, for /ʌ/, there was a modest AV benefit for the NE speaker (M = 0.17), whereas the AV benefit for the FR speaker was negative (M = -0.10).

4.5 Degree of Jaw Opening

To assess jaw opening cues, words including /e/ vs /æ/ were tested. Two-way ANOVAs for correct vowel identification in the AV and A conditions (modality factor) and for the vowel feature (jaw opening) were carried out for each speaker (NE and FR) separately, see Fig 12.

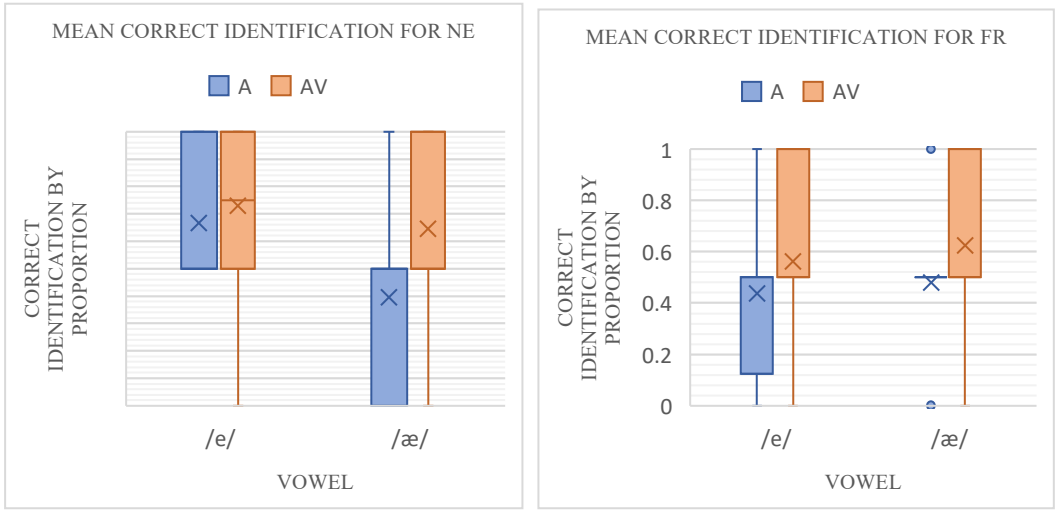

Figure 12: Intelligibility scores for degree of jaw opening /e/ vs / æ / across both modalities (A and AV) for native English speaker (left panel) and the French speaker (right panel).

Results of the two-way ANOVA for the NE speaker show a significant effect of vowel feature [$F_{(1, 92)}$= 9.17, p< 0.01], and a significant effect of modality [$F_{(1, 92)}$= 7.12, p< 0.01]. Participants performed better overall in the AV (*M*=0.69) compared to the A condition (*M*=0.53), and better overall for /e/ (*M*=0.70) than /æ/ (*M*=0.52), due to gaining high scores for /e/ in the audio-only condition (*M=0*.67) (see Fig 12, left panel).

The results for the French speaker show a significant effect of modality [$F_{(1,92)}$= 4.89, P=0.05], with AV scores being greater (*M*=0.60) than A scores *(M*=0.45*)*. However, there was not a significant effect of vowel feature. No interaction was observed (see Fig 12, right panel).

A second two-way ANOVA was carried out to assess the effect of the feature i.e., jaw opening, and speaker group (NE and FR) on the difference between the two modalities (AV- A), see Fig 13.

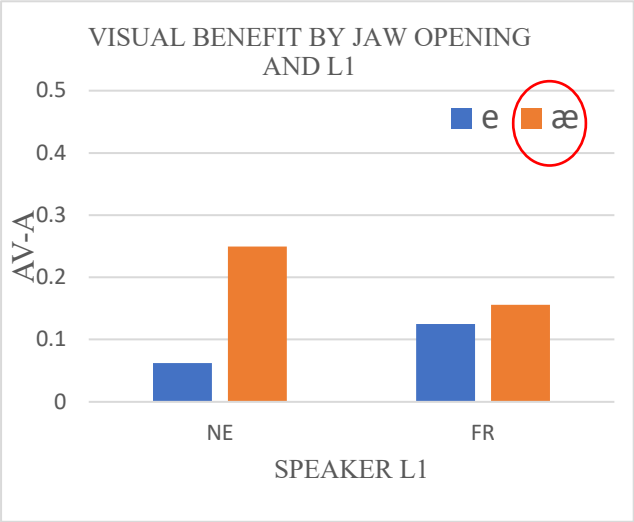

Figure 13: Visual benefit (y axis) for degree of jaw opening in vowels for native English and French speakers (x axis)

The results of the test show no significant main effects of jaw opening or L1, or significant interaction effects, on the 'visual benefit'. The improvement of the AV condition was greater for /æ/ than /e/ as shown in Figure 13, however this did not reach significance. Therefore, the improvement from the addition of the visual information did not significantly vary with degree of jaw-opening or L1 background.

4.6 Schwa Specific French Influence

To assess French productions of the schwa, words including /ə/ and /ɜː/ were tested. Two-way ANOVAs for correct vowel identification in the AV and A conditions (modality factor) and presence or absence of the feature (presence or

absence of schwa rounding) were carried out for each speaker (NE and FR) separately, see Fig 14.

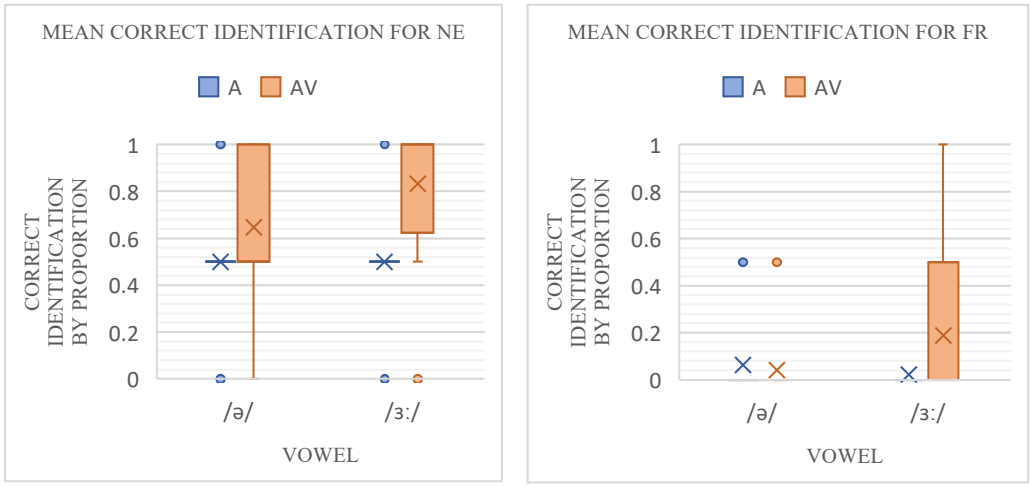

Figure 14: Intelligibility scores French productions of the schwa /ə/ and /ɜ:/ across both modalities (A and AV) for native English speaker (left panel) and the French speaker (right panel).

Results of the two-way ANOVA for the NE speaker show a significant effect of modality [$F_{(1, 92)}$= 14.78, p< 0.001], with participants performing better in the audio-visual condition (*M*=0.74) compared to the audio-only condition (*M*=0.50). No significant effect for feature or interaction were observed (see Fig 14, left).

For the French speaker, the results of the two-way ANOVA show a significant interaction effect [$F_{(1,92)}$= 5.95, P<0.05], which suggests that intelligibility varies depending on the vowel feature and modality combination (see Fig 14, right).

A second two-way ANOVA was carried out to assess the effect of the feature (lip-rounding in schwa) and speaker group (NE and FR) on the difference between the two modalities (AV-A), see Fig 15.

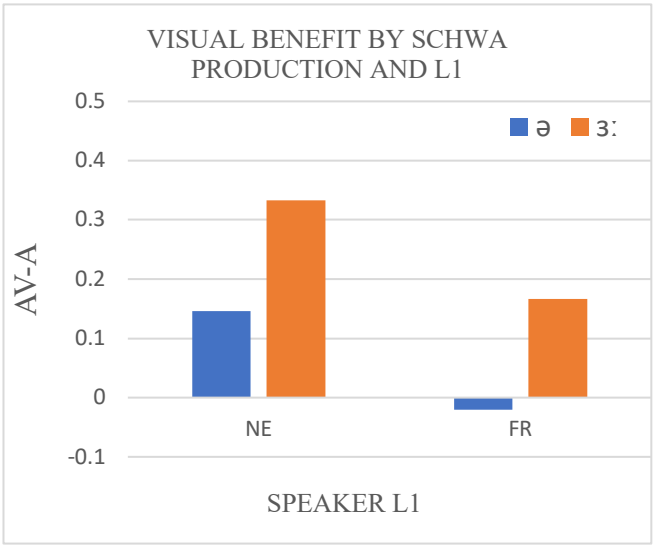

Figure 15: Visual benefit (y axis) for lip-rounding in schwa for native English and French speakers (x axis)

The results of the two-way ANOVA show a significant effect of feature i.e., lip-rounding in schwa [$F_{(1, 92)}$= 12.63, p< 0.001] on the improvement brought about by the visual modality, with participants gaining higher scores for /ɜː/ (*M*=0.25) than /ə/ (*M*=0.06). There was also a significant effect of L1 ($F_{(1, 92)}$= 9.98, p< 0.001) with the AV benefit results for the NE (*M* =0.24) being greater than those for the FR speaker (*M* =0.07). For the schwa /ə/, the results for the French speaker indicate that the visual information resulted in a decrease in correct identification relative to the audio-only condition (i.e., a smaller AV-A), as shown in Figure 15. That is, the visual cues for the French speaker were detrimental for schwa identification. No significant interaction was found.

5. Discussion

This paper investigated the intelligibility of English vowels in noise, comparing the benefit of the visual cues provided by non-native English speakers, specifically French (L1) speakers, with those provided by native speakers of English. Different types of visual cues were analysed. Firstly, visual cues that disambiguate contrasts or are secondary features of contrasts both in English and French vowels (lip-spreading, lip-rounding, frontness and jaw opening); secondly, a characteristic French feature, vowel rounding in schwa. Native English perceivers were used to test vowel intelligibility in noise in both conditions (audio-only and audio-visual). The contribution of visual cues in vowels, as opposed to those in consonants, has received relatively little attention, but the few studies that explore them did not find they significantly enhanced speech intelligibility (e.g. for Spanish learners of English, Ortega-Llebaria et al. 2001).

The present study hypothesised that the visual modality would improve speech intelligibility for both native and non-native English speakers, albeit with differing levels of effectiveness depending on the speaker's linguistic background. In addressing the first hypothesis, the results found that the native English perceivers showed significant improvements in the audio-visual condition for both the French speakers and the control English speaker, with larger effects for certain features for one speaker group over the other.

The findings of this study indicated that the visual cues provided by native and French non-native speakers of English were generally both beneficial for speech intelligibility. Some visual cues significantly enhanced intelligibility in noise (i.e., larger AV benefit) for both NE and FR speakers. These are lip-

spreading (/iː/ vs /ɪ/) and frontness (/æ/ vs /ʌ/), with a larger improvement in identification when the visual cue was present (i.e., in /iː/ and /æ/) than when it was not (i.e., in /ɪ/ and /ʌ/).

Other visual cues were found to increase intelligibility in noise more for the English than the French speaker. The presence of lip rounding in /ɔː/ in the AV condition increased intelligibility significantly more for English than for French speakers. Whilst improvements were observed in both speakers, the improvement of vowel identification for the native speaker was much greater for rounded /ɔː/ than /ɑː/. Therefore, the visual cue of lip-rounding from the native speaker was more beneficial than for the French speaker. However, unexpectedly, for the French speaker, although scores for the rounded vowel /ɔː/ did improve in the audio-visual condition, the improvements of the visual contribution for /ɑː/ were even more profound. This may be due to them producing a more open vowel. Also, improvements were observed for the open vowel /æ/ over /e/ in the audio-visual condition more for the English speaker than the French speaker, however this did not reach significance, which could be due to participants gaining high scores in the audio condition therefore not leaving much scope for improvement.

As per the second hypothesis, that the characteristic lip gestures and protrusion observed in French speakers may potentially compromise speech intelligibility in the audio-visual condition, supporting evidence was found in confusions caused by the unique lip movements and protrusions provided by the French speakers, which were found in both schwa /ə/ and /ʌ/.

The findings show that some visual cues provided by the French speaker, rather than enhancing intelligibility, were detrimental i.e., the visual cue resulted

in a decrease in intelligibility compared to the audio-only condition. For the test words looking specifically at any possible French influence (lip-rounding) on the production of schwa /ə/, not only did participants gain poor intelligibility scores across both modalities, but the visual modality did not prove to be beneficial for the French productions of the schwa. This suggests that the French speakers were providing misleading or unexpected visual cues, therefore confusing the native British English speaker (Fougeron and Smith, 1993). Interestingly, on more than one occasion, listening participants perceived the word 'audit' which starts with a rounded vowel, instead of 'a date' beginning with the schwa. In Figure 16, a visual still captures the mid-production of the schwa /ə/ vowel by both a French and English speaker. Confusions were also present in perceptions of the mid-open vowel /ʌ/ included in the front vs non-front feature. The visual benefit resulted in a negative score for /ʌ/. One possible explanation could be the fact that most dialects of French only have one open central /a/ vowel, and the French speaker may have been trying to exaggerate a difference using the wrong visual cue.

Figure 16: Lip and mouth movements during the production of the schwa. French speaker on the left with lips slightly protruded and English speaker on the right with relaxed mouth.

These findings parallel the results reported by Huang and Erickson (2019), whose study explored prosody rather than vowel productions. The atypical jaw movement patterns they observed serve as another articulatory difference that could potentially confuse a native perceiver among French speakers of English. Furthermore, as a by-product of this investigation, weight has been added to previous research that argues that non-native speech is often harder to understand in challenging auditory conditions, with the French speaker being less intelligible than the English speaker in all conditions (Kawase & Wang, 2014; Xie et al., 2014).

It should also be noted that the current study has some limitations. It was observed during the pilot test, that one of the less proficient French speakers exhibited more noticeable instances of lip protrusion compared to any of the speakers involved in the study proper. As mentioned earlier, the study specifically recruited French L1 speakers with a high proficiency in English. While this approach ensured consistency in the variable, a larger study could also explore the impact of varying proficiency levels. In a similar vein, the current study did not specifically investigate noise level as a variable. It is possible that different or more diverse results could have been obtained if the study had varied the signal-to-noise ratio (SNR) or included a condition with visual cues only. On some occasions results were limited due to high participant scores in the audio-only condition, this ceiling effect could have been avoided if the SNR had been more challenging.

6. Conclusion

The current study set out to investigate the benefit of visual cues provided by non-native English speakers in adverse auditory conditions and specifically to discover if these cues were as useful as those provided by a native speaker. French was selected due to its characteristic lip rounding, and vowels were chosen for examination given the predominant focus on consonants in previous similar studies. The findings of this research contribute to the understanding of the visual modality in speech perception and intelligibility by providing empirical evidence for the benefits of visual cues from native and non-native speakers of English. Since overall intelligibility scores were significantly higher over both modalities (A and AV) for the native speaker, this highlights the challenges faced by non-native speakers in producing L2 sounds correctly.

Notably, the study's main finding emphasises that in challenging listening environments, such as noisy cafeteria backgrounds, listeners benefit from audio-visual information provided by both native and non-native speakers. In most cases, there was no differential efficacy between cues from native and French English speakers. However, there was a significant speaker difference found for the AV benefit of productions of English (/ə/ and /ɜː/) and for productions of /ʌ/. Therefore, the visual cues provided by the French speaker were misleading or confusing to the native English perceiver.

Although visual speech information typically aids in the perception of non-native speech by native speakers, it can also hinder comprehension when incorrect visual gestures are present, resulting in reduced speech intelligibility. Even though some visual features appear to have the same status in both

languages (e.g., lip-rounding in back vowels), the visual gestures appear to differ across languages. There are language-specific features that make French-accented speech more difficult to understand even in face-to face interaction. For example, where participants performed better in the audio condition, it could be suggested that the incongruent visual cues are creating a type of McGurk effect. As a result, the implications of this research are significant for instructional strategies aimed at improving non-native pronunciation and visual gestures, speech recognition software development, and investigations into communication challenges in noisy work environments.

Recommendations for further research could include, but are not limited to, varying the proficiency level of French-speaking participants and adjusting the signal-to-noise ratio (SNR).

References

Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In E. A. Lopez-Poveda, A. R. Palmer, & R. R. Fay (Eds.), *Speech processing in the auditory system: Springer handbook of auditory research* (vol. 18, pp. 45-65). Springer. https://doi.org/10.1007/0-387-21575-1_5

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Audiovisual cues benefit recognition of accented speech in noise but not perceptual adaptation. *Frontiers in Human Neuroscience*, *9*(422), 1-13. https://doi.org/10.3389/fnhum.2015.00422

Barros, P. C. (2010). *"It's easier to understand": the effect of a speaker's accent, visual cues, and background knowledge on listening comprehension.* [Doctoral dissertation, Kansas State University]. K-Rex Repository

Benoit, C., Mohamadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, *37*(5), 1195-1203. https://doi.org/10.1044/jshr.3705.1195

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*, 1600-1610. https://doi.org/10.1121/1.1603234

Bicevskis, K., Derrick, D., & Gick, B. (2016). Visual-tactile integration in speech perception: Evidence for modality neutral speech primitives. *The Journal of the Acoustical Society of America*, *140*(5), 3531–3539. https://doi.org/10.1121/1.4965968

Borràs-Comes, J, and P. Prieto. (2011). 'Seeing tunes': The role of visual gestures in tune interpretation. *Laboratory Phonology, 2*(2), 355-380. https://doi.org/10.1515/labphon.2011.013

Bronkhorst, A. (2000). The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions. *Acta Acustica United with Acustica*, *86*, 117-128. https://doi.org/10.3758/s13414-015-0882-9

Brungart, D. S., Barrett, M. E., Cohen, J. I., Fodor, C., Yancey, C. M., & Gordon-Salant, S. (2020). Objective assessment of speech intelligibility in crowded public spaces. *Ear and Hearing*, *41,* 68S–78S. https://doi.org/10.1097/AUD.0000000000000943

Cebrián, J., & Carlet, A. (2012, January). Audiovisual perception of native and non-native sounds by native and non-native speakers. In Alegre, S. M., Moyer, M., Pladevall, E., & Tubau, S. (Eds.). *At a time of crisis: English and American studies in Spain*, Works from the 35th AEDEAN Conference (pp. 300-307).

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25, 975–979. https://doi.org/10.1121/1.1907229

Dubois, O., Otzenberger, H., Gounot, D., Sock, R., & Metz-Lutz, M.-N. (2012). Visemic processing in audio-visual discrimination of natural speech: A simultaneous fMRI–EEG study. *Neuropsychologia*, *50*(7), 1316-1326. https://doi.org/10.1016/j.neuropsychologia.2012.02.016

Flege, J. E., Munro, M. J., & Mackay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*(5), 3125-3134. https://doi.org/10.1121/1.413041

Fougeron, C., & Smith, C. L. (1993). French. *Journal of the International Phonetic Association*, *23*(2), 73–76. https://doi:10.1017/S0025100300004874

Gabbay, A., Shamir, A., & Peleg, S. (2017). Visual speech enhancement. *arXiv preprint arXiv:1711.08789*. https://doi.org/10.48550/arXiv.1711.08789

George, E. L., Festen, J. M., & Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *120*(4), 2295-2311. https://doi.org/10.1121/1.2266530

Girin, L., Schwartz, J.-L., & Feng, G. (2001). Audio-visual enhancement of speech in noise. *Journal of the Acoustical Society of America*, *109*(6), 3007–3020. https://doi.org/10.1121/1.1358887

Giuliani N. (2020, November 2) *For speech sounds, 6 feet with a mask is like 12 feet without*. ASHA Lead. https://leader.pubs.asha.org/do/10.1044/leader.AEA.25112020.26/full

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America*, *103*(5), 2677-2690. https://doi.org/10.1121/1.422788

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495-522. https://doi.org/10.1017/S0142716403000250

Hazan, V., Sennema, A., & Faulkner, A. (2002, September). Audiovisual perception in L2 learners. *Proc. Seventh International Conference on Spoken Language Processing.* 1685-1688, https://doi.org/10.21437/ICSLP.2002-426

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*(3), 1740–1751. https://doi.org/10.1121/1.2166611

Howard-Jones, P. A., & Rosen, S. (1993). The perception of speech in fluctuating noise. *Acustica, 78*, 258-272.

Huang, T., & Erickson, D. (2019). Articulation of English 'prominence' by L1 (English) and L2 (French) speakers. *In International Congress of Phonetic Sciences,* Melbourne, Australia (pp. 1-5),

Jin, S. H., & Liu, C. (2014). English vowel identification in quiet and noise: effects of listeners' native language background. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00305

Kawase, H. B., & Wang, Y. (2014). The influence of visual speech information on the intelligibility of English consonants produced by non-native speakers. *The Journal of the Acoustical Society of America*, *136*(3), 1352–1362. https://doi.org/10.1121/1.4892770

Léon, P. (1992). Phonetisme et prononciations dufranr; ais, coli. *Nathan Universite*, Nathan, Paris.

McGuire, G., & Babel, M. (2012). A cross-modal account for synchronic and diachronic patterns of /f/ and /θ/ in English. *Laboratory Phonology*, *3*(2), 251-272. https://doi.org/10.1515/lp-2012-0014

McGurk, H., Macdonald, J. (1976) Hearing lips and seeing voices. *Nature, 264*, 746–748. https://doi.org/10.1038/264746a0

McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Van Engen, K. J. (2018). Coping with adversity: Individual differences in the perception of

noisy and accented speech. *Attention, Perception, & Psychophysics*, *80*(6), 1559-1570. https://doi.org/10.3758/s13414-018-1537-4

Melguy, Y. V., & Johnson, K. (2021, April 1). General adaptation to accented English: Speech intelligibility unaffected by perceived source of non-native accent. *Journal of the Acoustical Society of America*, *149*(4), 2602–2614. https://doi.org/10.1121/10.0004240

Monod, P. A. (1971). French vowels vs. English vowels. The French Review, *45*(1), 88–95. https://www.jstor.org/stable/385695

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73-97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, *20*(2), 139-154. https://doi:10.1017/S0272263198002022

Ortega-Llebaria, M. M., Faulkner, A., & Hazan, V. (2001). Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English. Speech, Hearing and Language: Work in Progress, 13, 40-50.

Pattamadilok, & Sato, M. (2022). How are visemes and graphemes integrated with speech sounds during spoken word recognition? ERP evidence for supra-additive responses during audiovisual compared to auditory speech processing. *Brain and Language*, *225*, 105058. https://doi.org/10.1016/j.bandl.2021.105058

Robert-Ribes, J., Schwartz, J. L., Lallouache, T., & Escudier, P. (1998). Complementarity and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise. *The Journal of the Acoustical Society of America*, *103*(6), 3677–3689. https://doi.org/10.1121/1.423069

Rogers, C. L., Dalby, J., & Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech*, *47*(2), 139–154. https://doi.org/10.1177/00238309040470020201

Rogers, C.L., Lister, J.J., Febo, D.M., Besing, J.M., & Abrams, H.B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*, 465-485. https://doi.org/10.1017/S014271640606036X

Rosenblum, L. D., Pisoni, D. B., & Remez, R. E. (2005). *The handbook of speech perception*. Blackwell Pub. https://doi.org/10.1002/9780470757024

Saito, K., Tran, M., Suzukida, Y., Sun, H., Magne, V., & Ilkan, M. (2019). How do second language listeners perceive the comprehensibility of foreign-accented speech?: Roles of first language profiles, second language proficiency, age, experience, familiarity, and metacognition. *Studies in Second Language Acquisition*, 41(5), 1133-1149. https://doi.org/10.1017/S0272263119000226

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, *127*(1), 141-145. https://doi.org/10.1016/0304-3940(91)90914-f

Sekiyama, K., & Tohkura, Y. I. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*(4), 427-444. https://doi.org/10.1177/0033688220966635

Sekiyama, K., Tohkura, Y., & Umeda, M. (1996). A few factors which affect the degree of incorporating lip-read information into speech perception. *In Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP), 3*, 1481-1484. https://doi.org/10.1109/ICSLP.1996.607896

Sennema, A., Hazan, V., & Faulkner, A. (2003). The role of visual cues in L2 consonant perception. In Solé, M. J., Recasens, D., Romero. J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, *Barcelona.* (pp. 135-138). Causal Prod. Pty Ltd.

https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_intro.pdf

Stacey, P. C., Kitterick, P. T., Morris, S. D., & Sumner, C. J. (2016). The contribution of visual information to the perception of speech in noise with and without informative temporal fine structure. *Hearing Research*, *336*, 17-28. https://doi.org/10.1016/j.heares.2016.04.002

Sueyoshi, & Hardison, D. M. (2005). The Role of Gestures and Facial Cues in Second Language Listening Comprehension. *Language Learning*, *55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Sumby, W.H. and Pollack, I. (1954) Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*, 212-215. *https://doi.org/10.1121/1.1907309*

Summerfield, Q. (1992). Lipreading and Audio-Visual Speech Perception. *Philosophical Transactions: Biological Sciences*, *335*(1273), 71–78. https://doi.org/10.1098/rstb.1992.0009

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, *84*(3), 917–928. https://doi.org/10.1121/1.396660

Tranel, B. (1987) *The Sounds of French: An Introduction* (pp. 86-107). Cambridge University Press. https://doi.org/10.1017/CBO9780511620645.007

Van Dommelen, W. A., & Hazan, V. (2010). Perception of English consonants in noise by native and Norwegian listeners: Non-native speech perception in adverse conditions. *Speech Communication*, *52*(11-12), 968-979. https://doi.org/10.1016/j.specom.2010.05.001

Wheeler & Saito (2022). Second Language Speech Intelligibility Revisited: Differential Roles of Phonological Accuracy, Visual Speech, and Iconic Gesture. *The Modern Language Journal.* https://doi.org/106.10.1111/modl.12779

Xie, Z., Yi, H. G., & Chandrasekaran, B. (2014). Nonnative audiovisual speech perception in noise: Dissociable effects of the speaker and listener. *PloS one*, *9*(12), 114-139. https://doi.org/10.1371/journal.pone.0114439

Zerling, J. (1992). Frontal lip shape for French and English vowels. *Journal of Phonetics*, *20*, 3-14.

Zhao, C. (2022). Speech Intelligibility in Noise: The Role of Talker-Listener Accent Similarity and Second Language Experience. *Research and Advances in Education*, *1*(5), 11-25.

Appendix A:

Word Lists for Speakers:

	Native English Speaker Audio-Only SNR -15dB	Native English Speaker Audio-Visual SNR -15dB	French Speaker of English Audio-Only SNR -15dB	French Speaker of English Audio-Visual SNR -15dB
1	Tart	Gnat	Cart	Sack
2	Dork	Luck	Set	Hid
3	Deed	Taught	Heed	Knit
4	Well	Thigh	That	Bell
5	Nut	Seat	Sat	Hard
6	Sad	Tart	Suck	Tan
7	Gnat	Sit	Knit	Neat
8	Lass	Dark	Hard	Court
9	Been	They	When	Thin
10	Sit	Deed	Dud	Dad
11	Luck	Dork	Tan	Cart
12	Seat	Less	Dad	Suck
13	Less	Nut	Hoard	Set
14	Put	Put	Pray	That
15	Said	Did	Ten	Heed
16	Did	Lack	Sack	Sat
17	Dark	Said	Neat	Ten
18	Taught	Lass	Court	Dud
19	They	Been	Thin	Pray
20	Lack	Sad	Hid	Hoard
21	Thigh	Well	Bell	When
22	Upton	A test	Downton	A Date
23	A test	Upturn	A date	Downturn
24	Upturn	Her Test	Downturn	Her date
25	Her Test	Upton	Her Date	Downton

Appendix B

Participant Questionnaire

Name: _____

Age: _____

Sex _____

Where were you born? _____

Nationality: _____

Languages spoken: _____

Competency of 2nd language spoken (if any) _____

Where do you live? Please specify the country and/or region

Places lived outside of the U.K. (for longer than a few months) and for how long

_____

_____

_____

_____

Do you have any knowledge of French other than secondary school syllabus? (E.g., friend, partner or spouse from France or another French speaking country)

_____
_____
_____
_____

Do you have experience with French-accented English? (e.g., friends, colleagues)_____
_____

If so, for how long?_____