Foreigner Directed Speech

From Speech Adaptation to Cortical Tracking of the Speech Register Directed to Non-native Listeners.

Giorgio Piazza, M.A.

Supervised by

Prof. Clara D. Martin & Dr. Marina Kalashnikova

eman ta zabal zazu

Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

2023

Foreigner Directed Speech

From Speech Adaptation to Cortical Tracking of the Speech Register Directed to Non-native Listeners.

Doctoral dissertation by

Giorgio Piazza, M.A.

Thesis submitted for the degree of

Doctor of Philosophy

Supervised by

Prof. Clara D. Martin & Dr. Marina Kalashnikova

Basque Center on Cognition, Brain and Language (BCBL)

Euskal Herriko Unibertsitatea – Universidad del País Vasco (EHU/UPV)

Donostia-San Sebastián, 2023

The project that gave rise to these results received the support of a fellowship from ”la Caixa” Foundation (ID 10001043)[1]. The fellowship code is LCF/BQ/D I19/11730045. The research was supported by Spanish State Research Agency [PID2020-113926GB-I00] [2], the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 819093)[2]. This research was also supported by the Basque Government through the BERC 2022-2025 program and Funded by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010/AEI/10.13039/501100011033.

Giorgio Piazza

BCBL – Basque Center on Cognition, Brain and Language Paseo Mikeletegi, 69, 20009 - Donosotia-San Sebastián, Spain 2022

---

[1] Awarded to Giorgio Piazza

[2] Awarded to Clara D. Martin

Acknowledgements

I wish to thank my acquired family in Donostia, who made me feel at home outside Italy for the very first-time. Jose, Irene, Laura, Chiara, Coco, Sofia, Alberto, Giulia, Vicente, Asier, Sandy, Suheil, Pierma, Christoforos, Katerina, Lucilla, Dani, Inés, Hana, Marta, Matteo, I love you! Thank you for being with me during my PhD. Thank you for all the suggestions, chats, parties, caffes at Bizi, movie nights during curfew time, pintxo pote, covid at the Bukowski, vogueing at Dionis, trips in Sicily and Valencia, work-related and non-work-related help. Being with you always leads to surprises.

Eskerrik asko Donosti por los pintxos del Bar Sport y la Sebastiana, por la sidra y las sidrerías, por la txuleta, el kalimotxo en lo viejo, las cañas en Reyes Católicos, el Polviriña, la naturaleza, el atardecer en la Kontxa y el surf en Zurriola.

I would like to extend my gratitude to people at BCBL who helped me throughout these 4 years: researchers, PIs, including my two supervisors (Clara and Marina), admin, IT, and lab staff.

I thank my biological and first family, who supported me in pursuing my path, my career, long before I started my PhD. It is also thanks to `random.shuffle(["Papà", "Mamma", "Gigi"])` that I could grow up and move to Spain, where I found my way and so many good friends. All your sacrifices don’t go unnoticed.

Thanks to my friends and siblings in Italy*, in primis*, my brother Giancarlo Brandon, who edited so many drafts of my first academic paper. But also, Zuli, Jack, Franco, who I grew up with. After so much time we still have fun together. To Nicola, with whom I used to have deep conversations while doing wild camping (we survived a *Chupacabras* attack). To my rugby friends with whom I learned what teamwork and sacrifice are. To my friends in the Ballroom scene, such a new and unexpected thing to me, who showed me that another way is possible.

Grazie al mio cuore piccolo, Marta, per essere venuta a trovarmi quella prima volta.
Grazie di mostrarmi la parte migliore di me ogni giorno,
di coltivare la mia parte sensibile, di darmi coraggio,
e di insegnarmi quello che non sapevo di aver bisogno.

Table of Contents

Abstract

This thesis aimed to investigate the mechanisms underlying the perception and production of Foreigner Directed Speech. Foreigner Directed Speech (FDS) is the result of speech accommodation produced by native speakers of a certain language while addressing non-native listeners of that language. By studying FDS, we aimed to explore the intricate dynamics of native and non-native communication. The work included in this thesis sheds light on the processes involved in language accommodation, its resulting acoustic adjustments (production perspective), and its implications for effective communication and teaching (perception perspective).

To lay a solid foundation for our research, the first chapter delves into an exploration of the key features, definitions, and research gaps surrounding the FDS phenomenon. By examining existing literature and conducting a comprehensive review of previous studies, we aimed to provide a starting point to deepen unexplored aspects of FDS production and its consequences on second language (L2) perception and learning. Hence, in the first chapter, we identified the more characteristic FDS acoustic features, such as vowel hyperarticulation and low speech rate. Also, we observed that previous studies assumed that FDS is adjusted to meet non-native listeners' linguistic needs. Accordingly, FDS is hypothesised to be a promoter of L2 acquisition. We concluded this chapter by stressing that the didactic assumptions of FDS should be investigated by assessing whether production of FDS adapts to goals of teaching various aspects of the L2 (or communicative goals), such as writing or pronunciation. On the other hand, we also pointed out the lack of research on the impact of FDS on non-native listener's L2 perception and learning. These aspects were investigated in the next chapters.

In the second chapter, we aimed to shed light on whether FDS speech adaptations vary depending on the speaker's communicative goals and the listeners' profiles (i.e., native vs. non-native). We recorded native Spanish speakers naming novel objects to aid listeners'

performance in comprehension, pronunciation, and writing tasks. Each speaker interacted with a native (Native Directed Speech, NDS condition) and a non-native Spanish listener (FDS condition). We extracted measures of vowel hyperarticulation, duration, intensity, speech rate, and pitch height to assess speakers' speech adjustments. Vowel hyperarticulation results showed that speakers adapted their speech in the writing as compared to the comprehension condition, and in FDS compared to NDS. Pitch height, intensity, and duration results revealed that speakers provided acoustic cues adjusted to listeners' profile and conditions. Therefore, speakers indeed adjust acoustic features to communicative goals depending on the listener's profile.

After confirming our hypothesis that FDS is adapted to teach specific aspects of the L2, in chapter 3, we explored whether FDS facilitates L2 word learning and production as compared to NDS. Fifty Spanish participants were asked to learn 16 novel English words in an online experiment. Words contained either the /i - ɪ/ or the /æ - ɛ/ vowel contrasts, both non-native to Spanish speakers. Participants were divided in two groups: one being exposed to FDS and one to NDS. Participants carried out three tasks: a) Recognition task, b) Production task, and c) Continuum discrimination task. In a) participants were presented with novel objects and their auditory labels (e.g., "This is a deest") produced in FDS or NDS. Next, they were tested on their ability to recognize novel words. In b) participants were presented with the novel objects and asked to pronounce the associated labels. In c) participants were tested on perceptual discrimination of continua involving the four target vowels. We calculated participants' accuracy and response latencies. Also, for b) we measured participants' phonetic accuracy of word and vowel production. Growth curve analyses showed that the FDS group recognized novel words faster and learned the words of the /i - ɪ/ contrast better than the NDS group. Also, FDS group's production for the /i - ɪ/ contrast was more accurate than the NDS group. Conversely, the two groups did not differ for perceptual discrimination of vowel continua. These findings supported the didactic assumption of FDS, which should be considered in L2 teaching models.

However, experiments with controlled manipulations and single word/short sentences do not reflect the way in which non-native listeners receive L2 (and FDS) exposure. In chapter 4, we endeavour to explore how FDS impacts perception and comprehension in an ecologically valid experiment. By examining the ways in which FDS affects the interpretation of spoken language, we aimed to uncover its potential benefits and limitations on L2 perception. Here, we investigated the effects of exposure to FDS and NDS on speech perception in non-native listeners and native listeners. Naturalistic stimuli in the form of stories were employed in each speech register. Traditional behavioural studies typically examine language processing after it has occurred, while electroencephalography (EEG) allows for real-time analysis of language processing as it unfolds. One way to analyse EEG modulation to continuous speech is by measuring cortical tracking of speech features (such as envelope). That is, cortical tracking of low-frequency bands is associated with language processing since some studies identified that as markers of underlying linguistic processing. Two experiments were conducted: one involving 25 participants who were native Spanish listeners and non-native listeners of English, and another involving 18 participants who were native English speakers. In both experiments participants listened to English stories pronounced in FDS, NDS, and a control condition. The temporal response function was used to analyse the tracking of speech envelope (acoustic tracking) and semantic surprisal (semantic tracking). The results indicated Spanish listeners and English listeners showed respectively more efficient cortical tracking of speech envelope in FDS and NDS as compared to the other conditions. These findings suggest that FDS supports the perception of L2 speech in (and only in) non-native listeners. Also, the study highlights the importance of considering the speech register's target audience.

Through this thesis, we contributed to the existing body of knowledge regarding the intricate dynamics of speech production and perception within the context of native and non-native interactions. We found that speakers adjust FDS to support learning of specific L2 skills to learn, that FDS promotes L2 novel word learning (in particular for some vowel contrasts), both in recognition and production. Finally, we found that FDS also promotes perception of continuous speech, as cortical tracking results suggested. All these findings have implications for language learning and teaching, emphasising the significance of tailoring language input to the intended audience.

Resumen en castellano

El objetivo de esta tesis es investigar los mecanismos que subyacen a la percepción y producción del habla dirigida a oyentes no-nativos (o extranjeros). El habla dirigida al extranjero (FDS) es el resultado de la acomodación del habla producida por hablantes nativos de un determinado idioma al dirigirse a oyentes no nativos de ese idioma. Esta acomodación del habla es principalmente inconsciente y puede desempeñar una función didáctica al facilitar la comprensión, percepción y producción del idioma que se está aprendiendo. La adaptación del habla puede manifestarse en diversos aspectos, como cambios en el discurso, la sintaxis, el léxico y lo acústico (que es el enfoque de esta tesis). Al estudiar el FDS, pretendemos explorar la intrincada dinámica de la comunicación entre nativos y no nativos. Además, tenemos hemos analizado las implicaciones de nuestros resultados para los modelos psicolingüísticos de comunicación y aprendizaje de segundas lenguas. El trabajo incluido en esta tesis investiga sobre los procesos implicados en la acomodación lingüística, sus ajustes de las características acústicas resultantes (perspectiva de la producción) y sus implicaciones para una comunicación y una enseñanza eficaces (perspectiva de la percepción).

Para crear unas bases sólidas para nuestra investigación, el primer capítulo se adentra en una exploración de las características clave, las definiciones y las lagunas de la investigación en torno al fenómeno de la FDS. Mediante el examen de la bibliografía existente y la realización de una revisión exhaustiva de estudios anteriores, pretendíamos ofrecer un punto de partida para profundizar en aspectos inexplorados de la producción de FDS y sus consecuencias en la percepción y el aprendizaje de segundas lenguas (L2). Así, en el primer capítulo, identificamos los rasgos acústicos más característicos de FDS, como la hiperarticulación vocálica (triángulo vocálico alargado) y la baja velocidad del habla (ritmo lento). Además, observamos que los estudios anteriores asumían que el FDS se ajusta para satisfacer las necesidades lingüísticas de los oyentes no nativos. En consecuencia, se plantea la hipótesis de que el FDS es un promotor de la adquisición de L2. Concluimos este

capítulo subrayando que los supuestos didácticos del FDS deberían investigarse evaluando si la producción de FDS se adapta a objetivos de enseñanza de diversos aspectos de la L2 (u objetivos comunicativos), como la escritura o la pronunciación (primer experimento). Por otro lado, también señalamos la falta de investigación sobre el impacto del FDS en la percepción (tercer y cuarto experimentos) y el aprendizaje de la L2 por parte de oyentes no nativos (segundo experimento). Estos aspectos se investigaron en los siguientes capítulos y experimentos.

Primer experimento

En el segundo capítulo, pretendíamos entender si las adaptaciones del habla de los FDS varían en función de los objetivos comunicativos del hablante y de los perfiles de los oyentes (es decir, nativos frente a no nativos). Grabamos a hablantes nativos de español nombrando objetos novedosos (que no existen) para ayudar a los oyentes en tareas de comprensión, pronunciación y escritura. Cada hablante interactuó con un oyente nativo (condición Native Directed Speech, NDS) y un oyente no nativo (condición FDS). Se extrajeron medidas de hiperarticulación vocálica, duración, intensidad, ritmo del habla y altura del tono para evaluar los ajustes del habla de los hablantes. Los resultados de la hiperarticulación vocálica mostraron que los hablantes adaptaron su habla en la condición de escritura en comparación con la de comprensión, y en la condición FDS en comparación con la de NDS. Los resultados sobre la altura del tono, la intensidad y la duración de las vocales revelaron que los hablantes proporcionaron señales acústicas ajustadas al perfil y a las condiciones de los oyentes. Por lo tanto, los hablantes ajustan las características acústicas a los objetivos comunicativos en función del perfil del oyente.

Segundo experimento

Tras confirmar nuestra hipótesis de que el FDS se adapta para enseñar aspectos específicos de la L2, en el capítulo 3 exploramos si el FDS facilita el aprendizaje y la producción de palabras de la L2 en comparación con el NDS. Se pidió a 50 participantes españoles que aprendieran 16 palabras nuevas en inglés en un experimento online. Las palabras contenían los contrastes vocálicos /i - ɪ/ o /æ - ɛ/, ambos no nativos para los hispanohablantes. Los participantes se dividieron en dos grupos: uno expuesto a FDS y otro a NDS. Los participantes realizaron tres tareas: a) tarea de reconocimiento, b) tarea de producción, y c) tarea de discriminación continua. En a) se presentaron a los participantes objetos novedosos y sus etiquetas auditivas (por ejemplo, "*This is a deest*", *Esto es un deest*) producidas en FDS o NDS. A continuación, se evaluó su capacidad para reconocer palabras nuevas. En b) se presentaron a los participantes los objetos nuevos y se les pidió que pronunciaran las etiquetas asociadas. En c) se evaluó la discriminación perceptiva de los continuos que incluían las cuatro vocales objetivo. Calculamos la precisión y las latencias de respuesta de los participantes. Además, en b) medimos la precisión fonética de los participantes en la producción de palabras y vocales. Los análisis de la curva de crecimiento mostraron que el grupo FDS reconocía las palabras nuevas más rápidamente y aprendía las palabras del contraste /i - ɪ/ mejor que el grupo NDS. Además, la producción del grupo FDS para el contraste /i - ɪ/ era más precisa que la del grupo NDS. Por el contrario, los dos grupos no difirieron en la discriminación perceptiva de los continuos vocálicos. Estos resultados apoyaron el supuesto didáctico del FDS, que debería tenerse en cuenta en los modelos de enseñanza de L2.

Tercer y cuarto experimentos

Sin embargo, los experimentos con manipulaciones controladas y oraciones cortas/de una sola palabra no reflejan la forma en que los oyentes no nativos reciben la exposición a la L2 (y al FDS). En el capítulo 4, nos esforzamos por explorar cómo afecta el FDS a la percepción y la comprensión en un experimento ecológicamente válido. Al examinar las formas en que el FDS afecta a la interpretación de la lengua hablada, pretendíamos descubrir sus posibles beneficios y limitaciones en la percepción de la L2. Para ello, investigamos los efectos de la exposición a FDS y NDS en la percepción del habla en oyentes no nativos y nativos. Se emplearon estímulos naturalistas en forma de historias en cada registro del habla.

Los estudios conductuales tradicionales suelen examinar el procesamiento del lenguaje después de que se haya producido, mientras que la electroencefalografía (EEG) permite analizar en tiempo real el procesamiento del lenguaje a medida que se desarrolla. Una forma de analizar la modulación del EEG al habla continua es medir el seguimiento o sincronización cortical de las características del habla (como la envolvente). Es decir, el seguimiento cortical de las bandas de baja frecuencia se asocia con el procesamiento del lenguaje, ya que algunos estudios lo identifican como marcadores del procesamiento lingüístico subyacente. Se realizaron dos experimentos: uno con 25 participantes que eran oyentes nativos de español y oyentes no nativos de inglés, y otro con 18 participantes que eran hablantes nativos de inglés. En ambos experimentos los participantes escucharon historias en inglés pronunciadas en FDS, NDS y una condición de control. Se utilizó la función de respuesta temporal para analizar el seguimiento de la envolvente del habla (*tracking* acústico) y la sorpresa semántica (*tracking* semántico). Los resultados indicaron que los oyentes españoles e ingleses mostraron respectivamente un seguimiento cortical más eficiente de la envolvente del habla en FDS y NDS en comparación con las otras condiciones. Estos resultados sugieren que el FDS favorece la percepción del habla de L2 en (y sólo en) oyentes no nativos. Además, el estudio subraya la importancia de tener en cuenta el público objetivo del registro del habla.

Conclusiones

A través de esta tesis, contribuimos al conocimiento existente sobre la intrincada dinámica de la producción y percepción del habla en el contexto de las interacciones entre nativos y no nativos. Descubrimos que los hablantes ajustan el FDS para apoyar el aprendizaje de habilidades específicas de la L2 (por ejemplo, escribir o pronunciar), que el FDS promueve el aprendizaje de palabras nuevas de la L2 (en particular para algunos contrastes vocálicos), tanto en el reconocimiento como en la producción. Por último, descubrimos que el FDS también favorece la percepción del habla continua, como sugerían los resultados de EEG. Todos estos resultados tienen implicaciones para el aprendizaje y la enseñanza de idiomas, y evidencian la importancia de adaptar el input lingüístico al público al que va dirigido. Los resultados indican que la adaptación del habla es un fenómeno en constante actividad y constituye un componente integral de cualquier interacción verbal.

Además, nuestros resultados respaldan algunos modelos psicolingüísticos de comunicación y aprendizaje de segundas lenguas, específicamente encontramos coincidencias con el modelo "Interactive Alignment" (teoría de la alineación interactiva). Este modelo postula que, durante la comunicación, los individuos se esfuerzan por mantener el éxito de la interacción, y sugiere que la alineación es un proceso continuo y dinámico. Estas conclusiones coinciden con nuestros hallazgos.

Por otro lado, nuestros resultados respaldan también los modelos psicolingüísticos del aprendizaje de segundas lenguas, como la teoría sociocognitiva de la adquisición. Según esta teoría, el proceso de aprendizaje de una segunda lengua es un proceso natural y adaptable basado en la alineación ecológica. Nuestro estudio reveló que el aprendizaje de los participantes difería según la adaptación del habla (FDS vs. NDS). Estos hallazgos sugieren que la adaptación del habla FDS desempeña un papel significativo como promotor socialmente mediado en la adquisición de una L2.

Chapter 1 – Introduction: Literature Review

This dissertation aimed to give a well-rounded insight on the phenomena linked to the speech register known as Foreigner Directed Speech. Through this, we explored the processes involved in language accommodation and its consequences for L2 acquisition. To achieve this, Chapter 1 defines Foreigner Directed Speech, its features, and research gaps on this topic.

1.1. What is Foreigner Directed Speech?

Foreigner Directed Speech (henceforth, FDS) is a speech register that native speakers use in interactions with non-native speakers of their language. In recent literature, this register has also been referred to as "L2 speech accommodation". FDS is a broad phenomenon that can encompass changes at the discourse, syntactic, lexical, and acoustic levels (Chaudron, 1979; Long, 1981; Ramamurti, 1980; Uther et al., 2007). FDS is proposed to be a – mostly unconscious – speech accommodation that increases speech clarity and that could serve a didactic function by assisting non-native interlocutors to better understand, perceive, and articulate their L2 (Hatch, 1979; Tarone, 1980; Uther et al., 2007; Scarborough et al., 2007). In this chapter, we provide a critical scoping review of the extensive research investigating the didactic function proposal, focusing on the acoustic features of FDS. We propose that the *didactic function* of FDS comprises two related aspects: a didactic purpose, which is the function of teaching an L2, reflected on the acoustic features of FDS, and a didactic impact, which is the actual effect on L2 perception and learning. In light of this didactic function, we discuss whether FDS serves a purpose in increasing speech intelligibility, facilitating L2 learning, and whether L2 listeners may benefit from being exposed to FDS. The objective of this work is to review those aspects of FDS that are still under debate and to provide strong theoretical and methodological bases for future research into this speech register. An in-depth study of the features and function of FDS is expected to increase our understanding of communication between humans who do not share the same native language. This will enable researchers to build appropriate models of speech communication and social mediation. As linguistic diversity increases worldwide, making communication between native and non-native speakers ever more frequent, speech communication models need to account for FDS.

One of the earliest mentions of FDS as a speech register that serves a linguistic function is found in the 1930's, when Bloomfield (1933) proposed that FDS reflects native speakers' tendency to imitate the mistakes made by non-native speakers in order to assist their speech comprehension. Several decades later, FDS was positioned as a variant of *clear speech*, a term typically used to refer to registers that enhance speech clarity. Other clear speech variants include Infant Directed Speech (IDS, also known as baby talk) and speech directed to elders or to people with hearing impairments (Ferguson, 1975; Lam et al., 2012; see also Smiljanić & Bradlow, 2009 for a review on clear speech). It was only in the 1970's that FDS was identified as an independent speech register that – despite sharing some features with other registers (such as IDS, see Section 1.6.1) – manifests in speech specifically directed to non-native listeners and is uniquely suited to their linguistic needs (Ferguson, 1975). Ferguson (1971, 1975) coined the term *foreign talk* to implicitly compare this register to baby talk, suggesting that the two registers shared a didactic function (Hatch, 1979; Katz, 1977; Tarone, 1980; see also Kuhl et al., 1997 for a discussion on the didactic functions of IDS). FDS was proposed to convey articulatory instructions through a simplified register (as Ferguson, 1981 defined it) characterised by repetition and the use of high frequency words, reduced syntactic complexity, and lack of jargon or idiomatic expressions (Chaudron, 1979; Long, 1981, 1983; Ramamurti, 1980). Additional features, also assumed to assist L2 learners, have been proposed in contemporary studies, including low speech rate, exaggerated voicing of final stops, few vowel reductions, as well as exaggerated intonation and volume (Hatch et al., 1978; Hatch, 1979 reported by Tarone, 1980).

At present, there is more extensive knowledge of FDS, and research largely focuses on the acoustic features of this register. It has been shown that various acoustic features are the result of the FDS accommodation: vowel hyperarticulation, low speech rate, and long pauses are all proposed to help non-native listeners. Given that interest in this topic is growing, there is a need for a review that sums up and discusses the most relevant findings on FDS and sets goals for future research on this topic. In the past, research on FDS has focused on defining the acoustic features of this register; although some FDS features continue to be the subject of debate, future studies should focus on advancing our understanding of the factors that underlie these FDS adjustments, and the role that each FDS feature plays in non-native listeners' L2 acquisition. The present chapter provides the starting point for addressing these issues. This review includes all journal articles and conference proceedings available to date that (1) were written in English and (2) reported empirical findings. These were identified by an extensive literature search using the Google Scholar search engine (search terms: FDS, Foreigner directed speech, Foreign Talk, Foreigner Directed Speech, speech accommodation, listener-oriented speech) and complemented by including relevant references from the articles.

Here, we discuss important aspects of FDS research that help to discover its role in the native–non-native interaction. In Section 1.2, we focus on the acoustic features of FDS to explain *how* FDS improves speech clarity; in Section 1.3 we discuss the emotional valence of FDS that differs for native and non-native listeners. Section 1.4 frames the accommodation theories behind FDS, whereas Section 1.5 and 1.6 discuss whether FDS is adjusted to the listener's needs and thus supports L2 acquisition. Section 1.7 describes research on native and non-native listeners' perception of FDS – which will help us understand whether FDS is useful to non-native listeners. Section 1.8 presents our conclusions and recommendations for future research on this topic.

1.2. The acoustic features of FDS.

1.2.1. Vowel hyperarticulation

Compared to Native Directed Speech (NDS), the register used by peers sharing the same native language who have no need to further enhance intelligibility (Ferguson & Kewley-

Port, 2002; Smiljanić & Bradlow, 2009), FDS is characterised by an expanded vocalic space that is known as vowel hyperarticulation (Uther et al., 2007; Scarborough et al., 2007; Knoll et al., 2007; Knoll et al., 2009a). Most studies on FDS vowel hyperarticulation focus on native speakers' production of the three corner vowels /a/, /i/, and /u/. These vowels are considered important because they are located at the outer boundaries (low, frontal, and posterior) of a language's vocalic space, and they are present in the vocalic inventories of most languages in the world (Bradlow et al., 2003; Ladefoged & Maddieson, 1990). Usually, the averages of the first (F1) and second (F2) formant values are projected onto a two-dimensional cartesian plane to form the corners of the vocalic triangle. An expanded vocalic space corresponds to a vocalic triangle with a larger area (see Figure 1) since the corner vowels are produced at a greater distance from each other. This is proposed to enhance detection of vocalic contrasts and aid speech perception and comprehension (Bradlow & Bent, 2002; Ferguson & Kewley-Port, 2002; Smiljanić & Bradlow, 2005, 2009). F1 inversely relates to vowel height (the lower the vowel articulation, the higher its F1 value) and reflects articulatory effort, which is commonly higher in all clear speech styles than in conversational speech (Ladefoged, 2006; Smiljanić and Bradlow, 2009). F2 is affected by posteriority and lip rounding (F2 values are lower for vowels produced further back in the vocal tract; Ladefoged, 2006). F2 height is usually influenced by vowel hyperarticulation, but it also depends on whether speakers are expanding a front vowel (higher F2) or a back vowel (lower F2) (Ferguson & Kewley-Port, 2002).

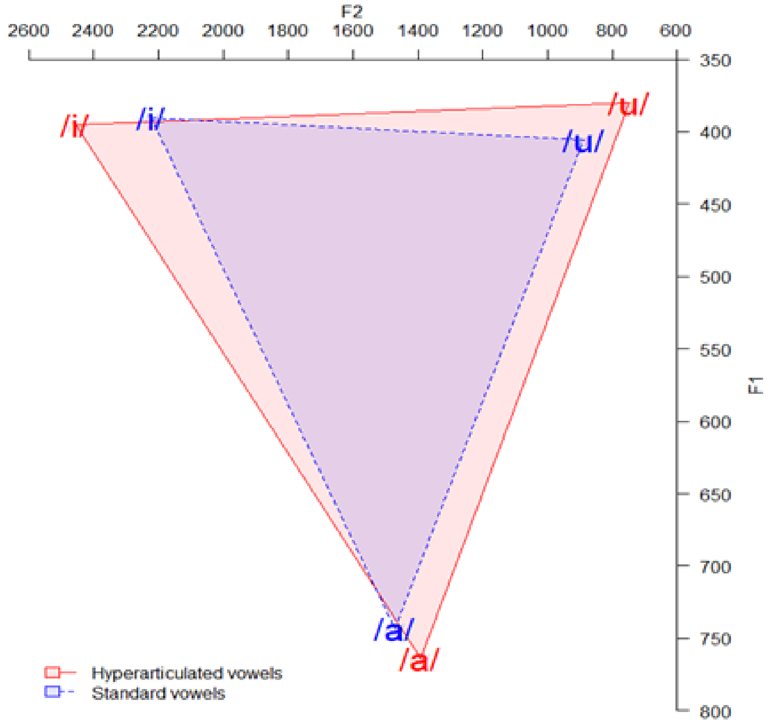


Figure 1. Example of a hyperarticulated vocalic triangle compared to the realisation of standard vowels (not based on real data). The X-axis represents F2 values (Hz); the Y-axis represents F1 values (Hz).

Of particular relevance to this review, vowel hyperarticulation is proposed to be the key acoustic feature that serves a didactic function (both purpose and impact) because it results in a clearer and more distinctive representation of vowel categories (Kuhl et al., 1997). The expansion of the vowel triangle allows speakers to create more discrete categories, thereby avoiding confusion and overlap between vowels and supporting vowel imitation and feature acquisition (Kuhl et al., 1997). Further supporting its proposed didactic purpose, vowel hyperarticulation occurs across clear registers associated with higher speech intelligibility (Bradlow et al., 2003; Krause & Braida, 2004; Picheny et al., 1986), but it is restricted to registers directed to audiences with perceived linguistic capacity (Burnham et al., 2002). For instance, vowel hyperarticulation has been reported in speech to infants (Kuhl et al., 1997) and to computer avatars (Burnham et al., 2010), but not in speech to pets such as cats and dogs (Burnham et al., 2002) – unless the pet is a parrot that “repeats” words (Xu et al., 2013).

Thus, vowel hyperarticulation might also be expected in FDS, since speakers' production is presumably modulated to support the fledgling linguistic abilities of the L2 learner.

We were able to identify 12 studies published to date that have investigated the presence of vowel hyperarticulation in FDS: 8 studies out of a total of 12 reported vowel hyperarticulation in FDS. Most studies focused on English FDS and identified vowel hyperarticulation as the main feature differentiating FDS from NDS (Knoll et al., 2007; 2009a; Uther et al., 2007; Scarborough et al., 2007; Hazan et al., 2015). At the best of our knowledge, only Kendi and Khattab (2019) have reported vowel hyperarticulation in a language other than English, providing some evidence for cross-linguistic generalisation. However, while some previous research had reported both F1 and F2 exaggeration, Kendi and Khattab (2019) found Arabic (Omani) speakers only used F1 to expand their vowel space (consistent with research by Dodane & Al-Tamimi, 2007). This is interesting in light of literature that claims that the degree of vowel hyperarticulation (in clear speech in general) depends on the vowel inventory size of each language. This would suggest that hyperarticulation is language-dependent, and likely to be predominant in languages with large vowel inventories (e.g., Andruski et al., 1999; see also Al-Tamami & Ferragne, 2005). Despite the relatively small number of vowels in Arabic (6 vowels as compared to 14 in English) (Saadah, 2011), Kendi and Khattab (2019) confirmed the presence of vowel hyperarticulation in Arabic FDS. It could be that modulation of hyperarticulation is not fully determined (Smiljanić and Bradlow, 2005) but instead varies in relation to the size of vowel inventories.

Despite the large number of studies confirming that vowel hyperarticulation is present in FDS, some studies have reported different result patterns even for FDS produced in English (Knoll & Scharrer, 2007; Knoll et al., 2009a, 2011a; Kangatharan et al., 2012). For instance, Kangatharan et al. (2012) indicated that they failed to find vowel hyperarticulation in English FDS. This may be because their assessment relied on a vocalic square (instead of triangle) that included vowels /e/, /iː/, /ɔː/ and the diphthong /ai/, which are usually not considered in these kinds of studies. Some of those mixed results derive from specific manipulations of the experimental design (e.g., use of imaginary listeners, see Section 1.5.1, see also Knoll & Scharrer, 2007; Knoll et al., 2009a). Despite such conflicting evidence, given the findings reported above, we conclude that vowel hyperarticulation is a robust feature of FDS. Nevertheless, it remains unclear how this feature relates to the vowel inventory size of a given language. Further cross-linguistic investigations including languages other than English will be needed to determine whether vowel hyperarticulation varies across languages (language-specificity).

In addition to adjustments to vowel formants, vowel hyperarticulation may also be manifested as vowel lengthening. Longer vowels may be easier to process and categorize, vowel lengthening thus having a didactic impact (see Biersack et al., 2005). In order to explore the potential relevance of vowel lengthening in FDS, Uther et al. (2007) analysed whether vowel hyperarticulation of FDS (and IDS) in their study was also associated with longer vowel durations. Instead, they found that vowel length in FDS did not differ from NDS (and was shorter than in IDS). This pattern has been confirmed in other studies where FDS vowel length was not associated with the expansion of vowel space (Knoll & Scharrer, 2007; Knoll et al., 2009a, 2011a; cf. Biersack 2005). Thus, it appears that hyperarticulation of vowels is a clear acoustic feature associated with FDS, but that these vowels are not produced with longer duration. This confirms the proposal from Biersack et al. (2005) that FDS should give non-native listeners more time to process sentences by lengthening pauses, but not vowels.

Some attention has also been dedicated to hyperarticulation of non-vowel phonological contrasts. For instance, Sankowska et al. (2011) were the first to find that the plosive durational difference (between voiced and voiceless consonants) was larger in FDS than in either NDS or Lombard Speech (LS), a register produced to help listeners cope with background noise. This finding suggests that hyperarticulation in FDS might not be relegated to vowel articulation only. Similarly, several studies have reported hyperarticulation of lexical tone categories in Foreigner (Papoušek & Hwang, 1991) and Infant Directed Speech (Han et al., 2018; Liu et al., 2007; Xu et al., 2013) in lexical tone languages. Lexical tones are based primarily on modulations of pitch height and contour, and their realisation is not independent of segments (tones are carried by vowels as well as the adjacent consonants, so they can be considered supra-segmental; Burnham et al., 2011). However, similarly to consonant and vowel segments, lexical tones mark phonemic contrasts, and so it is not surprising that lexical tone categories are also exaggerated in clear speech registers.

1.2.2. Low speech rate and long pauses

Another characteristic acoustic feature of FDS is low speech rate, measured as an increase in pauses between utterances and the duration of individual words within utterances (Ferguson, 1975; Biersack et al., 2005; Scarborough et al., 2007; Kangatharan, 2015; Lorge & Katsos, 2019; Bobb et al., 2019). FDS tends to have a lower rate of words per minute than NDS (Biersack et al., 2005; Hatch, 1979; Nelson, 1992; Rodriguez-Cuadrado et al., 2018; Scarborough et al., 2007). Similar to vowel hyperarticulation, this phenomenon is proposed to benefit comprehension and processing of speech by non-native listeners: L2 listeners might benefit from having more time to parse, segment, and analyse linguistic information when speech rate is slower (see Biersack et al., 2005).

While vowel hyperarticulation across languages awaits further research, speech rate adjustments in FDS have been explored directly in cross-linguistic studies (e.g., English vs. French). Warren-Leubecker & Bohannon (1982) and Hazan et al. (2015) reported that English FDS has a lower word rate per minute than NDS. Kühnert and Antolík (2017) provided evidence from French using a *tandem* paradigm: participants with two different native languages (L1s) practised language exchange to help each other learn an L2; each participant was a native speaker of the L2 the other participant wished to learn. Using this paradigm in English and French, Kühnert and Antolík (2017) found that French native speakers accommodated their production to the English listeners (French L2 learners) by slowing down their speech rate. However, native English speakers did not significantly lower their speech rate when they interacted with the French (English L2). This finding on English speakers contrasts both with the results for the French participants in this study and with previous studies on English FDS. The authors explained this incongruency by suggesting that speech tempo adjustments may be language specific, possibly related to the faster natural speech rate in French than English, or to French participants having higher proficiency in English than their English counterparts had in French. In fact, slower speech was directed to L2 listeners with lower proficiency (English native speakers), whereas the faster speech rate was directed to highly proficient L2 listeners (French native speakers). This suggests that low speech rate may be a feature of FDS directed to naïve L2 learners in order to support their L2 comprehension.

1.2.3. Intensity and Pitch

Several suprasegmental features including acoustic intensity, pitch height, pitch range, and pitch contours have also been investigated in FDS. It remains unclear whether enhancement or exaggeration of these features serve independent didactic purposes or occur as by-products of the phonetic exaggeration already noted in this register. For instance, vowel hyperarticulation may be accompanied by increased intensity and heightened pitch whereas the independent exaggeration of either of these two features may not result in enhanced clarity.

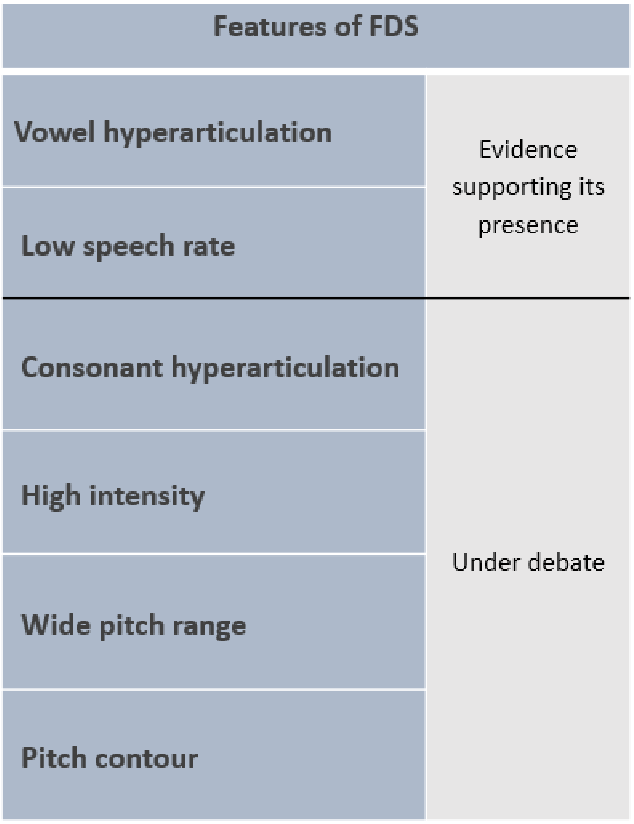
Intensity corresponds to the amount of energy carried by sound waves, and loudness is its primary perceptual correlate. High intensity is a prosodic cue for emphasis that in FDS might correlate with vowel hyperarticulation. Vowel hyperarticulation might be the result of articulatory effort (see Lindblom, 1990), while intensity and loudness are secondary correlates of effort (see also Smiljanić and Bradlow, 2009). Rodriguez-Cuadrado et al. (2018) analysed Spanish FDS by measuring the intensity of repeated words, which are usually hypoarticulated in conversational speech. They observed higher word intensity for repeated words in FDS than NDS. Kendi and Khattab (2019), in their study on Arabic FDS, also reported significantly higher vowel intensity in FDS than NDS (in line with Hazan et al., 2015; cf. Knoll et al., 2015). Thus, higher intensity could be a relevant feature of FDS, but the evidence for this claim to date is not robust, and no studies have directly tested whether it is an acoustic correlate of another FDS feature, specifically vowel hyperarticulation (see Ferguson & Quené, 2014 for similar a hypothesis regarding clear speech).

Pitch is a suprasegmental feature that is used to mark prosody conveying prominence and/or information structure. Pitch is the perceived acoustic product of the vibration rate of the vocal cords (Ladefoged, 2006). Pitch range corresponds to the maximal/minimal excursions of pitch (i.e., the difference between pitch *Max* and *Min*), whereas pitch contour is the curve of the perceived pitch change over time. Very few studies have investigated pitch range in FDS, and existing studies have reported both wider pitch excursions compared to NDS (Smith, 2007) and comparable pitch ranges for the two registers (Knoll et al., 2015). It would not be surprising to find an expanded pitch range in FDS since emphatic pitch excursion is proposed to form part of hyperarticulatory phenomena, stressing relevant words and assisting word segmentation (as research on IDS suggests: Fernald & Kuhl, 1987; Thiessen et al., 2005).

Although exaggerated pitch contour is linked to pitch range, the two features are not equivalent. Specifically, the same pitch range value could be associated with both bell and rising contour shapes and vice versa. In fact, despite some evidence for a wider pitch range in FDS than NDS (Papoušek & Hwang, 1991; Smith, 2007), the few experiments that assessed pitch contour reported little evidence for exaggerated pitch contours in FDS (Papoušek & Hwang, 1991; Knoll et al., 2006, 2007; Knoll & Costall, 2015). Knoll et al. (2006) studied contour shape and found that FDS did not contain exaggerated shapes compared to NDS (Knoll et al., 2007). Most contours in FDS were level shapes (flat) or falling contours like those found in NDS (Knoll & Costall, 2015). Further qualitative analysis in Knoll and Costall (2015) highlighted the fact that the participants (students) sometimes produced FDS with rising contours, which are similar to the contour shape of questions (Fernald, 1989; Knoll et al., 2006, 2007; Knoll & Costall, 2015). However, it is likely that native speakers were trying to assess the L2 listener's comprehension so as to adapt their production accordingly and used a rising contour typically associated with a questioning tone (a silent “did you understand?”) to implicitly interrogate listeners’ comprehension. This strategy invites the listener to provide continuous feedback, either verbally or nonverbally (e.g., through nods, confused expressions; see section 1.5.3 for a discussion of listeners’ feedback). This interpretation of the rising contour converges with the reported results on pitch range in IDS, which show mostly bell shape contours. The main hypothesis is that exaggerated pitch contour and wide pitch excursion serve the functions of emotional transfer and requesting attention (Ferguson, 1971; Papoušek & Papoušek, 1981; Trainor & Desjardins, 2002). The results on FDS fit with this view since speakers are not expected to employ FDS to convey emotions but rather to draw the listener’s attention to meaningful words.

Mean pitch exaggeration corresponds to raised fundamental frequency (F_0), which is found in other clear speech registers such as IDS. Exaggerated mean pitch is mostly interpreted as a strategy to convey emotion and a non-threatening attitude in speech (Ohala, 1984a), and for this reason this feature is not expected to be prominent in FDS. In fact, the many studies sustain that FDS does not have a high pitch correlate (Biersack et al., 2005; Bobb et al., 2019; Knoll et al., 2009a, 2011a; 2011b; Lorge & Katsos, 2019; Uther et al., 2007). For example, Biersack et al. (2005) and Uther et al. (2007) found no significant difference between FDS and NDS. It is important to note that although some studies underscored the absence of heightened mean pitch in FDS, other studies have disclosed a different pattern of results. In recent work, Kendi and Khattab (2019) demonstrated that pitch average midpoints were higher in FDS than NDS (Hazan et al., 2015). However, this significant increase in mean pitch in FDS might be due to specific aspects of their study design. In Kendi and Khattab (2019), the addressees of FDS were domestic helpers who had been working for the participants (i.e., speakers) for an extended period of time (from 2 months to 4 years). Indeed, previous evidence has shown that familiarity and emotional closeness can alter speech realisation (Bänziger & Scherer, 2005; Costa et al., 2008; Farley et al., 2013), and this could have resulted in exaggerated pitch in speech produced in the interactions between these dyads. On the other hand, Knoll & Scharrer (2007) and Knoll et al. (2015) found a similar effect with higher mean pitch in FDS than NDS, which was not due to familiarity between interlocutors. Still, procedural differences may play a role. The 2007 study used a specific procedure, in which participants had to imagine non-native listeners and speak *as if* they were addressing them (see Section 1.5.1). In the 2015 study, all the participants had just been speaking to people with hearing loss. This could have elicited higher pitch and resulted in carry-over effects when they switched to using FDS.

In summary, this section reviewed evidence on vowel hyperarticulation, low speech rate, high intensity, and high pitch correlates in FDS (see Figure 2 for a summary of the FDS features). Most research has found vowel hyperarticulation and low speech rate in FDS. Conversely, there is less evidence of intensity and different pitch features (range, contour, and mean) being different between FDS and NDS. FDS mainly employs flat contours, but in some cases rising contours occur, resembling the contours of interrogative utterances. FDS may be characterised by a wide pitch range, which could reflect hyperarticulation, although little research effort has been dedicated to exploring this feature. Lastly, several studies failed to report higher mean pitch in FDS in comparisons to NDS, although some supportive evidence has been noted (e.g., Kendi & Khattab, 2019).



Features of FDS	
Vowel hyperarticulation	Evidence supporting its presence
Low speech rate	
Consonant hyperarticulation	Under debate
High intensity	
Wide pitch range	
Pitch contour	

Figure 2. Summary of the features of FDS. “Evidence supporting its presence” means that there is evidence in favour of this feature. Features reporting “Under debate” mean that there is still little or mixed evidence.

1.3. Emotional valence of FDS

The emotional or affective properties of FDS have also received attention in the literature. In studies exploring this question, participants were asked to listen to continuous speech samples and rate how positive or negative they sounded. The emotional valence of a speech signal is a complex combination of several acoustic features including, but not limited to, speech tempo, pitch height and range, and intensity (see Liscombe et al., 2003; Tursunov et al., 2019). While the perceived affect of a speech register can be directly related to speakers’ desire to transmit emotion or to their communicative intent, it can also be a by-product of the exaggeration of prosodic and acoustic components intended to enhance a register’s clarity or its didactic purpose. As discussed below, these components can elicit negative perceptions from listeners, which in turn can have an effect on the register’s effectiveness as a linguistic tool. Critically, perception of the emotional valence of FDS appears to vary depending on the linguistic profile of the listener, so next we separately consider studies in which ratings were provided by either foreign or native listeners.

Non-native listeners. Bobb et al. (2019) identified a positive correlation between median pitch in FDS and ratings of speakers' competence and friendliness (see Lynch, 1988). In this study, FDS consisted of sentences read to an imaginary audience rather than naturally produced speech to a foreigner. FDS was compared to other speech registers including NDS, clear speech, and IDS, yielding several interesting results. FDS was perceived as friendlier than NDS. IDS contained the highest level of median pitch, followed by FDS, clear speech, and NDS (IDS > FDS > clear > NDS), suggesting that positive emotional affect is (at least partially) driven by pitch height. Each register was produced by native speakers, but naïve (foreign) raters – who were not aware of speakers' language background and proficiency – considered speakers in the FDS condition to be overall less competent than speakers in the generic clear speech condition. On the other hand, speakers who produced higher median pitch in their FDS were rated as more competent and less condescending. Bobb et al. (2019) concluded that intelligibility did not positively correlate with perceived condescension, meaning that speaking clearly does not entail sounding condescending and patronising.

Native listeners. Native listeners are expected to have different perceptions of FDS than the intended non-native audience because they would not derive any linguistic benefit from this register (unless they hear it in challenging listening conditions). That is, a positive or negative evaluation of FDS may depend on whether the listeners feel themselves to be the intended and appropriate addressees for the register adopted (Austerlitz, 1956; Ferguson, 1975). DePaulo and Coleman (1981) recruited 91 native English listeners to rate communications directed to non-native listeners (as well as to infants, adult native listeners, and people with cognitive disability). These participants perceived FDS as less friendly, less respectful, and less encouraging than NDS, but they considered speakers using FDS to be more competent than speakers of all the other speech registers. Surprisingly, in a later study using the same methodology and a combination of measures, DePaulo and Coleman (1987) instead found that FDS was considered to be warmer than NDS. The authors stressed that listeners displayed a remarkable ability to recognize differences between registers even without any explicit/external cues as to the identity of the addressees (see also Knoll et al., 2011a). This indicates that FDS is clearly differentiated from other registers, suggesting that it serves a communicative function and conveys psychological and sociological information (DePaulo and Coleman, 1981).

In more recent work, Uther et al. (2007), Knoll and Scharrer (2007), and Knoll et al. (2011b) used low-pass filtered segments of vocal interactions between two native speakers of English, a native speaker and their infant, a native speaker and a non-native confederate (a native Chinese speaker) to elicit ratings of negative and positive vocal affect from naïve native listeners. This band filter removed all frequencies above 1000Hz rendering speech unintelligible, while leaving prosodic and emotional features unaffected (Scherer, 1971; Scherer et al., 1972; Starkweather, 1967), so that raters had to rely on acoustic features for their emotional evaluations, without considering semantic content. In Uther et al. (2007) and Knoll and Scharrer (2007), FDS received the lowest ratings for positive vocal affect and the highest ratings for negative affect (see also Knoll et al., 2009a), whereas in Knoll et al. (2011b), FDS received lower rating than NDS for positive vocal affect only. In addition, Knoll et al (2009b) tested the consistency of emotional ratings across various band filters, in addition to the 1000 Hz cut-off, to measure the contribution of higher frequency bands to rating scores. They found that across most filter levels (except the unfiltered and the 1200Hz filter version), NDS was rated as having higher positive vocal affect than FDS, but it was considered to request the same level of attention.

A number of acoustic parameters may be responsible for the different impressions made by different registers, especially the negative perception of FDS by native listeners. For instance, it has been argued that speech rate (Stewart, Bouchard-Ryan, 1982; Knoll et al., 2009a), vowel hyperarticulation (Uther et al., 2007), and pitch (Knoll et al., 2015) modulate the degree of negativity associated with FDS (see Rothermich et al., 2019). Rothermich et al.

(2019) reported that in Uther et al. (2007) more hyperarticulated vowels in IDS got more positive ratings, whereas in FDS greater vowel hyperarticulation corresponded to more negative ratings. High pitch and wide pitch range also seem to play a role in eliciting positive emotional evaluations (cf. Knoll et al., 2015). Knoll et al. (2009b) found that the positive vocal affect of IDS decreased as low-pass band filters excluded higher frequencies, therefore reducing the contribution from the high pitch cue. We know that high pitch is a typical feature of IDS, hence it is likely that the exclusion of higher frequencies was the cause of the reduction in the perceived positive vocal affect in this register. This indicates that high mean pitch of IDS has some influence over its positive affect. However, it is worth noting that in Knoll et al. (2009b) pitch was higher in NDS than FDS, and this could have accounted for its higher positive vocal affect ratings and partially account for the rating discrepancy between these two registers (Biersack et al., 2005; Uther et al., 2007).

Most experiments with native raters have found that FDS is perceived negatively, even when semantic content is obscured by using various band filters. However, studies to date have not assessed listeners' beliefs about the intended audience for each register that they were asked to rate. We suggest this assessment should be included in future perceptual studies: listeners' ratings may be influenced by their perceptions of the register used and its intended audience. For instance, when native raters think that FDS is addressed to them, they might find it condescending and rate it negatively; conversely, if they believed it was addressed to someone else (especially a non-native listener), they might rate it more positively. Since FDS is a register directed to adults, it could be used to convey a disrespectful message to native listeners and be negatively rated for this reason (see Starkweather, 1967; Clyne, 1981). It is likely – but this proposal needs deeper investigation – that low speech rate and vowel hyperarticulation in the absence of a positive emotional contribution from high mean pitch both play a role in eliciting negative judgments of FDS from native raters.

1.4. The theory behind FDS.

The discussion above indicates that speakers adjust their speech when addressing non-native listeners in a way that is proposed to assist speech processing and comprehension. Now, we turn to the theories that have tried to account for these accommodations in FDS. Earlier theoretical approaches hypothesised that FDS was an example of a simplified register, largely characterised by syntactic and lexical simplifications (Henzl, 1973; Tweissi, 1990). FDS was assumed to be the result of a communication strategy determined by cultural rules (see Ferguson, 1975; Canale & Swain, 1980; Tarone, 1980). More recently, FDS has generally been interpreted as the result of speech accommodation by L1 speakers who want to maintain successful communication with L2 listeners (Giles et al., 1991; Hazan et al., 2015; Scarborough et al., 2007; Smith, 2007; Snow et al., 1981; Zuengler, 1991; see also Costa et al., 2008). This view assumes that speakers are sensitive to the addressee's need to receive linguistic clarifications and learn phonological contrasts.

The Hyperarticulation & Hypoarticulation (H&H) theory of speech accommodation (Lindblom, 1990) is the main theoretical framework adopted to interpret FDS research findings in the recent literature. This theory sustains that the main source of speech variability is accommodation to listeners, situations, and contexts. According to this theory, speakers continuously regulate their speech production along the hypo-/hyper-articulation continuum, in order to meet their communicative aims, listeners' demands, and to maintain successful communication. The articulation continuum spans the range from least to most effortful articulation, where the least effort is put into interaction with peers (i.e., NDS), following the natural tendency to save articulatory energy as much as possible without losing category distinctiveness (effort-based approach to Optimality Theory, Kirchner, 1997; Theory of Adaptative Dispersion, Lindblom, 1990; Diehl & Lindblom, 2004). The Communication Accommodation Theory (CAT) offers an alternative but similar perspective that accounts for previous experience with interlocutors, non-linguistic cues (such as smiling), and adoption of listeners' communication behaviours (Beebe & Giles, 1984; Coupland et al., 1988; Giles, 2016; Giles et al., 1991; Zhang & Giles, 2017). The latter element is part of a so-called *convergence* strategy, which enhances communication success through modifications of segmental and suprasegmental properties. The opposite strategy is called *divergence*, and is designed to maintain social distance by eschewing speech adjustments and demonstrating indifference to effective communication (for additional theoretical frameworks, see Nekvapil & Sherman, 2015; Wooldridge, 2001; Zuengler, 1991).

Both the H&H and CAT speech accommodation frameworks entail that most FDS adjustments are regulated by didactic intentions. There is, therefore, widespread theoretical consensus on the didactic purpose of FDS (Biersack et al., 2005; Smith, 2007; Margić, 2017; Rothermich et al., 2019; Bobb et al., 2019). However, as we discuss in later sections of this review, despite this consensus, there is little direct evidence establishing that FDS is effective in achieving its proposed didactic impact, and whether any positive effects associated with this register actually enhance non-native listeners' subsequent L2 perception or production. In short, there is a pressing need for further research into the didactic intentions, functions, and impacts of FDS.

The next section of this review focuses on the factors that can influence or modulate acoustic adjustments and the proposed didactic purpose of FDS described in previous sections. According to these theoretical accounts, FDS speech accommodations are based on the listener, the speaker, the situation and communicative context. We discuss the effects of each of these factors on the properties of FDS below.

1.5. Factors that influence speech accommodation.

1.5.1 Listener characteristics

Language proficiency and accentedness. Language proficiency and the accentedness of the listener may be the first factors that influence the level of accommodation in FDS and its acoustic realisation. To our knowledge, no systematic investigations of the effect of listener's proficiency on FDS are available, but previous studies allow us to speculate that perceived proficiency modulates the extent to which native speakers are inclined to adjust their speech for L2 listeners. For instance, Snow and collaborators (1981) suggested that the perceived language proficiency of the non-native listener (in addition to their perceived social status and intelligence) is responsible for the magnitude of FDS effects (see also Gaies, 1979; Chaudron, 1978; Dahl, 1981; Liu, Kuhl, & Tsao, 2003 for evidence on IDS). Kühnert and Antolík (2017) results on different speech rate adaptations made by native French and English speakers (presented in Section 1.2.2) suggest accommodation differences across speakers largely depend on the listener's language proficiency. In fact, in this study, English listeners had about 3 years less L2 experience than their French counterparts (6.4 vs. 9.2 years, respectively). The authors acknowledged that this might be the reason why only French participants adapted their speech rate to help their listeners. Furthermore, other studies have found low speech rate in English FDS (Warren-Leubecker & Bohannon, 1982; Hazan et al., 2015), suggesting that Kühnert and Antolík's null result for English FDS was likely due to listener characteristics rather than a cross-linguistic difference in adaptation between English and French. In sum, speech rate may vary as a function of an addressee's proficiency and future research should aim to assess correlations between the foreigner listener's proficiency and the native speaker's speech rate in English and other languages.

Several studies suggest that accentedness has a tight negative correlation with language proficiency (Gallego, 1990; Kang et al., 2010; Munro & Derwing, 1995), meaning that a strong foreign accent corresponds to a low level of proficiency. While some low-proficiency addressees may have less obvious foreign accents (Munro & Derwing, 1995), it is still the case that speakers may be biased to interpret a strong L2 accent as a symptom of low proficiency, and consequently adapt their register to help these listeners (see Kang et al., 2010). To date, no experiments have explored proficiency versus accentedness in orthogonal designs; instead, experiments have relied on the generic perception of the confederate's strong foreign accent without differentiating between accent and proficiency (e.g., Uther et al., 2007). Furthermore, previous research has not provided many details or objective measurements of listeners' accentedness and/or proficiency with few exceptions (Lynch, 1988; Hazan et al., 2015; Kendi and Khattab, 2019). In the only study to specifically address accentedness, Lorge and Katsos (2019) asked a Greek confederate to emphasise her foreign accent while producing grammatically correct speech in English, yet even in this case, no measure of resulting accentedness was reported. In this study, speakers hyperarticulated vowels only when the confederate simulated a strong foreign accent. This suggests that accentedness has some influence on the properties of FDS.

These findings stress the importance of using consistent measurements across studies that allow for direct comparisons of proficiency and accentedness. If the listener's proficiency level affects the realisation of FDS properties (i.e., speech rate) independently from accentedness, this would be strong evidence that this register has a generic didactic purpose. Conversely, if accentedness only drives FDS features, or some of them, FDS likely provides articulatory information to help highly proficient L2 listeners who nevertheless retain a strong foreign accent.

Perception of foreignness. Another listener characteristic that may influence FDS production is the perception of foreignness. This construct is largely linked to listeners' physical appearance, which can lead speakers to assume they are foreigners with linguistic difficulties even when there is no evidence of strongly accented L2 speech or low L2 proficiency. Of course, it is noteworthy that a speaker's L1 and language background (e.g., bilingualism) are not directly related to their physical aspect, but some studies find that perception of foreignness is influenced by physical appearance, even when that is not justified by the interlocutor's language identity or proficiency (Bernstein et al., 2007; Ito et al., 2004; Rubin, 1992). If physical appearance (only) drives hyperarticulation in FDS (see Long, 1983), then the register would not serve a didactic purpose but rather reflect an intention to emphasise social distance and linguistic superiority (Biersack et al., 2005; Clyne, 1981; Valdman, 1981). Results from FDS ratings (Uther et al., 2007; Hazan et al., 2015) seem to be in line with this idea because native listeners might perceive FDS to be disrespectful and to have lower intelligibility than other clear registers (see Valdman, 1981 for a similar hypothesis). Ratings made by native listeners might be based on their perception of an imbalance in speaker-listener interactions; that is, adjustments that are not made to accommodate listeners, but rather due to prejudices about perceived foreignness.

To address this issue, Kangatharan et al. (2012) calculated the vowel hyperarticulation of a square vocalic area of speech directed to foreign-looking and native-looking confederates with or without foreign accents. They observed that the accent of the addressee did not induce vowel hyperarticulation, but the addressees' foreign appearance did. Surprisingly, native-looking listeners with foreign accents did not elicit any acoustic adjustments in speech. This finding, however, was not replicated by the same team in a study with different target vowels and a larger sample size (Kangatharan et al., 2015). Here, the physical aspect of the listeners had no effect on FDS, whereas a foreign accent elicited hyperarticulated vowels (Arthur et al., 1980). In order to shed light on this matter, future research on FDS is required to consider possible confounding factors derived from speakers' biases toward different ethnicities. At present, most studies conducted on FDS did not clearly state whether their confederates had both foreign physical appearances and foreign accents (i.e., Uther et al., 2007; Hazan et al., 2015; Kendi & Khattab, 2019), which hinders the distinction between the influence of these two factors. Based on this limited evidence, it appears that physical appearance alone, namely the perception of foreignness, is not sufficient to elicit FDS. However, further evidence that directly compares appearance and accentedness is required.

Imaginary addressees. Many studies have opted to use 'imaginary listeners' by eliciting FDS in the absence of a live non-native listener (Papaousek and Hwang, 1991; Biersarck et al., 2005; Scarborough et al. 2007; Knoll and Scharrer, 2007; Knoll et al., 2009a, 2009b, 2011a, 2015; Bobb et al., 2019). This approach is advantageous in terms of controlling interactions, by reducing the individual differences that inevitably arise from live face-to-face interactions. Still, this design choice prevents researchers from controlling for factors such as accentedness and language proficiency, which as discussed above, can influence the realisation of FDS, and may explain some of the contradictory findings in the current literature. The validity of these paradigms may also be questioned: the interaction may not be as natural as one with a live interlocutor present and each participant might imagine a different ‘stereotypical’ foreigner, possibly depending on their previous personal experiences (Snow et al., 1981). Any of these imagined differences could elicit different degrees of acoustic adaptation in FDS.

In fact, findings from studies that employed imaginary addressees to elicit FDS have been mixed. Some studies have reported adjustments similar to those reported in the presence of real addressees (Biersack et al., 2005; Bobb et al., 2019; Scarborough et al., 2007), while others report no differences from NDS (Knoll et al., 2009a, 2011a; Knoll & Scharrer, 2007). Scarborough et al. (2007) compared the results of interactions with imaginary and real non-native listeners and found that the imaginary listeners elicited low speech rates to a greater degree than real addressees (see also Knoll et al., 2009a). This demonstrates that real and imaginary audiences can elicit unequal manifestations of FDS features (Knoll & Scharrer, 2007; Knoll et al., 2009a). It seems likely that imaginary addressees lead to inauthentic speech modifications, and that such adjustments likely vary across participants due to their own performance abilities (i.e., actresses vs. students in Knoll et al., 2009a), experience with L2 speakers, or the instructions they received on the experimental task (Snow, 1981; Lam et al., 2012; Knoll et al., 2011a). To counteract some of these issues, further research involving fictitious listeners could consider employing simulations of live interactions, i.e., making participants believe they are talking to a real foreigner. This option still does not require actual addressees, making it practically feasible, and in turn has several benefits. A simulation, for instance, guarantees stable comparisons across participants, thanks to the standardisation of the fictitious listener’s behaviour (see Buz et al., 2016). It also allows the researcher to control various factors such as the physical appearance and accentedness of the listener, simultaneously.

1.5.2 Speaker characteristics

Production of FDS acoustic features seems to be speaker dependent. In fact, Knoll et al. (2011a) compared students and actresses' FDS production, and observed that, with the same amount of exposure to non-native listeners, the actresses hyperarticulated vowels more than the students (see also Knoll et al., 2009a). In addition, the nature of the relationship between interlocutors might also induce speakers to tailor their speech to listeners' needs.

Previous experience and bilingualism. Experience communicating with L2 listeners is one factor that appears to increase speakers' sensitivity to listeners (Snow et al., 1981), and makes them more likely to accommodate their speech (i.e., to use FDS). For instance, language teachers, who are used to dealing with L2 learners' difficulties, might be particularly prone to employ effective speech adjustments, which would result in specific adaptations in their speech production matched to their students' L1 phonological inventory. Another population that has been studied in this regard are bilingual speakers. Lorge & Katsos (2019) found that bilinguals hyperarticulated vowels more than monolinguals in FDS. This suggests that bilingualism shapes FDS, and that individuals who are immersed in multilingual environments, and who may have been L2 learners themselves, are more prone to use this speech register.

Emotional closeness, familiarity, and relationship. Speakers may differentially adjust their speech depending on the nature of their relationship with their interlocutor. People are likely to behave differently with elders as compared to same-age peers, with people whom they know, or have a close relationship to outside the experimental context such as romantic partners (Bänziger & Scherer, 2005; Caporael, 1981; Farley et al., 2013; Kemper et al., 1995). Young people or caregivers often overaccommodate their speech when talking to elders or people with disabilities, conveying condescension (Coupland et al., 1988; Ryan et al., 1994; Ryan et al., 1986). This demonstrates that both age and familiarity shape the relationship between interlocutors and influence their speech adjustments. Kendi and Khattab (2019) used foreign addressees (age not reported) who had been working for the participants of the study (the native speakers) for at least 2 months. Speakers and listeners knew each other, some for up to four years. Native speakers produced FDS characterised by wider vocalic space, limited to F1 movement, and higher pitch than NDS. This finding is in line with studies on articulation and pitch modulation in speech addressed to lovers or intimate friends. This suggests that social aspects of the relationship between interlocutors, like distance or closeness (i.e., relationships with superiors or peers) may be relevant to the way in which FDS is delivered. In line with this possibility, other work on FDS has employed strangers as listeners, and did not find higher mean pitch, but instead observed an expanded vocalic triangle manifested as both F1 and F2 exaggeration (Uther et al., 2007; Kangatharan et al., 2015). Note that, as far as we know, the latter set of studies used English as the target language, which may limit the generalisation of these findings to other languages. Future studies should probe the influence of different types of relationships between interlocutors on FDS, especially in languages other than English, and consider age, social distance, emotional closeness, and immersion in a multilingual environment as factors that potentially play a role in shaping FDS features.

1.5.3 Situational and contextual factors

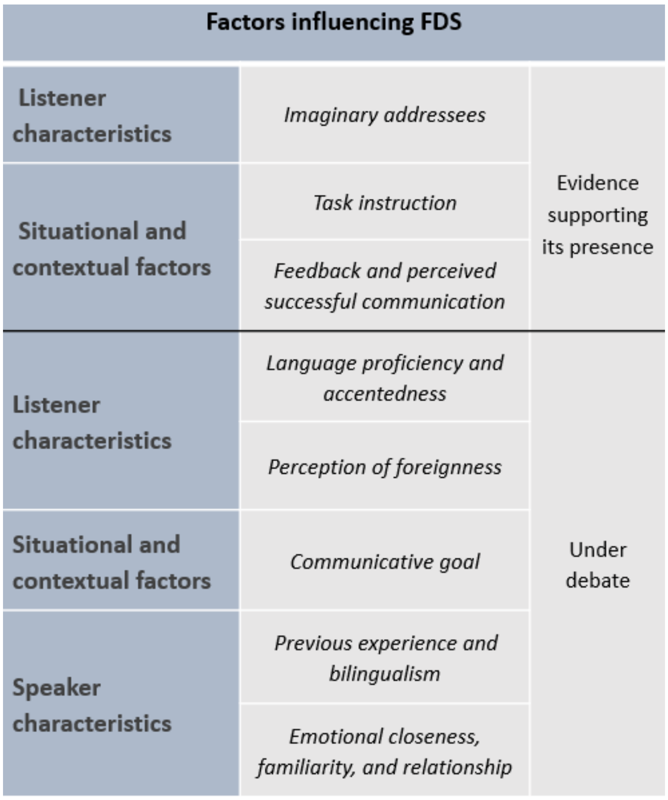
Situational factors: task instruction and communicative goal. Situational factors influence the properties of FDS and include, but are not limited to, the instructions given to the speakers to elicit FDS (Knoll et al., 2011a), and the purpose of the conversations in which they use FDS. There are numerous potentially important situational elements, and they are often tightly bundled together, making it difficult to disentangle their effects. At present there is still little evidence for the influence that experimental tasks have on the acoustic properties of FDS, but it is likely they have important effects. For instance, previous research indicates that the instructions used to elicit a clear register directly influence its realisation (Lam et al., 2012; see also Smiljanić & Bradlow, 2009; Knoll et al., 2011a). Knoll et al. (2011a) investigated FDS features by employing “simulated free speech” and “standardised sentences”. In the former task, speakers were provided with three target toys (a shark, a sheep, and a shoe) so that they could invent their own scenarios (centred on the target words) to address imaginary listeners. In the latter task, speakers used fixed sentences like ”Look at the ‘target word’” to address the same imaginary listeners (e.g., “Look at the shark”). The authors observed differences in the acoustic features elicited in the two tasks and concluded that the reproduction of some FDS features depends on the task employed. Hence, the type of task may induce peculiar speech modifications and communicative strategies; for instance a task where participants have to read aloud (as in Bobb et al., 2019) likely results in FDS with peculiar phonetic characteristics that are different than those of spontaneous speech tasks (see Blaauw, 1992; Hazan & Baker, 2010; Laan, 1992). A *tandem* situation where two interlocutors practise language exchange (as in Kühnert & Antolík, 2017), provides for a free ranging, natural, conversational situation. By contrast, requiring a speaker to give directions to a listener over the phone (as in Smith, 2007) entails a strictly defined situation and a possibly hierarchical relationship between interlocutors. In the former case, the target audience probably feels freer to ask for repetitions when the speaker’s enunciation is not sufficiently clear. There are other factors, not directly related to listeners or speakers themselves, that are not fixed, but rather vary dynamically. For instance, it may be hypothesised that speakers provide slightly different acoustic cues depending on the immediate goals of communication; for instance, if they are using FDS to teach the proper pronunciation of phonemes or the spelling of ambiguous words. These different didactic goals may require different approaches to situational difficulties and possibly different articulatory strategies.

Contextual factors: feedback and perceived successful communication. Feedback from listeners during communication may induce speakers to dynamically regulate their speech rates or increase the perceptual distance between ambiguous phonemes in a specific word. In fact, feedback from the interlocutor seems to be fundamental in eliciting hyperarticulatory phenomena in communicative interactions (Buz et al., 2016; Maniwa et al., 2009; Ohala, 1984; Stent et al., 2008; see also Smith & Trainor, 2008 for evidence on IDS).

Two studies demonstrate that speakers make dynamic adjustments to their speech properties in response to interlocutors' feedback (Burnham et al., 2010; Buz et al., 2016). Burnham et al. (2010) observed that speakers hyperarticulated vowels when addressing a computer avatar (rather than a human), who repeated their sentences, and they did so to a greater extent after the avatar pronounced their sentence incorrectly than when it did so correctly (see also Schertz, 2013 for a study on exaggerated contrasts after listener's misheard speech). Buz et al. (2016) observed similar results in a simulated native-native interaction where speakers hyperarticulated plosive consonants after receiving negative feedback, and then maintained this alteration across several trials. In fact, some speech adjustments emerge only if speakers perceive that successful communication is useful to the listener (Kuhlen & Brennan, 2013; Lockridge & Brennan, 2002). In short, feedback appears to shape speech registers because it induces the production of clear features such as vowel hyperarticulation.

With regards to FDS, only one study to date has assessed the role of feedback on its realisation. Warren-Leubecker and Bohannon (1982) directly explored the role of feedback on the online adjustment of speech during native speaker and non-native listener interactions. They found that regardless of non-natives' L2 proficiency levels, lower FDS speech rates were mostly driven by feedback indicating that communication had failed. This result, together with previous findings from NDS studies, suggests that feedback may play a significant role in FDS production (see Suffill et al., 2021 for evidence on lexical alignment with non-native listeners). Alternatively, negative feedback could demonstrate comprehension difficulties, indicating the listener has low proficiency. In this case, FDS might be elicited mainly as a function of the listener's language proficiency level. To test these possibilities, future research should focus on disentangling main effects and the interaction between feedback and proficiency in order to define the role of each of those factors on FDS production. Warren-Leubecker and Bohannon's (1982) evidence on the feedback mechanism supports the H&H hypothesis (see Section 1.4) that FDS serves a didactic purpose in response to a listener's linguistic needs. However, future research should aim to establish whether didactic purpose and feedback determine the acoustic changes in FDS to a similar degree and whether their effects can indeed be disentangled to provide a more precise explanation for the observed properties of FDS.

In Section 1.5, we showed that FDS features are influenced by multiple factors, some related to listeners and others to speakers (see Figure 3 for a summary). FDS is the result of a complex set of factors, which include, for instance, adaptations to low proficiency listeners and the nature of the personal interaction between interlocutors. Lastly, we discussed that speech is adapted according to feedback from listeners in line with the goals of successful communication. The information presented in Section 1.5 indicates that FDS is not a static register, but rather adapts to situations, context, and interlocutors' interactions. The dynamic nature of FDS is strictly bound to its didactic purpose, and Section 1.6 provides a discussion of FDS features in relation to other clear speech registers that will further help to understand this aspect.



Factors influencing FDS		
Listener characteristics	*Imaginary addressees*	Evidence supporting its presence
Situational and contextual factors	*Task instruction*	
	Feedback and perceived successful communication	
Listener characteristics	*Language proficiency and accentedness*	Under debate
	Perception of foreignness	
Situational and contextual factors	*Communicative goal*	
Speaker characteristics	*Previous experience and bilingualism*	
	Emotional closeness, familiarity, and relationship	

Figure 3. Summary of the factors influencing FDS realisation. "Evidence supporting its presence" means that there is evidence in favour of this factor. Factors reporting"Under debate"mean that there is still little or mixed evidence.

1.6. Differences and commonalities between FDS and other clear speech styles

A comparison between FDS and IDS suggests that both registers serve a didactic function: they are produced to enhance language acquisition. By contrast, the clear features of Lombard Speech cannot serve this function; this register simply reflects the need to communicate clearly in a noisy environment. But similar acoustic adjustments might result from different underlying purposes, adaptation to listeners' needs, and communicative goals.

1.6.1 Comparing FDS and IDS

IDS is the speech register that adults use in interactions with young infants (Golinkoff et al., 2015). It has a number of linguistic, emotional, and acoustic characteristics that differentiate it from Adult Directed Speech (ADS, which is equivalent to NDS), including simplified grammar (Soderstrom, 2007), warm positive affect (Kitamura & Burnham, 2003), changes in speech timbre (E. A. Piazza et al., 2017), low speech rate (Panneton et al., 2006), exaggerated pitch height and range (Fernald et al., 1989), and acoustic exaggeration of vowels (Kuhl et al., 1997). Vowel hyperarticulation in IDS (Burnham et al., 2015; Cristia & Seidl, 2014; Kalashnikova & Burnham, 2018), has been proposed to serve a specific linguistic function, similar to FDS. Compared to ADS, caregivers using IDS significantly expand the acoustic space between the three corner vowels /i/, /u/, and /a/. This is proposed to result in clearer speech that helps infants discriminate the phonetic categories of their native language and later reproduce them in their own vocal tract. However, this proposal has been debated in IDS literature (Cristia, 2013). First, while vowel hyperarticulation in IDS has been reported for a number of languages including English (Adriaans & Swingley, 2017; Burnham et al., 2002; Kalashnikova & Burnham, 2018), Russian and Swedish (Kuhl, 1997), Spanish (García-Sierra et al., 2021), and Mandarin Chinese (Liu et al., 2009), it has not been detected in Dutch (Benders, 2013), German (Audibert & Falk, 2018), or Norwegian (Englund & Behne, 2005)

IDS. Second, vowel categories in IDS are more variable than those in ADS, so despite the expansion of the space between corner vowels, overall vowel clarity is reduced, and non-corner vowel categories are less discriminable (Cristia & Seidl, 2014; Martin et al., 2015; McMurray et al., 2013). This evidence has led to the proposal that vowel hyperarticulation in IDS does not facilitate language acquisition but is instead a by-product of other affective adjustments made in this register such as changes in voice quality and smiling (Benders, 2013; Miyazawa et al., 2017).

Hence, it is possible that the acoustic exaggeration of vowels observed in IDS and FDS stem from different speaker intentions and articulatory mechanisms. Kalashnikova, Carignan, and Burnham (2017) provided the first direct evidence for this possibility. In their study, mothers spoke to an adult in a typical manner (ADS), to an adult in a clear and exaggerated manner (exaggerated speech, ES), and to their infant (IDS) while their tongue and lip movements were measured using electromagnetic articulography. Acoustic analyses of maternal speech indicated that ES and IDS contained more hyperarticulated vowels than ADS. However, mothers exaggerated their tongue movements, that is, actually hyperarticulated speech, only in ES and not IDS. Acoustic exaggeration of vowel F1 and F2 in IDS, was instead explained to result from the significantly greater reduction in the size of the vocal tract through laryngeal raising in IDS compared to both ES and ADS. This adjustment is typically observed when a speaker wants to appear smaller and less threatening. The authors proposed that the acoustic exaggeration of vowels in IDS may have originated as a by-product of a maternal intent to sound friendly and comfort infants. However, while not originally aimed at facilitating infants' language development, this 'accidental' component of IDS may have acquired a secondary linguistic function.

In line with this proposal, there is evidence that vowel exaggeration is modulated by infants' linguistic and processing needs, and that infants benefit from this component of speech input. First, reduced vowel exaggeration has been reported in IDS to infants who are unable to hear their mothers' speech (Lam & Kitamura, 2012), or when infants are at-risk for a language processing disorder such as dyslexia (Kalashnikova, Goswami et al., 2018; 2020). Thus, mothers appear to adjust the vowel properties of their IDS to the specific needs of their infant audiences. Second, hyperarticulated vowel sounds elicit more mature neural responses and more successful sound discrimination in nine-month-old infants (Peter et al., 2016) and facilitate word recognition in 19-month-olds (Song et al., 2010). Critically, these relations are observed at the level of individual mother-infant dyads: mothers who exaggerate vowels to a greater extent in their IDS have infants with more advanced speech perception skills (Kalashnikova & Carreiras, 2021; Liu, Kuhl, & Tsao, 2003) as well as larger concurrent and future vocabularies (Hartman et al., 2017; Kalashnikova & Burnham, 2018; Lovcevic et al., 2020).

The prosodic components of IDS, namely slow rate, pitch height, and pitch range also facilitate speech processing and lead to positive language acquisition outcomes in young infants (Spinelli et al., 2017). This is interesting given that these speech components are typically associated with the affective function of the register and are not consistently found in other clear speech registers including FDS (Biersack et al., 2005; Uther et al., 2007; Knoll et al., 2009a, 2011a; 2011b; Lorge & Katsos, 2019; Bobb et al., 2019). For instance, speech stimuli with the prosodic properties of IDS have been shown to facilitate infants' neural encoding of speech (Kalashnikova, Peter, et al., 2018; Zangl & Mills, 2007), vowel discrimination (Trainor & Desjardins, 2002), segmentation of continuous speech (Thiessen et al., 2005), and word learning (Graf Estes & Hurley, 2013; Ma et al., 2011). Adults also benefit from these properties as they are more successful at learning novel words when they are produced in IDS than in ADS (Golinkoff & Alioto, 1995; Ma et al., 2020).

As can be seen, some but not all IDS components overlap with FDS, and these similarities and differences have been used to support the claim that these components can occur independently of each other and are dynamically adjusted according to the specific emotional and linguistic needs of each audience (Burnham et al., 2002; Kalashnikova, Goswami, et al., 2018; Uther et al., 2007). However, more recent research has identified more nuanced similarities and differences between these two registers that help us better understand their possible didactic functions and roles in facilitating language acquisition and processing. It appears that vowel hyperarticulation and low speech rate are manifested to a similar degree in FDS and IDS (Uther et al., 2007; Lorge & Katsos, 2019; Martin et al., 2016). The main difference between the registers consists of the lack of pitch exaggeration in FDS compared to IDS, particularly with regards to the exaggeration of pitch contours (Knoll et al., 2015) and overall pitch height (Uther et al., 2007).

This review suggests that IDS and FDS share several components that may assist speech processing and language learning in their intended audiences. Infants benefit from the acoustic components of IDS when they occur in isolation or in unison, and there is some evidence that these components can also lead to processing benefits in adults. However, the presence of individual components in IDS is modulated by infants' age and linguistic experience. It is plausible that similar effects due to language proficiency can be observed in FDS. In fact, neglecting the importance of L2 proficiency may have led to inconsistent findings regarding the individual components of FDS and how they facilitate L2 perception and learning.

1.6.2 Comparing FDS and LS

Lombard speech (LS) is a register elicited when speakers have to counter background noise (Lombard, 1911). Compared to NDS, its characteristic articulatory and acoustic features include loudness, articulatory effort, low speech rate, and hyperarticulation (Garnier et al., 2006; Garnier et al., 2018; Sankowska et al., 2011; Hazan et al., 2015). Most of these features are shared with FDS, including loudness, low speech rate, and hyperarticulation (Hazan et al., 2015; Sankowska et al., 2011), but research has also uncovered several key differences. With regards to loudness, some research reported FDS to be louder than NDS, as we noted in Section 1.2.3. However, to the best of our knowledge, there is one study that has compared intensity across LS and FDS and it found no significant difference between the registers

(Hazan et al., 2015). Whereas the difference between FDS and NDS could be predicted, the lack of distinction between FDS and LS is surprising. In fact, we would instead expect LS to be significantly louder than FDS since the latter is not specifically intended to overcome noise. This predicted difference also aligns with accommodation theories, which predict speakers adjust to better accommodate listeners' needs.

Sankowska et al. (2011) explored other aspects of speech that distinguish FDS from LS. In this study, the authors found that FDS emphasises phoneme duration contrasts that help distinguish short from long speech sounds, whereas LS emphasises duration differences less than FDS. Hazan et al. (2015) compared NDS, FDS and two acoustic barriers, namely, vocoded and noisy speech. Compared to NDS, speakers modified their speech across all other conditions, but in the vocoded condition they lowered speech rate, lengthened words, and hyperarticulated vowels more than in FDS. As can be seen, LS and FDS share similar acoustic features, which are manifested to different extents. Specifically, the available results to date suggest that LS uses more hyperarticulated vowels and slower speech rates (lengthened words) than FDS, but FDS uses length contrastively to highlight phoneme differences to a greater extent than LS.

LS and NDS share a native audience, but LS is a clearer register produced to counteract interference that lowers the intelligibility of the message for native listeners, who otherwise (without interference) would understand it perfectly. In fact, LS is only designed to overcome acoustic interference; the addressee faces no linguistic difficulty and has no need to learn the language. In short, LS does not have a didactic purpose. This is in line with the differences described in the features of LS and FDS and the perception of LS by both native and non-native listeners. In fact, Cooke & Lecumberri (2012) discovered that non-native listeners are not able to take advantage of LS clarity like native speakers (see also Bradlow and Benet, 2002 for a similar effect on clear speech), suggesting that LS and FDS fulfil different functions.

The comparison between FDS and LS shows that the aims of communication and addressees are crucial for eliciting speech registers. The didactic function of FDS is not limited to hyperarticulation, given that LS also exhibits this feature, but it does not serve a didactic purpose. The fact that both FDS and LS are characterised by low speech rates and vowel hyperarticulation does not make them similar versions of clear speech. Rather, speech registers result from the sum of various factors such as the speaker’s intention and the specific communicative goal (e.g., to overcome linguistic difficulties in the case of FDS vs. noise in the case of LS). These factors, together with the addressee’s linguistic needs and identity, seem to be the most relevant factors in eliciting specific speech styles and their respective acoustic features (see Knoll et al., 2015 for similar results on FDS vs. speech directed to people with hearing impairments).

In Section 1.6, we discussed the differences and commonalities between FDS and two other clear registers. By comparing and contrasting acoustic features, we established that FDS and IDS are both likely to serve didactic purposes that nevertheless differ in some respects. In fact, we saw that the origin of the didactic function of these two registers is regulated by the specific needs of their audiences: addressee identity plays an important role in defining the characteristics and purpose of each register. As for FDS and LS, the evidence suggests that the two registers have highly similar acoustic features (loudness, low speech rate and vowel hyperarticulation), but that specific manifestations of these features may derive from different speaker intentions and listener needs. Loudness in LS is justified by its need to overcome background noise, which is not the case in FDS. Perceptual studies would help to untangle the differences and similarities between these registers in an objective manner. Ratings of clarity and other types of subjective ratings could help advance our understanding of the differences between these registers and their purposes (see Rothermich et al., 2019). With this in mind, the next section turns to existing research that has investigated the perception of FDS by native and non-native listeners.

1.7. Perception of FDS

Without the appropriate level of speech accommodation, non-native listeners experience frustration and lose interest in L2 learning (Zuengler, 1991; Kemper et al., 1995; Margić, 2017). The appreciation of FDS may depend on whether it meets their needs without being either overaccommodating or patronising (Perdue, 1984; Coupland et al., 1988; Lindblom, 1990). According to the didactic view of FDS and our discussion above, both affect and clarity perceptions of FDS depend on the non-native listener's L2 proficiency (Chaudron, 1978; Dahl, 1981; Snow et al., 1981; Xu et al., 2013). However, studies on listener's perceptions of FDS are scarce and most focus only on the emotional valence of FDS using listener affective ratings (as discussed in Section 1.3). On the other hand, very few studies have focused on FDS intelligibility, or on how clear and useful L2 listeners consider FDS to be. Here we point out the most relevant results regarding the perceived clarity of FDS first by non-native listeners and then by native listeners.

Non-native listeners. Congruent with theories of accommodation and the didactic function hypothesis (Lindblom, 1990; Uther et al., 2007), non-native listeners tend to rate FDS as being clearer than NDS, possibly because FDS meets their needs for language learning (Hazan et al., 2015). In Bobb et al. (2019), participants had to assign clarity scores to FDS and NDS without knowledge of what register they were hearing; the non-native listeners rated FDS as clear speech, whereas NDS was rated as less intelligible. In Kangatharan et al. (2015), early and late L2 learners listened to samples of FDS and NDS with low to high levels of added noise and assessed their clarity by using a Likert scale. All participants perceived FDS to be clearer than NDS regardless of their L2 proficiency, and an interaction between noise level and speech register showed that FDS clarity was less affected by noise than NDS. Such results are crucial because they demonstrate that FDS is sharply differentiated from NDS (Depaulo & Coleman, 1981; Knoll et al., 2011a), and that it possibly boosts L2 intelligibility for non-native listeners. This is in line with the finding that non-native listeners do not consider LS to be as clear as native listeners do (Cooke & Lecumberri, 2012), supporting the assumption that LS lacks any didactic function. It appears that non-native listeners perceive FDS to be clearer than LS since only the former is intended to meet their linguistic and communicative needs. It is important to underline that, although most general features of FDS (e.g., low speech rate) likely enhance speech clarity for non-native listeners of any L1, it is also probable that this clarity effect partially depends on the non-native listeners' L1 and on whether production of FDS is oriented to accommodate listeners of that specific language.

The studies reported above indicate that FDS supports speech clarity at any level of L2 proficiency, but all conclusions are based on subjective survey ratings, and objective measurements of actual speech processing and comprehension are missing. Neuroimaging techniques would provide a way to obtain measurements that do not depend on raters' metacognitive skills and would directly assess the benefits of FDS. The only study to date that has employed this approach is Uther et al. (2012) who used electroencephalography (EEG). They recorded event related potentials (ERPs) derived from the perception of hyperarticulated words and measured mismatch negativity (MMN), which is an index of auditory discrimination. To assess whether vowel hyperarticulation helped L2 listeners to discern vowel contrasts, native and non-native listeners were tested in a word listening task, in which words were produced with either standard or hyperarticulated vowels. Results showed that the phonetic changes were detected regardless of the listener's language status: MMN was elicited by hyperarticulated vowels in both native and non-native listeners. This finding leaves open the questions of whether non-native listeners benefit from hyperarticulated vowels to perceive L2 phonemic contrasts and whether FDS enhances L2 perception as compared to NDS. In fact, the non-native participants had a high level of proficiency (about 9 years of L2 use) and were living in the country where their L2 was used at the time of the experiment. Hence, non-native participants had likely already acquired the phonological contrast used in the experiment, and they did not need hyperarticulation to aid its detection. Therefore, the question of whether L2 listeners benefit from vowel hyperarticulation remains open and requires further research, especially on non-native listeners with low levels of L2 proficiency.

Native listeners. Kangatharan et al. (2015) also asked native listeners to rate how clear they found FDS and NDS (with or without noise). Like non-native listeners, native listeners perceived FDS to be clearer than NDS, indicating that FDS is indeed a type of clear speech. In line with this finding, Hazan et al. (2015) explored native listeners' perception of clarity in FDS compared to NDS and Lombard Speech, which, as discussed above, exhibits a similar degree of vowel hyperarticulation as FDS but lacks its proposed didactic purpose. To do this, the authors used naturally elicited LS, FDS, and NDS (e.g., LS was elicited in a native-native conversation with added noise) and calculated the number of words produced by speakers to complete the task as an index of communicative difficulty. Hazan et al. (2015) reported that the speakers experienced greater communicative difficulty in the FDS compared to the LS condition (see Knoll et al., 2011a for similar results). Nevertheless, naïve raters who listened to those conversations in the absence of noise considered FDS to be clearer than NDS but less clear than LS. Hazan et al. (2015) offered the interesting explanation that register features that make speech clearer do not merely depend on the level of communicative difficulty. Such results support our discussion in Section 1.6.2 (comparing FDS and Lombard speech) by demonstrating that difficult listening conditions *per se* and the need for clarity are not sufficient for eliciting FDS.

Understudied aspects of FDS perception. In Hazan et al. (2015), native listeners and non-native listeners with low and mid-level proficiency rated FDS to be clearer than LS. Crucially, all three groups considered FDS to be clearer than NDS (as in Kangatharan et al., 2015), suggesting that FDS is perceived differently (and perceived to be clearer) at all levels of language proficiency. However, the lack of difference among the three listener groups, most importantly between the native and non-native listeners, does not offer support to the hypothesis that FDS has a didactic impact. In fact, research has not yet addressed whether FDS enhances language acquisition for L2 learners or whether, on the contrary, the sole way to improve L2 perception and production is exposure to native and peer to peer register (Margić, 2017). That is, no perceptual studies to date have explored the effects of FDS perception on L2 learning directly. As suggested above, research must address the effects of FDS exposure on non-native listeners' speech processing in order to understand its actual role in the process of L2 acquisition. If there is evidence that this register performs a didactic impact, non-native listeners would be expected to gain greater benefits from listening to FDS than native listeners, and to learn more when exposed to FDS compared to NDS. Note that perceptual ratings may also fail to highlight such differences because of intrinsic limits to subjective evaluations. One possibility is to expand research on FDS perception with neuroimaging techniques that provide more objective measurements of the phenomenon and to use them in combination with behavioural methods. We discuss this possibility in the next section.

1.8. Future directions and conclusion

1.8.1. Open questions

This chapter reviewed the evidence for the FDS function of increasing speech intelligibility and facilitating L2 learning, by considering the main acoustic features and the factors influencing them. Low speech rate and vowel hyperarticulation were identified as the main features of FDS. We also examined research on additional acoustic features in FDS, such as wide pitch range and high intensity, which are still debated in the literature. Evidence revealed that FDS is a register based on listeners' identities, communicative needs and goals, and situational factors, such as the instructions provided for performing an experimental task. This suggests that FDS has a didactic purpose. Our discussion was grounded in the leading theoretical frameworks that account for the acoustic properties of FDS and supported by comparing FDS to two other clear speech registers, IDS and LS. We also reviewed empirical literature that has assessed the perception of FDS by native and non-native listeners, which yielded the following main findings. First, FDS is positively perceived only by non-native listeners, who also rate it to be clearer than NDS. Although native listeners rated FDS negatively, they still consider it to be clearer than NDS. This consensus further backs up the status of FDS as a clear speech register that is easily differentiated from NDS (Hazan et al., 2015; Uther et al., 2007). Second, FDS reduces vowel ambiguity in speech, which may provide listeners with useful information on how to perceive foreign sounds and (perhaps) produce them. Finally, clarity ratings of LS – a clear register meant to overcome communication noise – highlight that non-native listeners give LS lower clarity scores than native listeners. Taken together, this evidence from clarity ratings suggests FDS has a didactic impact in contrast to LS and NDS. However, further work is required to produce conclusive evidence for these possibilities and to understand whether non-native listeners benefit from FDS in the process of L2 acquisition. In this section, we discuss the open questions that we consider to be the most relevant directions for future research.

The first open question regards the typical acoustic characteristics of FDS (see Section 1.2), and whether these features are universally present in this register. For instance, it is not clear if speech rate is a feature of FDS present across languages, given that some research has shown mixed results for French and English (Smith, 2007; Kühnert & Antolík, 2017). Other features such as wide pitch range have simply not been adequately investigated in languages other than English to assess their universality. Further investigation is also needed on speaker status to determine whether speaker’s identity (see Section 1.5) plays a role in eliciting FDS; it remains unclear whether all speakers produce FDS features to a comparable extent. Such factors could, for instance, include gender, based on evidence from one study (Lorge & Katsos, 2019) showing that women tend to hyperarticulate speech more than men. Another and highly interesting factor to investigate is speakers’ bilingual status. It is plausible that bilinguals are better than monolinguals at adapting their speech to L2 listeners and at responding to audience needs and feedback (Lorge & Katsos, 2019). Moreover, it is probable that bilinguals who are also L2 teachers are particularly good at adapting their speech to their students’ L1. Therefore, future experiments should investigate FDS production in bilinguals, expanding research in this field to languages other than English, and employing research designs that control for additional factors such as interlocutors’ identities, listener’s feedback, and adopt ecologically valid interactions for elicitation of FDS.

Another outstanding question relates to the independence of the acoustic features of FDS. That is, one important venue for future research is to explore whether all the acoustic features of this register systematically co-occur, or whether they manifest independently from each other, serving different purposes for the speaker and the addressee. This is an important theoretical point for a better understanding of all clear registers, namely FDS. A parallel with other audience-oriented styles may point to an answer to this question (see Section 1.6). For instance, if we turn to developmental changes in IDS, we notice that several properties of IDS undergo drastic reshaping as the baby grows. In fact, pitch and speech rate in IDS seem to be adjusted to the infant’s increasing age and linguistic ability, and become more adultlike in the second and third years of the child’s life (Narayan & McDermott, 2016). Importantly, unlike other features, vowel hyperarticulation in IDS does not vary with infants’ age, possibly reflecting the infants’ continuing need to acquire language (Kalashnikova & Burnham, 2018; Liu et al., 2009). This suggests that all the typical features of IDS do not need to co-occur and manifest to similar extents. This may also be the case for FDS. If future studies confirm the same pattern of vowel hyperarticulation and speech rate adjustments in FDS based on the characteristics of the listener (i.e., L2 proficiency and accentedness), this will reinforce the link between IDS and FDS and their didactic purposes.

Relatedly, the possibility that FDS may be characterised by a continuum of speech adjustments and accommodation requires attention in future research. Is FDS an on/off register, or does it occur on a continuum that goes from no adaptation (when the foreigner has a high level of phonological and linguistic competence) to the maximum grade of speech modifications with naïve L2 listeners? The results reported in Section 1.3 suggest that the latter may be the case, and that speech rate and vowel hyperarticulation might be modulated as a function of listeners’ L2 proficiency. This aspect could be clarified via longitudinal studies on L2 acquisition and exposure to FDS, from naïve learners to proficient speakers. Alternatively, this could be achieved via a study that makes native speakers address non-native listeners with various levels of proficiency.

1.8.2. Research gaps addressed in this thesis

In addition to the topics described in the previous section, there are further open questions that are relevant to uncover the phenomena related to FDS. Here, we introduce these questions, which will be the main anchor points for the different studies of this thesis.

FDS features may be subject to other contextual variables, apart from those described above, such as the nature of the interaction, speaker familiarity, and communicative context. We hypothesise that contextual factors, such as listeners' feedback and communicative goals, are similarly relevant for determining the acoustic modifications in FDS, as described in Section 1.5. Evidence that communicative goals shape FDS, and that feedback due to miscomprehension induces further exaggeration of FDS features is essential for strengthening the claim that FDS serves a didactic purpose. In addition, that would help confirm that FDS is the outcome of the speakers' unconscious goal to teach phonological contrasts to a non-native audience (Uther et al., 2007). This will be the scope of Chapter 2. We will explore the acoustic features of Spanish FDS and investigate whether FDS production adapts to communicative goals, such as the intention of teaching certain language skills (e.g., writing and pronouncing L2 words).

Another open question relates to the impact of FDS on non-native listeners. The similarities between FDS and IDS, in their acoustic features, suggest a didactic impact for FDS too. However, it is unknown so far whether non-native listeners who are exposed to FDS benefit in terms of language learning (i.e., phonemes perception and pronunciation), as is proposed for IDS. Also, does this assumed benefit expand to both L2 perception and production learning? In Chapter 3 we will address these questions, by assessing whether non-native listeners' L2 learning is promoted by the exposure to FDS in both perception and production of L2 novel words.

One last remaining question is whether FDS promotes perception of continuous speech. Combined behavioural and electrophysiological designs offer an avenue for answering many open questions on FDS, like this one. The use of EEG provides insight into the cognitive processes involved in listening to and interpreting hyperarticulated vowels and the effects of low speech rate on the brain activity of L1 and L2 speakers and listeners. As a complementary measure, a comprehension questionnaire of FDS by both native and non-native listeners will provide information on intelligibility, which is useful for interpreting the electrophysiological data. Thus, in Chapter 4 we investigated whether FDS promotes non-native listeners' L2 perception and comprehension by means of an EEG study.

1.8.3. Conclusion

To summarise, FDS likely boosts non-native listeners' speech perception and comprehension. This is probably due to speakers' accommodating listeners' linguistic needs and results in adjustments such as vowel hyperarticulation and low speech rate. Nevertheless, further evidence that speakers adapt FDS to factors such as listener proficiency, listener feedback, and the specific aims of communication, is required to confirm theories that propose that FDS supports speech accommodation. Crucially, a deeper understanding of the factors that influence FDS production and perception is relevant for models of second language acquisition and can inform theoretical and practical approaches to second language instruction. If FDS serves a didactic purpose, it is imperative to assess how it benefits non-native listeners' perception and/or production of L2 phonological contrasts, and more generally, helps them learn their L2. Establishing that naïve L2 listeners appreciate and learn better when exposed to FDS would suggest that FDS is an important tool for second language teaching and for understanding all listener-oriented registers.

Supplementary material

A supplementary table summarising designs and findings of the published studies that have assessed vowel hyperarticulation, speech rate, or pitch correlates of FDS can be consulted in Appendix 1.1.

Chapter 2 - Speakers' Communicative Intentions Lead to Acoustic Adjustments in Native and Foreigner Directed Speech.

2.1. Introduction

2.1.1. Factors influencing acoustic features of speech

Speech is characterised by various acoustic features, such as speech rate, pitch height, and intensity. Various factors influence speech production and these acoustic features. These factors include, but are not limited to, the listener's linguistic profile, task instructions, and listener's feedback (Buz et al., 2016; Piazza, Martin, et al., 2022; Uther et al., 2007). As we discussed in the previous chapter, a long line of research has shown that speakers accommodate acoustic features to their listeners' profile (i.e., being a native, a non-native listener, or an infant). For instance, adult speakers tend to produce speech with hyperarticulated vowels, slow speech rate and exaggerated pitch when addressing infants (known as infant-directed speech: Kalashnikova & Burnham, 2018; Knoll et al., 2006; Kuhl, 1997; Trainor & Desjardins, 2002; Uther et al., 2007). Speakers also adapt the acoustic features of their production when addressing computer avatars (Burnham et al., 2010) and even pets (Xu et al., 2013). The type of listener-oriented speech adaptation that is the focus of this thesis is Foreigner Directed Speech (FDS), which occurs when native speakers address non-native listeners (Hazan et al., 2015; Kendi & Khattab, 2019; Uther et al., 2007). In the first chapter, we showed that FDS is a clear speech register that increases speech clarity, and that is assumed to serve a didactic purpose by assisting non-native listeners to better perceive, understand, and articulate their second language (L2) (Biersack et al., 2005; Bobb et al., 2019; Piazza, Martin, et al., 2022; Rothermich et al., 2019; Scarborough et al., 2007; Uther et al., 2007). In recent literature, FDS has also been referred to as "L2 speech accommodation" (Cheng et al., 2021). As implied by this term, speech adjustments are proposed to be the result of the native speaker's accommodation to meet their listener's linguistic needs and to improve communication effectiveness (Lindblom, 1990; Giles, 2016, Uther et al., 2007). FDS is usually described in comparison with Native Directed Speech (NDS), which is the register used among native speakers with no intention of enhancing speech clarity (Ferguson & Kewley-Port, 2002). Compared to NDS, the FDS accommodation is expected to result in more hyperarticulated vowels and reduced speech rate (Hazan et al., 2015; Scarborough et al., 2007; Uther et al., 2007). Even though FDS has been interpreted as a case of accommodation to the communicative goal of teaching an L2, there is a lack of evidence on this didactic aspect of FDS. FDS has been investigated as a fixed factor, derived from the presence/absence of a non-native listener. But, to build more realistic models of speech communication, it is also important to understand whether speakers dynamically adjust FDS to specific skills being taught or communicative needs.

Adjustments to the acoustic features of speech do not only occur in speech to language learners (e.g., FDS or infant-directed speech), but also in speech to native adult listeners, following listener feedback, and according to specific task instructions. Buz et al. (2016) and Warren-Leubecker & Bohannon (1982) reported that speakers adapt their speech to listener feedback regarding the success of the communication. Both studies found that acoustic adjustments, such as slower speech rate and consonant lengthening, were driven by feedback of failed communication. That is, feedback may induce speakers to dynamically regulate their speech to enhance intelligibility (Burnham et al., 2010; Maniwa et al., 2009; Stent et al., 2008). Moreover, certain speech adjustments only emerge if speakers perceive that communication helps listener's performance (Kuhlen & Brennan, 2013; Lockridge & Brennan, 2002). That is, acoustic features are different in live than in imaginary interactions, when speakers *act as if* they were addressing real listeners (Scarborough et al., 2007). For example, Knoll et al. (2009) assessed the validity of simulated FDS in a study involving students and actors. For both groups, they found that the simulated interaction did not elicit vowel hyperarticulation, which is proposed as one of the core acoustic features of FDS (Hazan et al., 2015; Kendi & Khattab, 2019; Uther et al., 2007).

Another factor that influences speech production is the experimental task speakers are asked to perform (Knoll et al., 2009, 2011; Lam et al., 2012). Knoll et al. (2011) compared speech to imaginary and real addressees as speakers performed two tasks: one in which they had to spontaneously describe three toys (a shark, a sheep, and a shoe), and another in which they produced fixed sentences containing the names of the toys (e.g., ”Look at the shark”). These two tasks resulted in different levels of adjustments of acoustic features in speech to imaginary vs. real addressees. For example, in the imaginary addressee condition, speakers´ pitch height and vowel duration were more similar to natural interactions with real listeners in the standardised sentences task than in the free speech task.

The results described above demonstrate that speech is dynamically adapted to various factors and specific needs of the listeners. In line with this, speakers may adjust certain acoustic features in relation to specific communicative goals (e.g., teaching how to pronounce or spell a word). One such communicative goal, which is the focus of this study, is the intention of delivering didactic information through speech, specifically teaching how to pronounce or spell a word. Piazza et al. (2022) hypothesised that communicative goals together with situational/contextual factors (i.e., task instructions, adaptation to feedback) influence the acoustic features of FDS, which are used as cues to meet the listeners’ needs and promote successful communication (Knoll et al., 2011). It is possible that speakers adjust their speech differently when they are teaching a word, and according to whether they are teaching the listener to pronounce or spell that word, independent of the listener´s linguistic profile. However, to the best of our knowledge, no research to date has directly investigated this aspect of speech production. For this purpose, we examined whether speech adjustments vary according to specific communicative goals, and whether such adjustments are further modulated by the listener´s linguistic profile (FDS vs. NDS).

2.1.2. Target acoustic features of speech

This study focuses specifically on the acoustic features that have been most reliably identified in clear speech registers, namely vowel hyperarticulation, lengthening, and intensity increase, as well as speech rate reduction and pitch height raising. Previous studies examining whether and how these features vary in FDS compared with NDS are briefly reviewed below.

Vowel hyperarticulation refers to the acoustic expansion of the vowel space by producing the three corner vowels /a/, /i/, /u/ with greater distance than standard vowels (vowels produced without intention to enhance clarity). Vowel hyperarticulation is usually measured on the averages of the first (F1) and second (F2) formant values of the three corner vowels (/a/, /i/, and /u/), which are projected onto a two-dimensional cartesian plane to form the corners of the vocalic triangle (Kuhl, 1997; Uther et al., 2007). FDS has more hyperarticulated vowels than NDS (Hazan et al., 2015; Lorge & Katsos, 2019; Scarborough et al., 2007; Uther et al., 2007). This FDS feature has been proposed to have a didactic function, being a promoter of language learning that enhances detection of vocalic contrasts (Ferguson & Kewley-Port, 2002; Kuhl, 1997; Smiljanić & Bradlow, 2005, 2009) and possibly improves both perception and production of L2 phonemes (Piazza et al., 2022).

Another acoustic feature characteristic of FDS is speech rate reduction. FDS tends to have a lower rate of words per minute compared to NDS due to the elongation of vowels and pauses (Biersack et al., 2005; Bobb et al., 2019; Rodriguez-Cuadrado et al., 2018; Scarborough et al., 2007). This feature is proposed to enhance processing and comprehension of speech in non-native listeners who might benefit from having more time to parse and analyse linguistic information (Biersack et al., 2005). Although vowel lengthening is associated with speech rate reduction, these two acoustic features may be dissociable; therefore, we decided to measure both speech rate reduction and vowel lengthening.

Vowel intensity increase is an acoustic cue of emphasis that has been reported in FDS, although evidence is limited as only a few studies have investigated this feature (Kendi & Khattab, 2019; Rodriguez-Cuadrado et al., 2018). It has also been found to vary in Lombard speech, which is the speech register used to contrast background noise (Garnier et al., 2016, 2018). Increased intensity could be interpreted as a generic resource for enhancing speech clarity (without providing specific phonetic cues), as indicated by higher clarity ratings of loud speech (Cooke & Lecumberri, 2012; Tjaden et al., 2014).

Pitch is the perceptual result of the vibration rate of the vocal cords (Ladefoged, 2006), and its exaggeration (pitch raising and expansion of pitch range) is often interpreted as a strategy to attract attention, signal affection, or enhance salience, without directly serving a linguistic function (Kalashnikova et al., 2017; Ohala, 1984; Trainor & Desjardins, 2002). Exaggerated pitch height has been found in Arabic FDS (Kendi & Khattab, 2019) but not in English FDS (Hazan et al., 2015; Knoll et al., 2009, 2011; Lorge & Katsos, 2019; Uther et al., 2007; see Piazza et al., 2022 for a discussion). Despite the lack of consensus on this feature modulation in FDS, pitch height is modulated by the presence of other listeners (infants, Uther et al., 2007; pets, Xu et al., 2013).

Importantly, we assume that vowel hyperarticulation provides phonetic cues to enhance phonemic specificity and provide linguistic information (in line with Uther et al., 2007). In contrast, we assume that acoustic features such as speech rate reduction, vowel lengthening, intensity increase, and pitch height raising have salience enhancement functions, by providing acoustic cues only (in line with Kalashnikova & Burnham, 2018; Trainor & Desjardins, 2002). This suggests these features (and their functions) may be employed in different ways depending on the listener's profile and the speaker's communicative goal.

2.1.3. Communicative goals

In addition to the effects of speakers' accommodations to their listeners' linguistic profile on acoustic features, there is some evidence that such features may also vary according to other communicative factors both in speech to language learners (e.g., FDS, infant-directed speech) and native adult listeners. The present study considers three communicative goals, which are relevant when teaching novel words irrespective of the learner's language profile (i.e., in NDS as well as FDS). The three goals consist of speaking for enhancing the listener's (1) comprehension, (2) pronunciation, and (3) writing (spelling). As reviewed below, research shows that these goals place increasing processing demands on the listener; thus, we predicted the three goals may elicit different levels of speech adaptation by speakers (Gershkoff-Stowe & Hahn, 2013; Hendriks & Koster, 2010; Klimova, 2014). Previous research indicates that comprehension precedes production of linguistic knowledge: most of the time, learners understand linguistic elements before they can produce them (Childers & Tomasello, 2002; Flege, 1995; Gershkoff-Stowe & Hahn, 2013; Hendriks & Koster, 2010). For instance, Gershkoff-Stowe & Hahn (2013) investigated comprehension and production of novel words by comparing participants' performance in a recognition task (comprehension) and an oral production task. They observed a comprehension–production asymmetry during novel word learning with an (initial) advantage for comprehension. Moreover, within the two modalities for production, oral production (pronunciation) is easier than written production (spelling). Writing in an L2 is a skill that requires deeper linguistic competence and involves higher cognitive load compared with comprehension or speaking (Bourdin & Fayol, 1994, 2002; Chan et al., 2015; Grabowski, 2005; Klimova, 2014). For example, Bourdin & Fayol (1994; 2002), asked children (~7 years old) to recall series of words, either orally or in writing, and found an advantage for the oral modality. Similarly, in a longitudinal case study, Chan et al. (2015) compared the changes in complexity of oral and written sentence production by two non-native speakers during L2 acquisition. Sentence complexity was higher for the oral modality than the written modality in initial learning stages. A possible explanation of these findings regarding production is that both the correct pronunciation and spelling of recently learnt words require handling a high level of phonemic specificity.

2.2. The present study

The present study aims to identify the acoustic adjustments that speakers make when they interact with listeners who have different linguistic profiles, and to investigate whether and how these adjustments vary as a function of the communicative goal of the interaction. To achieve this, speakers (Spanish L1) addressed a native listener (Spanish L1) and a non-native listener (American English L1) in Spanish. The experiment was conducted online, and both native and non-native listeners who interacted with the speakers were simulated. This was achieved using a series of audio recordings to simulate the listener's presence throughout the task. This resulted in an experimental task with a realistic human interaction where speakers were also provided with feedback, which made them perceive their speech adaptation to be relevant for successful communication. Speakers were informed that their goal was to use novel Spanish words in three types of interactions with the listeners. Each interaction type aimed at achieving a different communicative goal: 1) helping the listener recognise the recently learnt novel words (Comprehension condition), 2) helping the listener correctly pronounce those novel words (Pronunciation condition), 3) helping the listener correctly write those novel words (Writing condition). The five acoustic measures described above (speech rate, vowel hyperarticulation, vowel intensity, length, and pitch height) were assessed for each register and communicative goal.

The present study aimed to answer the following questions:

1) Are communicative goals reflected in the acoustic features of speech? If so, what are the acoustic adjustments observed in Spanish word production?

2) Do modulations of acoustic features for communicative goals depend on the listener's linguistic profile?

We developed specific hypotheses regarding acoustic feature adaptations and their modulation in NDS and FDS, and according to the communicative goals. First, we hypothesise that speakers will produce phonetic cues through greater vowel hyperarticulation in the Writing condition, followed by Pronunciation, followed by the Comprehension condition (Writing > Pronunciation > Comprehension). This is because when speaking to enhance comprehension only, the listener can access the lexical form based on partial phonetic cues. In contrast, teaching speech production requires a higher level of phonemic specificity. In the case of oral production, access to the full phonemic structure of the word and articulatory information is required (Bock & Griffin, 2000). Also, Pronunciation articulation should not be exaggerated to avoid inaccurate pronunciation (non-native like). Articulatory information is also required for written production, but it involves the additional load for mapping individual phonemes to their corresponding graphemes (Bourdin & Fayol, 1994, 2002).

Second, we hypothesise that the remaining acoustic features (vowel lengthening, intensity increase, pitch height raising, and speech rate reduction) have salience and clarity enhancement functions rather than providing phonetic cues for production (or for phoneme mapping). Hence, speakers are expected to modulate acoustic cues to enhance clarity and salience in both the Pronunciation and Writing conditions as compared to the Comprehension condition due to the greater demands of Pronunciation and Writing compared to Comprehension (Writing = Pronunciation > Comprehension; Childers & Tomasello, 2002; Gershkoff-Stowe & Hahn, 2013).

Regarding the *Speech register* factor, we hypothesise that we will observe more hyperarticulated vowels, greater intensity increase, pitch raising, and speech rate reduction in FDS compared to NDS. Predictions FOR vowel hyperarticulation are in line with evidence from other languages, such as English, French and Arabic (Hazan et al., 2015; Kendi & Khattab, 2019; Kühnert & Kocjančič Antolík, 2017; Uther et al., 2007), which suggests that phonetic cues are provided via such features to non-native listeners for supporting L2 learning. In addition, non-native listeners would benefit from increased clarity, and for this reason, we also expect greater speech rate reduction and vowel lengthening, higher intensity, and pitch in FDS compared to NDS.

We expect a *Communicative Goal* effect to be present in both speech registers. Our experimental paradigm employs novel words, which were 'recently learned' by both native and non-native listeners. Thus, speech accommodations would benefit both the non-native and native listeners, supporting their recognition, pronunciation and/or writing of novel words. However, we also expect a Communicative Goal by Speech Register interaction: speakers may put more effort into enhancing phonetic cues when addressing a non-native listener compared to a native listener. Greater differences between communicative conditions are expected in FDS than NDS as a reflection of speakers' sensitivity to the non-native listeners' lower linguistic competence compared to native listeners. We predict that this sensitivity may manifest as greater vowel hyperarticulation adjustments in FDS than NDS in the Pronunciation and Writing conditions but not the Comprehension condition. In Comprehension, the speaker´s focus is directed to enhance the overall perception of novel words, implying that FDS and NDS may be similar in this condition.

2.3. Method

2.3.1 Speakers

The final sample size for analysis included 24 participants (Female = 21, M_{age} = 25.33 years, SD = 4.33). Additional 10 speakers were excluded for: technical reasons (n = 1), excessive noise in their recordings (n = 4), failure to complete both sessions (n = 2) and declaring that they did not believe they were addressing real listeners or for reporting that the interaction with the listeners was highly unnatural (n = 3). All speakers reported knowledge of at least one L2, but with low level of proficiency (12 English, 9 Basque, 1 German, 1 Portuguese, 1 Catalan; see Supplementary Material for detailed demographic information). All participants signed an informed consent form before starting the experimental procedure, and the study was approved by the Basque Center on Cognition, Brain and Language (BCBL) Ethics Committee. Speakers were paid 40 euros for taking part in the study.

2.3.2 Materials

Novel words and objects. Target words were 36 disyllabic and trisyllabic novel Spanish words, which were designed to contain one of the 3 target sounds each: a stressed vowel /a/, /i/, or /u/ in the initial syllable (e.g., *anfes,* [‘anfes]). Those vowels were chosen in line with previous research on FDS (e.g., Uther et al., 2007). The target vowels were included to extract intensity, pitch height, formant values (F1 and F2), and duration, and each vowel appeared in 12 novel words. To create some variability in the stimuli, different spelling patterns were used, but always ensuring that the target vowel would appear in the stressed position in each word. As a result, the vowels could carry a graphic or a non-graphic accent (i.e., *áfaro,* [‘afaro] and and *isbos,* [‘isbos]), or be preceded by an H (i.e., *hasme,* [‘asme]), which is not articulated in Spanish (see Appendix 2.2 for the complete list of the target words). Eighty-four additional novel words with a different structure were added to the 36 items of interest to be used as fillers in the task. Twenty-four of these were trisyllabic novel words beginning with a non-stressed vowel (e.g., *imoste,* [i’mɔste]), 48 were disyllabic or trisyllabic novel words beginning with a consonant (e.g., *jeñofo,* [xe’ɲɔfo), and 12 were disyllabic words ending on -d (e.g., *rosed*, [‘roseθ]). In total, a set of 120 novel words were created for use in the interactions and were presented superimposed on the object images (Figure 4). Two Spanish native speakers, naïve to the present study, checked and confirmed that all novel words were not real Spanish words and followed Spanish phonotactics, making them sound like real words.

A set of 120 novel objects was selected to match the 36 target novel words and 84 filler novel words. The images were taken from five novel object databases (Anderson et al., 2018; Antón et al., 2015, 2020; Barry et al., 2014; Horst & Hout, 2016) and from a pool of pictures freely available online (see Supplementary material for a list of experimental stimuli). The images represented unknown objects and unfamiliar tools to simulate listeners’ word learning.

To create the object-word pairings while avoiding any effects derived from specific relations between words and objects in our stimuli, we created 3 lists of pseudo-random associations, and the presentation of these word-object lists was counterbalanced across speakers. Pseudo-randomisation was used to prevent repetition of the same word-object pair across lists, and to prevent the same target vowel appearing more than twice in a row.

Each word–object pairing was presented twice per condition (36 x 2 repetitions x 4 conditions); thus, the total number of target vowel-initial words was 576 (288 x 2 sessions) per speaker. The same novel objects and words were used for all 4 conditions (Alone, Comprehension, Pronunciation, Writing). All words were displayed in black font size 1.3vw of the screen, default CSS font style, and presented on a white background. The material was strictly identical across conditions and experimental sessions.

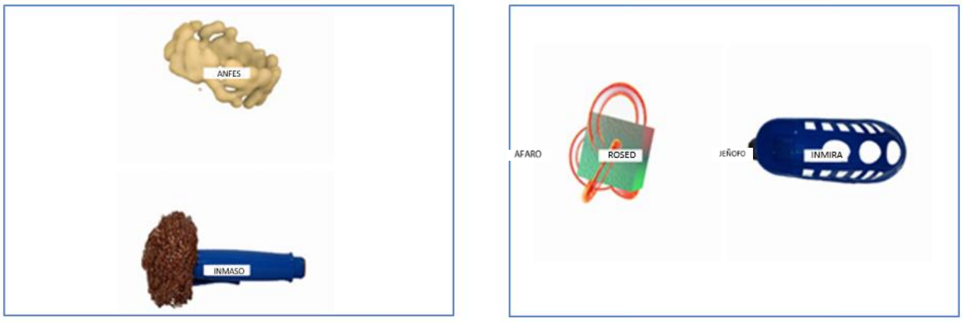


Figure 4. Trial examples, with vertical (left) and horizontal (right) alignment. The target word is "ANFES" in the left image and "ÁFARO" in the right image.

Listener simulation. Throughout the experiment, speakers were led to believe that they were engaged in a live interaction with an interlocutor who could hear them and who was relying on their speech to successfully complete various tasks involving the novel objects and novel words. We decided to simulate the presence of a listener instead of involving live confederates (Uther et al., 2007) to standardise the fictitious listener's behaviour and feedback to ensure this was not a confounding factor between speakers (G. Piazza, Martin, et al., 2022). Although previous studies have asked speakers to imagine a listener (Knoll et al., 2009; Scarborough et al., 2007), more recent studies indicate the importance of 'live' listeners to trigger speech adaptations to their linguistic needs (Buz et al., 2016; Kuhlen &

Brennan, 2013; Piazza et al., 2022). The online experimental design facilitated the simulation of a realistic interaction via computers (as in Buz et al., 2016), which would be difficult in a traditional lab environment.

To simulate the presence of live listeners and to elicit both target speech registers (FDS and NDS), we audio-recorded a non-native and a native speaker of Spanish and included the audios in the experimental task. The non-native speaker was a native male speaker of American English who introduced himself in Spanish for about 1 minute. He spoke naturally, so his speech was accented, and there were some grammatical mistakes, which corresponded to this person’s level of Spanish proficiency. When comparing speech directed to native and non-native listeners, native speakers are usually referred to as highly proficient speakers, monolingual or dominant in their preferred language. The definition of non-native listeners is not as rigorous or well reported: this group typically includes low to mid proficient L2 learners (Cheng et al., 2021; G. Piazza, Martin, et al., 2022; Uther et al., 2007). For this reason, five native Spanish speakers rated accentedness and proficiency of the American interlocutor on a 5-point scale (1 = low proficient/strong foreign accent, 5 = almost native-like proficiency/almost native accent). All raters considered the non-native interlocutor to have intermediate proficiency (mean = 3) and a moderate foreign accent (mean = 2.4). The native interlocutor was a native male speaker of Spanish, and he introduced himself in Spanish for about 1 minute. Both speakers covered similar topics in their introductions (speaking about their age, profession, place of living, etc.). To further create an illusion of a live interaction, each simulated interlocutor was recorded answering potential questions from the experimenter, which could arise during the task, such as: “Are you ready?” or “Can you hear me?”. They also produced different utterances stating that they were ready to begin, they were tired and asked for extra time to rest, and they could hear well. These utterances were used during the experiment breaks, before resuming each experimental condition.

2.3.3 Procedure

The experiment was administered online via PennController for Ibex (Zehr & Schwarz, 2018), which is a JavaScript-based platform. During the session, speakers remained connected with the experimenter via Zoom™, but video streaming was disabled throughout the experiment. This allowed us to verify that speakers' microphone worked properly and that they stayed focused on the task. We asked participants to wear a head-mounted microphone if available, but any type of microphone with acceptable quality was allowed. Before the start of the experiment, speakers recorded and played back their own voice to self-check audio quality.

Speakers completed two sessions (conducted 7-12 days apart), each consisting of a short practice, and four experimental conditions (the Alone condition, and the Comprehension, Pronunciation, and Writing critical conditions) (see Figure 5 for a representation of the experimental procedure).

General task procedure. In each experimental session, speakers were instructed to first name novel objects without a listener (Alone condition), and then to help their listener achieve a specific goal involving these words and objects (comprehension, pronunciation, or spelling) in the following parts. On each trial of all 4 conditions, participants saw two novel objects on the screen and were asked to name each object within sentence frames and then press Enter to send their response. Speakers were asked to embed novel word into sentences instead of reading the individual words superimposed on the object images. This was done to a) measure sentence speech rate and b) reduce the effect of clear speech derived from reading (Blaauw, 1992; Hazan & Baker, 2011; Laan, 1992). To control for the sentential and lexical contexts in which the target labels appeared, speakers were asked to always name the object on the top or left side of the screen first, and to only use masculine grammatical gender for all labels (the determinate article 'el' or the contracted form of 'de el', 'del'). For example, if presented with the trial in Figure 4, the speaker was instructed to say "El ANFES está encima del INMASO" (*the ANFES is above the INMASO*) or "El ÁFARO está a la izquierda del JEÑOFO” (*the ÁFARO is on the left of the JEÑOFO*). To simulate the cooperative task between interlocutors, on half of the trials, the objects were aligned vertically and on the other half, horizontally (counterbalanced across trials). The target novel words always appeared on the top or left positions, but the speakers were not aware of this assignment, so the target label was named first on each trial.

Order	Condition	Instruction	On participants' screen	Participants' task	(Simulated) Listener's task
1	**Alone**	Describe the position of the novel objects			No listener
2	Interaction with a listener (Native or Non-native)		In turns the participant and simulated listener introduced themselves		
Counterbalanced	**Comprehension**	Help recognition of the novel words			To order objects on the screen
	Pronunciation	Help correct pronunciation of the novel words			To repeat the novel words
	Writing	Help correct spelling of the novel words			To type the novel words

Figure 5. Graphical representation of the experimental procedure. Last column “(Simulated) Listener’s task” represents what the speakers were told the listener’s task was. Speakers could not see the listener’s screen.

Practice. At the start of each experimental session, speakers were first familiarised with the visual arrays of the trials and their task of naming the novel objects. After reading the instructions, speakers completed 3 practice trials that were not included in the final analyses.

Alone condition. After practice trials, speakers were instructed to name the objects within the fixed sentence structures, without addressing a listener (and no communicative goals). Participants were informed that these were practice trials for the teaching task they would perform next.

Introduction of the interlocutor. Next, speakers were informed that their interlocutor would join the session. They were told that this person had previously learned the words seen in the Alone condition, and that their task was to further assist the listener to learn the novel words of Spanish throughout three tasks explained next. To simulate a conference call, speakers clicked a button to begin the call and waited for the connection to be established. Then, the experimenter asked the listener to shortly introduce himself. After that, the speaker was asked to introduce him or herself as well. Thus, speakers had the opportunity to detect the listener's linguistic profile (native vs. non-native speaker), and it was established that the two interlocutors could hear each other.

Comprehension, Writing, and Pronunciation conditions. Next, the three experimental conditions began following almost identical procedures. Speakers were asked to name the objects to (1) help the listener correctly match the words to the objects' images (Comprehension condition, 72 critical trials), (2) help the listener pronounce the target words correctly (Pronunciation condition, 72 critical trials), and (3) help the listener type the novel words correctly (Writing condition, 72 critical trials). Instruction of each condition was provided separately, after each break, and the order of each condition was counterbalanced across participants. Speakers believed that (a) the listener saw the two objects in the correct positions on the screen, but could not see their labels in the Writing and Pronunciation conditions, and (b) the listener saw the two objects disarranged on the screen and could not see their labels in the Comprehension condition (and thus their task was to arrange them into the correct position).

Simulated Feedback. In all conditions, after the speaker pressed Enter to send their response, the microphone continued recording for 500ms to avoid early button presses to trim off responses. In order to make the interactions with the listeners as realistic as possible, on every trial of all conditions but Alone, speakers saw a green checkmark or a red cross for

1500ms, indicating that the listener’s response was correct or incorrect. To simulate individual variability in the listeners’ response time (as in Buz et al., 2016), the time between pressing Enter and the feedback was adjusted to each condition: in Writing this interval time was a random value between 6500ms-10000ms; in Pronunciation between 4500ms-8000ms; in Comprehension between 3000ms-5000ms. These response intervals were created due to Kuhlen and Brennan’s review (2013), which showed that realistic interlocutor’s response timing is fundamental for eliciting speaker speech adaptation and maintaining successful communication. To keep the speakers engaged throughout the entire experiment, on 70% of trials the response was rated as correct and 30% as incorrect. We kept this rate constant across conditions to maintain comparability. Moreover, negative feedback (“X”) was provided if the participant pressed Enter within 2500ms from the beginning of the trial, as no correct response could be elicited by an utterance that was not correctly or entirely pronounced (~1% of the trials).

The two sessions (addressing a native or non-native listener) followed the exact same procedure and the order of NDS and FDS (session 1 or 2) was counterbalanced across participants. At the end of the second session, speakers were asked to leave a comment about the experiment (if any) and to indicate on a Likert scale (1-5) how natural they found the interaction with each of the two listeners. We did not directly ask questions about the interaction with the listeners to avoid getting spurious answers (Buz et al., 2016). As noted above, speakers were discarded from the analyses (n = 3) if they spontaneously reported that they did not believe that the interaction involved real listeners or provided a rating of 1 to the question on how natural the interaction was. These criteria were adopted to ensure that the acoustic features were naturally elicited and to reduce the risk of over/underarticulated or otherwise ungenuine speech (Scarborough et al., 2007).

2.3.4 Statistical analyses

Offline spectral analyses were conducted. All target sounds were marked in TextGrid files and extracted by custom scripts with PRAAT software (Boersma & Weenink, 2001). We measured intensity, pitch height and formant values at the mid-point of each /a/, /i/, /u/ vowel. Duration was calculated for target vowels. Formant values were used to calculate vowel triangle area, as a measure of vowel hyperarticulation, with the package *PhonR* in R (R Core Team, 2012). Moreover, another script in PRAAT calculated speech rate as the number of syllables in the trial sentences divided by sentence duration (de Jong & Wempe, 2009).

All missed, mispronounced or noisy responses were discarded from the analysis. In addition, we removed all responses exceeding 2.5 standard deviations above or below the speaker´s individual mean for all variables. Before analysing vowel hyperarticulation, we also removed formant values that did not fall within the acceptable values for each vowel category for Spanish (Bradlow, 1995). Across the dependent variables, for Alone, Comprehension, Pronunciation, and Writing conditions, 8.82%, 7.19%, 5.79%, and 5.86% of trials were discarded, respectively, whereas based on the FDS and NDS condition distinction, 5.97% and 6.58% of the trials were discarded, respectively. The number of discarded trials did not differ significantly across conditions (all p values > .05). Values for each dependent variable in each experimental condition (i.e., Comprehension, Pronunciation, Writing) were divided by values of the Alone (control) condition, used as baseline to obtain scores normalised to each session and for each speaker. The resulting normalised scores were used as dependent variables in the statistical tests. The baseline was set to 1 and the scores for all variables refer to the normalised increase or decrease from the baseline for all analyses.

Each dependent variable was separately added to a linear mixed-effects model in R using the *lme4* package (Bates et al., 2015). Both *Speech register* (NDS, FDS) and *Condition* (Comprehension, Writing, Pronunciation) were within subject factors. We set a priori contrasts so that within *Speech register,* NDS was the referent level for FDS, whereas in the *Condition* factor, Comprehension was the referent level for Pronunciation and Writing. Regarding

Session (first or second session), we used sum contrasts, so that the first session was set to -1 and the second session to 1. Starting with the maximal structure, various models were created and their fit to the data tested with the package *Performance* (Lüdecke et al., 2021) and its function *model_performance* (model fit with Maximum Likelihood method). The models that converged and showed the best fit were chosen to test our hypotheses. All the models included (at least) *Speech register* and *Condition* as fixed effects and by-Speaker as a random effect. Where possible, by-Word and by-Condition random slopes were added. *Session* was introduced as a random effect to account for any differences between the first and the second session, unless differently specified. All final formulas employed for these analyses can be found in the Appendix 2.1. To determine significance of the models we used the type II Wald chi-square tests included in the CAR package (see Fox, 2015; Fox & Weisberg, 2019). For exploring interactions between main effects, we ran post-hoc analyses using the *emmeans* package with Tukey HSD correction for multiple comparisons. We used the Kenward-Roger degrees-of-freedom calculation method for the vowel hyperarticulation variable. For all the other variables the asymptotic method was employed instead, because the number of observations exceeded 3000.

2.4. Results

First, we explored whether the presence of a listener induced speech adaptation (changes in acoustic features) compared to speech aimed at no addressee. We used one-sample t-tests to compare production of each of the five target acoustic features in the condition expected to show the least speech adaptation, NDS-Comprehension, with the baseline Alone condition (i.e., we assessed whether the values for these five features were significantly different to 1). This analysis showed that all NDS-Comprehension scores were significantly > 1, indicating that addressing a listener induced acoustic changes compared to the Alone condition. Table 1 provides a summary of the t-test results.

Variable	t	df	p
Vowel hyperarticulation	2.15	23	0.044
Speech rate reduction	6.95	2598	< 0.001
Vowel intensity increase	11.30	3110	< 0.001
Vowel lengthening	9.50	3122	< 0.001
Pitch height raising	24.63	3078	< 0.001

Table 1. Results from one-sample t-test analyses comparing vowel intensity increase, pitch height raising, vowel hyperarticulation, vowel lengthening, and speech rate reduction in the NDS-Comprehension condition to the baseline value of 1.

Vowel hyperarticulation

The final model included the main effect of *Session* as a fixed effect and by-subject random effect. By-word random effect was not included in the model because the vowel space measure was collapsed across words and calculated by speakers and conditions only. The main effect of *Condition* was also significant ($\chi^2 = 18.94$, $p < .001$). The effect of *Session* was significant ($\chi^2 = 4.38$, $p = .036$), revealing that vowels were more hyperarticulated in Session 1 (+17.04%) than in Session 2 (+7.33%). Post-hoc analyses revealed that the difference between the normalised scores of Comprehension and Writing was significant, with vowels in Writing being pronounced with a wider acoustic space ($t = -4.26$, $p = .001$), +17.87% over the Alone condition compared to +6.63% in Comprehension (see Figure 6). The normalised scores of Pronunciation (+12.07%) did not differ from those of Comprehension ($t = -2.06$, $p = .10$), but there was a trend for Pronunciation to be less hyperarticulated than Writing ($t = -2.20$, $p = .076$). In addition, the main model detected an effect of *Speech register* ($\chi^2 = 5.96$, $p = .015$), reflecting that, relative to the Alone condition, vowel hyperarticulation was present to a greater extent in FDS (+17.03%) than in NDS (+7.36%). No significant interaction was identified.

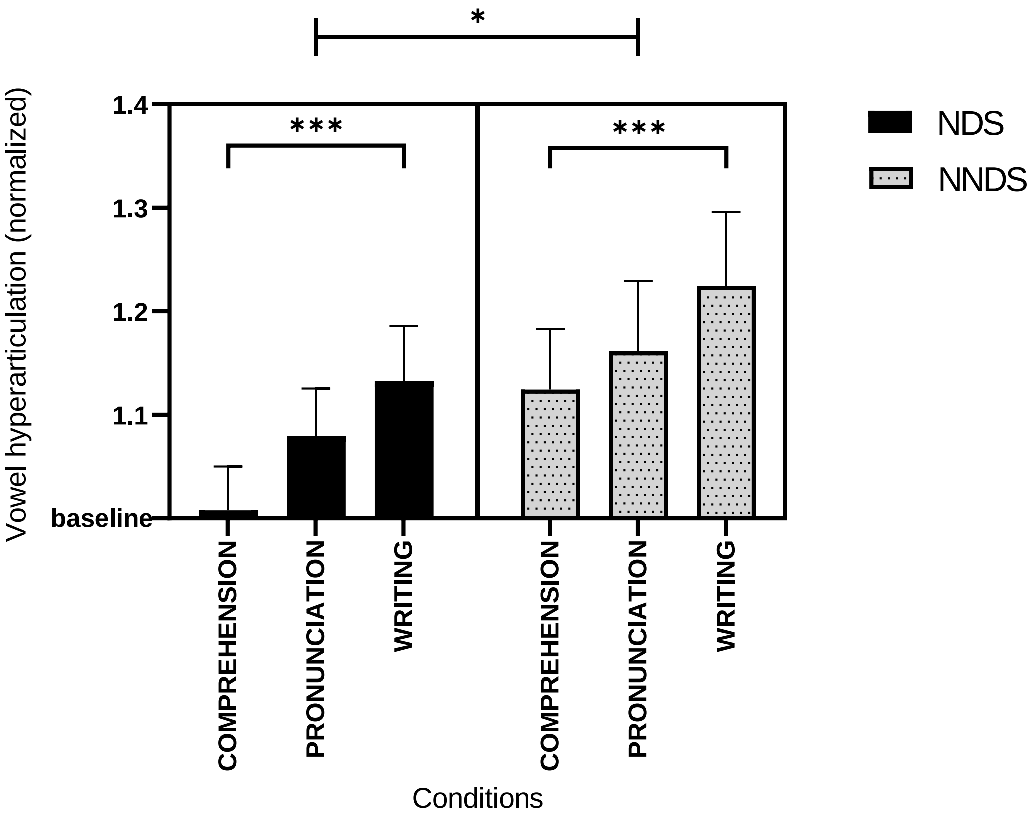


Figure 6. Vowel hyperarticulation by Speech register (NDS = Native Directed Speech; FDS = Foreigner Directed Speech) and Condition. Y axis reports normalised intensity values relative to the baseline value of 1 (alone). Bars represent SEM. Asterisks indicate significant differences (*p <.05, **p < .01, ***p < .001).

Speech rate reduction

The final model for Speech rate included random variation in intercept among speakers within communicative goals and the fixed effect of *Trial number*. Relative to the alone condition, FDS had slower speech rate than NDS ($\chi^2 = 195.11$, $p < .001$); Speech rate was reduced by 6.34% compared to the Alone condition in FDS, but only by 3.27% in NDS. The effect of *Trial number* was significant ($\chi^2 = 95.56$, $p < .001$), indicating that speech rate was slower in the first trials and increased throughout the session. No effect of *Condition* ($\chi^2 = 0.82$, $p = .66$) was observed but there was a *Condition* by *Speech register* interaction ($\chi^2 = 23.15$, $p < .001$). Post-hoc comparisons confirmed that FDS had slower normalised speech rate than NDS (NDS-FDS Comprehension: $z = 11.32$, $p < .001$; NDS-FDS Pronunciation: $z = 8.16$, $p < .001$; NDS-NND Writing: $z = 4.64$, $p < .001$), but there were no significant differences between conditions within FDS and NDS. In both registers, there were no significant differences as a function of communicative goal: Comprehension - Writing (FDS: $z = -0.08$, $p = 1$; NDS: $z = 0.96$, $p = .93$), Comprehension - Pronunciation (FDS: $z = 0.63$, $p = .99$; NDS: $z = 1.13$, $p = .87$), Pronunciation - Writing (FDS: $z = -0.72$, $p = .98$; NDS: $z = -0.17$, $p = 1$; see Figure 7).

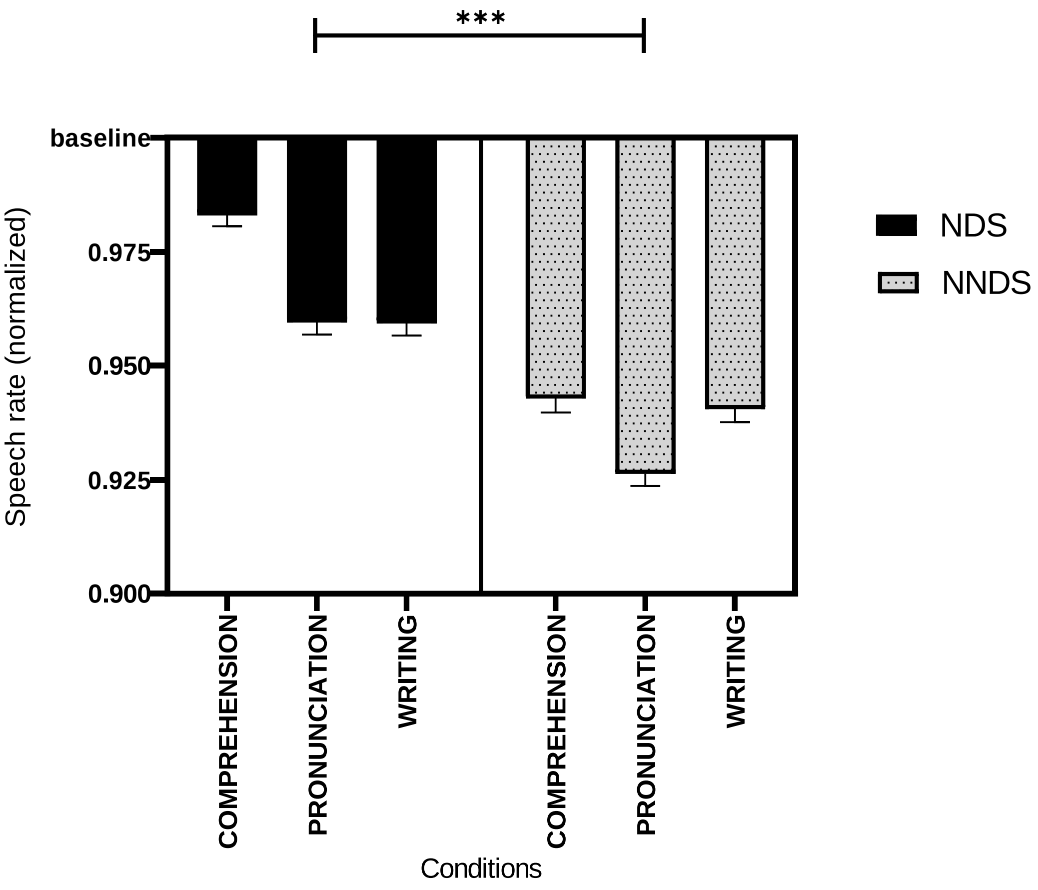


Figure 7. Speech rate reduction by Speech register (NDS = Native Directed Speech; FDS = Foreigner Directed Speech) and Condition. Y axis reports normalised intensity values relative to the baseline value of 1. Bars represent SEM. The bar and asterisks above the graph indicate significant main effect of Speech register (*p <.05, **p < .01, ***p < .001).

Vowel intensity increase

In the final model, the *Condition* factor yielded a significant effect for vowel intensity ($\chi^2 = 10.07$, $p = .007$), but *Speech register* did not ($\chi^2 = 0.26$, $p = .613$). There was a significant *Condition* by *Speech register* interaction ($\chi^2 = 17.17$, $p < .001$). Relative to the Alone condition, vowels in FDS were louder in Comprehension (+1.67%) than Writing (1.39%, *z* = 3.03, *p* = .030) but not than Pronunciation (+1.64%, *z* =0.58, *p* = .992), which in turn did not significantly differ from Writing (*z* = 2.46, *p* = .135). Conversely, in NDS vowels for Pronunciation (+1.67%) were louder than for Comprehension (+1.24%, *z* = -4.08, *p* < .001; see Figure 8), whereas Writing (+1.15%) did not differ significantly from Comprehension (*z* = -2.41, *p* = .154) and Pronunciation (*z* = 1.68, *p* = .543).

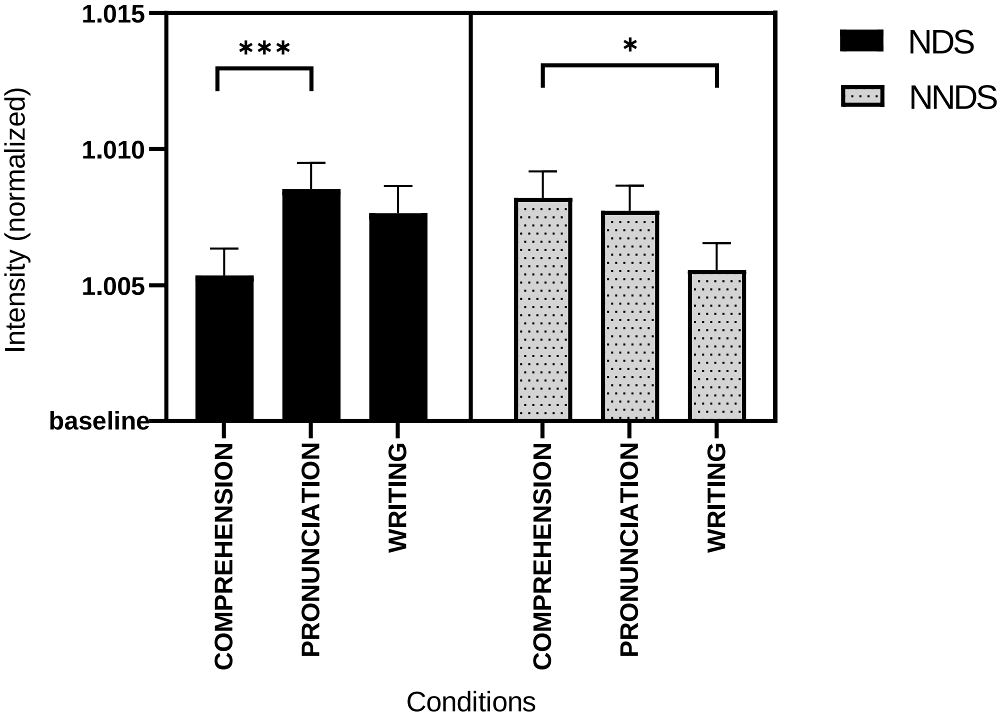


Figure 8. Vowel intensity increases by Speech register (NDS = Native Directed Speech; FDS = Foreigner Directed Speech) and Condition. Y axis reports normalised intensity values relative to the baseline value of 1 (Alone). Bars represent SEM. Asterisks indicate significant differences (*p <.05, **p < .01, ***p < .001).

Vowel lengthening

The final model for vowel duration included *Sex* as a fixed effect as its main effect was significant (χ^2 = 4.59, *p* = .030): female speakers exaggerated vowel length to a greater extent (+5.19%) than male speakers (+0.93%). The main effect of *Speech register* was not significant (χ^2 = 2.71, *p* = .10), yet the main effect of *Condition* was significant (χ^2 = 41.18, *p* < 0.001), and so was the interaction (χ^2 = 12.54, *p* = .002). Post-hoc analyses showed that for FDS, in

Pronunciation (+12.14%) normalised scores were higher than in Comprehension (+10.06%, *z* = -3.44, *p* = .008) and in Writing (+9.00%, *z* = 4.41, *p* < .001), while normalised scores of Comprehension and Writing did not differ (*z* = 1.20, *p* = .84; see Figure 9). In NDS, both Writing (+7.08%, *z* = -4.05, *p* < .001) and Pronunciation (+7.98%, *z* = 5.47, *p* < .001) elicited longer vowels than Comprehension (+4.21% relative to the Alone condition), but the former two did not differ (*z* = 1.43, *p* = .710). The interaction was thus driven by distinct patterns of duration adjustments across communicative goals between the two registers (FDS: Pronunciation > Comprehension = Writing; NDS: Pronunciation = Writing > Comprehension).

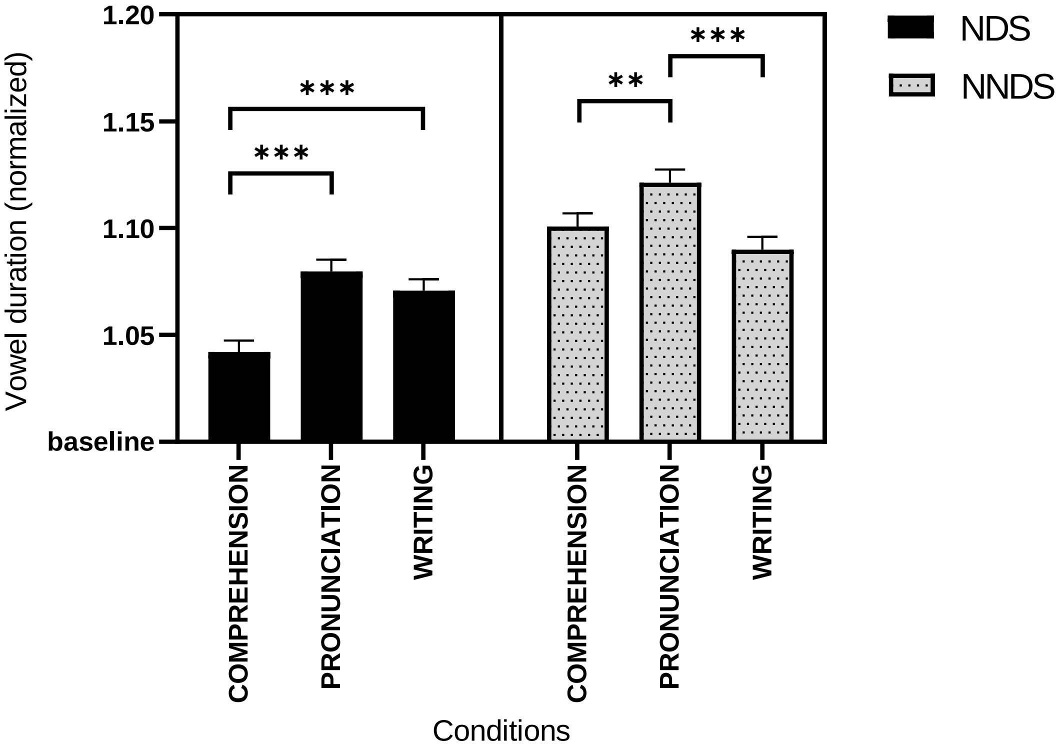


Figure 9. Vowel lengthening by Speech register (NDS = Native Directed Speech; FDS = Foreigner Directed Speech) and Condition. Y axis reports normalised lengthening values relative to the baseline value of 1 (Alone). Bars represent SEM. Asterisks indicate significant differences (*p <.05, **p < .01, ***p < .001).

Pitch height raising

The model revealed only a significant effect of *Condition* (χ^2 = 32.26, *p* < .001), whereas there was no significant effect of *Speech register* (χ^2 = 1.32, *p* = .252) or interaction

(χ^2 = 3.96, p = .138). Post-hoc analyses on *Condition* revealed that, relative to the Alone condition, Comprehension (+6.66% normalised score) had lower pitch than Pronunciation (+7.53%, z = -3.55, p = .001) and Writing (+8.01%, z = -5.62, p < .001), but the normalised scores of the two former conditions did not differ (z = -2.10, p = .089). See Figure 10 for a graphical representation of the effect.

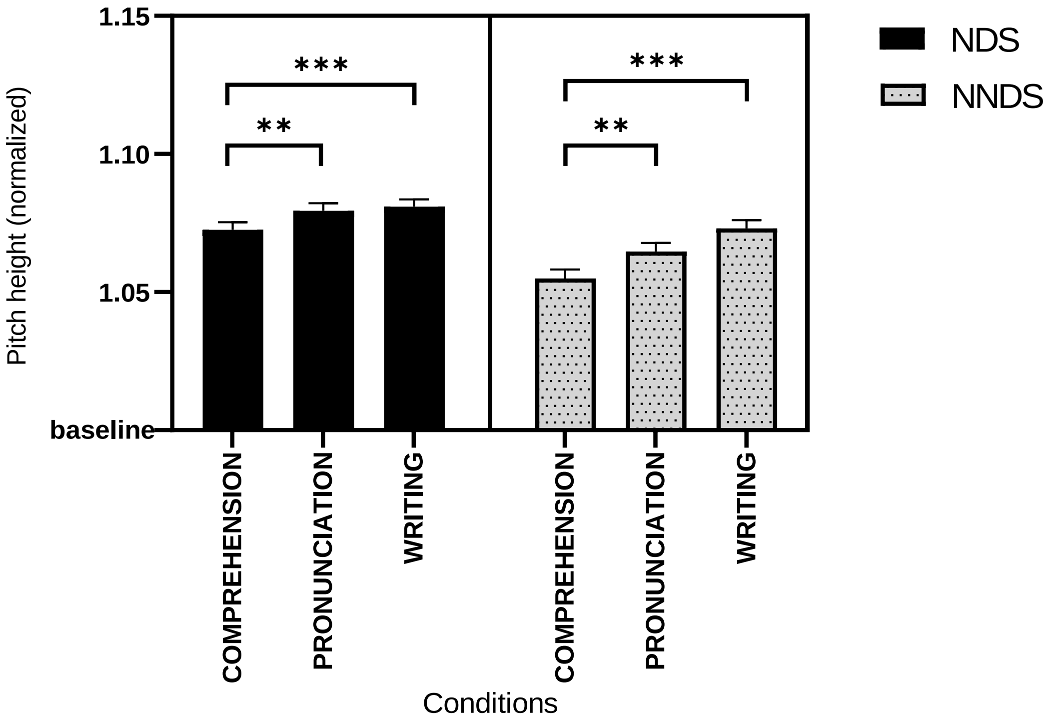


Figure 10. Pitch height raising by Speech register (NDS = Native Directed Speech; FDS = Foreigner Directed Speech) and Condition. Y axis reports normalised pitch height values relative to the baseline value of 1 (Alone). Bars represent SEM. Asterisks indicate significant differences (*p <.05, **p < .01, ***p < .001).

In sum, we found that FDS and NDS differ for vowel hyperarticulation (vowel space) and speech rate reduction. We show that, in both registers, vowels are less hyperarticulated in Comprehension than in Writing and have (normalised) lower pitch in Comprehension than both Pronunciation and Writing. Moreover, we observed a complex set of interactions between the main effects, which are reported in Table 2.

Features	Hypotheses	Dependent variable	Results
Phonetic cues	Writing > Pronunciation > Comprehension FDS > NDS FDS [Writing, Pronunciation] **>** NDS [Writing, Pronunciation] FDS [Comprehension] **≈** NDS [Comprehension]	**Vowel hyperarticulation**	Writing **>** Comprehension Writing **≈** Pronunciation Pronunciation **≈** Comprehensio FDS **>** NDS
Acoustic cues	Writing ≈ Pronunciation > Comprehension FDS > NDS FDS [Writing, Pronunciation] **>** NDS [Writing, Pronunciation] FDS [Comprehension] **≈** NDS [Comprehension]	**Pitch height raising**	Writing **≈** Pronunciation **>** Comprehension
		Intensity increase	FDS [Comprehension **>** Writing NDS [Pronunciation **>** Comprehension]
		Vowel lengthening	FDS [Pronunciation **>** Comprehension **≈** Writing] NDS [Writing **≈** Pronunciation > Comprehension]
		Speech reduction	FDS > NDS

Table 2. Summary of the hypotheses and results by phonetic and acoustic cues.

2.5. Discussion

Previous studies have shown that speakers adapt acoustic features of their speech according to contextual factors, such as the listeners' linguistic profile (Knoll et al., 2011; Uther et al., 2007). In this study, we investigated acoustic features of speech, to assess whether they varied between FDS and NDS and whether any such effects were modulated by the communicative goal (teaching new words for comprehension, pronunciation, or writing). We discuss these findings in more detail below and then consider differences in FDS and NDS.

Vowel hyperarticulation. Speakers were expected to produce greater vowel hyperarticulation (phonetic cues) when their communicative goal was to facilitate the listener's language production (Writing and Pronunciation) as opposed to comprehension. Speakers were also expected to limit vowel hyperarticulation in the Pronunciation compared to the Writing condition. This strategy was hypothesised to be the most suitable to avoid vowel *over-articulation*, which deforms natural formant values and hinders native-like pronunciation. Accordingly, speakers would help the listeners to produce native-like phonemes in the Pronunciation condition and to map phonemes onto graphemes in the Writing condition. This prediction was mostly confirmed since vowels were more hyperarticulated in Writing than in Comprehension, and values for Pronunciation fell in between. This result supports our assumption that speakers provide phonetic cues to the listener depending on their communicative goals. This indicates that: a) speech registers are not fixed, b) speech production is influenced by communicative goals, which suggests that c) speakers adapt to listeners' needs to meet communicative goals, in addition to their linguistic profile (Kalashnikova et al., 2017; Piazza et al., 2022; Uther et al., 2007). Also, in line with our predictions, speakers produced FDS with more hyperarticulated vowels than NDS, as previously reported for English (Hazan et al., 2015; Uther et al., 2007). We did not observe different degrees of hyperarticulation across conditions between the two registers (i.e., no interaction between the two factors). This suggests that speech directed to both native and non-native listeners is adjusted to communicative goals in a similar way (but with greater magnitude for FDS), at least regarding phonetic cues (e.g., vowel hyperarticulation).

Regarding acoustic cues, we predicted greater exaggeration of features in the Writing and Pronunciation conditions compared to the Comprehension condition, and no difference between the former two. This was partly confirmed, at least in NDS. We will discuss these features in the following paragraphs.

Pitch height raising. Pitch height modulations are argued to derive from the intention of raising salience of speech segments relevant to successful communication, without providing phonetic cues. In line with our predictions, in both registers, vowels were produced with greater pitch height in the Writing and Pronunciation conditions compared to the

Comprehension condition. Hence, in our study, pitch height raising is presumed to reflect increasing salience of words that are relevant to specific communicative goals, without conveying linguistic information itself. This result is in line with previous literature showing that a listener's linguistic profile (e.g., infant vs. adult) modulates pitch height with (mainly) attention grabbing functions (Kalashnikova et al., 2017; Ohala, 1984; Trainor & Desjardins, 2002). Here, we report for the first time that speakers adapt pitch height to grab listeners' attention, not only depending on their profile but also depending on communicative goals. These results are in line with the theories of accommodation (Lindblom, 1990; Giles 2016) and suggest that speech and pitch height are indeed adjusted to communicative goals and to maintain communication successful.

In contrast with our expectation, we did not find differences between the two registers or an interaction between communicative goals and register for pitch height raising. This is in line with previous findings (Biersack et al., 2005; Knoll et al., 2011; Uther et al., 2007). Those previous studies claim that high pitch does not serve a teaching function, does not provide acoustic cues useful for L2 learning, and is not a critical feature of FDS (Biersack et al., 2005; Knoll et al., 2006; Uther et al., 2007). Our results support the assumption that pitch height raising is not useful for L2 learning, it is used for salience enhancement, and it is not a critical feature of FDS. We still cannot exclude with certainty that high pitch serves a teaching function, given its modulation in Writing and Pronunciation conditions compared to Comprehension.

Vowel lengthening and *intensity raising.* In NDS, speakers produced longer and louder vowels in Pronunciation than Comprehension. In FDS, vowel length was also increased for Pronunciation compared to Comprehension, but the effect was absent for vowel intensity. These results are consistent with our assumption that modulation of vowel length and intensity occurs to enhance salience in those conditions where this is useful. In the Pronunciation condition, speakers used lengthened vowels (and louder in NDS) to provide the listener with more time to catch the correct vowel uttered (Biersack et al., 2005). Conversely, the

Comprehension condition does not require this enhancement of acoustic cues because word recognition is possible with partial cues and partial phoneme identification.

We also expected vowels in the Writing condition to be longer and louder compared to the Comprehension condition. This was true only in NDS (significant only for vowel lengthening and not for intensity increase). Unexpectedly, in FDS, those effects were absent for vowel length (Comprehension ≈ Writing) and reversed for vowel intensity (Comprehension > Writing). A possible explanation for this finding is that non-native listeners need more teaching support than native listeners, this support being provided via acoustic cues such as vowel hyperarticulation. Thus, when speakers addressed the non-native listener in the Writing condition (the more demanding condition), they resorted mainly to hyperarticulated vowels as a teaching strategy. As an aftereffect of hyperarticulating vowels, speakers reduced their effort to make vowels longer or louder in this same condition. Also, it could be the case that vowel hyperarticulation is never accompanied by length and intensity increase, or it is even accompanied by a reduction of those acoustic cues. For instance, Uther et al., 2007 found that lengthened vowels are not directly associated with vowel hyperarticulation of FDS. In line with this, we tested the correlations between vowel hyperarticulation and vowel length and intensity, and found no significant correlation with either feature (length: Pearson's $r = -0.016$, $t = -0.191$, $p = .849$; intensity: Pearson's $r = 0.094$, $t = 1.125$, $p = .262$). However, further research is needed to shed light on whether speakers independently adapt various speech features to produce the more suitable speech depending on both communicative goals and listener's need. Also, this would suggest that speakers employ the most suitable speech features – intuitively adjusted in the most relevant way – to the listener's needs in relation to their linguistic profile. Whether such feature modulations actually have an impact on the listener's performance in comprehension, pronunciation and writing should be explored in future studies (but see Kangatharan et al., 2023).

In contrast with our prediction, speakers did not produce overall longer and louder vowels in FDS than in NDS. However, the vowel lengthening result is in line with most previous studies, which show no direct link between vowel length and hyperarticulation in FDS (Uther et al., 2007; Knoll & Scharrer, 2007). Conversely, intensity increase results are in contrast with previous findings, which pointed at higher intensity in FDS than NDS (Kendi & Khattab, 2019; Hazan et al., 2015; Rodriguez-Cuadrado et al., 2018; cf. Piazza et al., 2022). Such a lack of effect could be due to the fact that, contrarily to previous studies, our experiment employed a task where both native and non-native listeners had to learn novel words. Likely, intensity increase is a strategy to which speakers turn to in situations with great need of clarity (e.g., difficult tasks or adverse listening conditions), but that does not provide phonetic cues to meet any specific linguistic need (i.e., learning to write/pronounce a novel word).

Speech rate reduction. Contrarily to our expectation, speakers did not vary their speech rate according to the three conditions. Instead, FDS was produced with lower speech rate than NDS, which is in line with previous literature and with our hypothesis (Biersack et al., 2005; C. Smith, 2007). Reduced speech rate in FDS is not fine-tuned to communicative goals, but it is assumed to give listeners more time to parse speech (Biersack et al., 2005), pointing at enhancing non-native listeners' overall perception.

The role of phonetic and acoustic cues. Across the various features considered here, the results overall confirmed our hypothesis that speakers dynamically adjust speech and its relevant features to communicative goals and listeners' profile. These findings suggest that speech accommodation is a complex phenomenon that is influenced by a variety of factors, including communicative goals. Our study contributes to a deeper understanding of the cognitive processes involved in speech accommodation and sheds some light on the role of phonetic and acoustic cues. That is, it appears that through phonetic cues, speakers provided linguistic information, whereas through the acoustic cues, speakers enhanced salience of relevant parts of speech, without providing linguistic information. Speakers intuitively use those cues selectively to guarantee successful communication.

Non-native and Native Directed Speech. To our knowledge, the current study is the first systematic investigation of the acoustic features of Spanish language FDS. Interestingly, most feature modulations in Spanish FDS were comparable with those previously observed in English and other languages. Vowels were hyperarticulated despite the Spanish small vowel inventory (c.f., Al-Tamami & Ferragne, 2005; Andruski et al., 1999; in line with Piazza et al., 2022). This finding contradicts previous claims that only speakers of languages with dense vowel inventory produce hyperarticulated vowels in clear speech (Al-Tamami & Ferragne, 2005; Andruski et al., 1999). Moreover, it is of note that we observed differences in FDS and NDS, despite both profiles of listeners (native and non-native) being addressed in a didactic task, which may have encouraged similar adaptations in both groups. Speakers were informed that both the native and non-native listeners were learning the novel words presented in the task. Interestingly, to our knowledge, this is the first time that FDS hyperarticulation is observed in comparison to NDS in a novel word learning task similar for both native and non-native listeners. Despite the didactic purpose in both groups, FDS was produced with greater hyperarticulation and speech rate reduction than NDS. This is in line with speech accommodation theories (Giles, 2016; Lindblom, 1990) and the didactic purpose hypothesis of FDS (Piazza et al., 2022; Uther et al., 2007). According to these frameworks, speech is adjusted to the listeners' linguistic profile and to maintain successful communication. Thus, our study suggests that such accommodations in FDS compared with NDS are robust even when the communicative aim is to teach new vocabulary.

2.6. Limitations and future directions.

The present chapter was conducted online and used a simulated listener based on voice recordings, rather than a real listener. This could have impacted speech accommodations, meaning we failed to detect differences which might arise in more naturalistic interactions. However, our design was completely within subjects and, if the online environment affected acoustic features, this effect should be equally distributed across all conditions and sessions. Thus, we are confident in sustaining that the differences we found across speech registers and communicative goal conditions were genuine and not driven by the online setting. In addition, previous research suggests that speech production can be assessed online (Fairs & Strijkers, 2021; Piazza, Kartushina, et al., 2022; Vogt et al., 2021). Yet, we cannot exclude with certainty that speech accommodation to speakers and communicative goals differs significantly between live and simulated (online) interactions. Future research should replicate these results with real listeners to dispel any doubts that speech accommodations to communicative goals differs between online-simulated and onsite-natural interactions.

In addition to the main analyses, we explored whether speaking to an addressee entails more exaggerated acoustic and phonetic features compared to the Alone condition, where speakers did not speak to anyone. The results showed that addressing an interlocutor, even a native listener, causes speech changes. In fact, we found out that speaking to an addressee induces acoustic adaptations of all dependent variables, even in the least adjusted condition (Comprehension of NDS). This result urges future studies to carefully consider the use of tasks with live or simulated listeners as this choice may have consequences on experiment outcomes. However, we hesitate to generalise this claim to all experimental manipulations because 1) the Alone condition was at the beginning of each session and did not appear in a counterbalanced order, 2) we employed novel words in order to equate didactic needs for native and non-native listeners, but this manipulation does not represent a typical native-native interaction. Thus, there is need for future research that compares the listener/no-listener (Alone) conditions with real words at play, and ascertains whether the mere presence of a listener induces acoustic changes of speech.

2.7. Conclusion

Across the various features considered here, we found that speakers dynamically adjust their speech to adapt to the communicative goal and their listeners' linguistic profile. These findings suggest that speech accommodation is a complex and dynamic phenomenon that is influenced by a variety of factors. Our study contributes to a deeper understanding of the cognitive processes involved in speech accommodation, and sheds some light on the role of phonetic and acoustic cues. That is, it appears that through phonetic cues, speakers provided linguistic information, whereas through the acoustic cues, speakers enhanced salience of relevant parts of speech, without providing linguistic information. Speakers seem to intuitively use those cues to guarantee successful communication. These results support the didactic function assumption of FDS and extend its use to Spanish speakers.

Discovering that FDS adapts to factors such as communicative goals shows that FDS is endowed with a didactic purpose. Moreover, this finding allows us to predict that FDS has an important impact on language acquisition, thus having a didactic impact on L2 learning. This prediction will be tested in Chapter 3.

Supplementary Material

Material, data, experiment script and analysis code can be found at https://osf.io/9gmp3/?view_only=3d0f81510cc24836a52db13749d5b9aa. Instead, a list of the target novel words and statistical formula can be found in Appendix 2.1 and 2.2

Chapter 3 - L2 Speech Accommodation of Foreigner Directed Speech Supports L2 Word Learning and Pronunciation.

3.1. Introduction

In the previous chapters, we showed the features of FDS and that it is adapted to speakers' communicative goals. For this reason, FDS has also been referred to as "L2 speech accommodation" because it is assumed to be the result of the speaker's accommodation to the listener's low L2 proficiency and learning needs (see also Giles, 2016; Lindblom, 1990 for theoretical frameworks). Such theories of accommodation entail that FDS is a register derived from the intention of meeting the listener's linguistic need – in this case – L2 learning. In the first chapter, we proposed that the didactic function of FDS comprises two related aspects: a *didactic purpose* and a *didactic impact*. The former is the function of producing clear speech to support L2 teaching, reflected in the acoustic features of FDS, whereas the latter is the actual effect on L2 learning, perception, and production. While we found evidence for the didactic purpose indicating that speakers systematically adjust their FDS (Chapter 2), resulting in clearer speech (Piazza et al., under review; Uther et al., 2007), so far, the didactic impact of FDS has never been directly explored. In the present chapter we investigated whether L2 learners benefit from being exposed to FDS to test its didactic impact on perceiving, learning, and pronouncing L2 words.

3.1.1. From high clarity to the didactic impact of FDS

FDS is characterised by speech adaptation to the non-native listener. Compared with NDS, such an adaptation leads to production of several acoustic features that enhance FDS clarity and potentially supports L2 learning. The most typical features are speech rate reduction and acoustic exaggeration of vowels, i.e., vowel hyperarticulation (Piazza, Martin, et al., 2022). Vowel hyperarticulation is assumed to be the key acoustic feature that serves a didactic function because it results in a clearer and more distinctive representation of vowel categories (Kuhl et al., 1997). These features together are proposed to support speech perception, comprehension and even production (Bradlow & Bent, 2002; S. H. Ferguson & Kewley-Port, 2002; Piazza, Martin, et al., 2022; Smiljanić & Bradlow, 2005, 2009). Several studies provide indirect evidence that FDS supports speech comprehension by asking listeners to rate the intelligibility of this register. For instance, Knoll et al., (2009) and Uther et al., (2007) asked naïve native listeners to rate FDS and NDS audio samples. They found that FDS is clearer than NDS but less than other clear speech registers, like Lombard speech, which is a register produced to contrast background noise during native-native interactions (Garnier & Henrich, 2014; Hazan et al., 2015; Lombard, 1911). Conversely, in other studies L2 learners reported to understand FDS better than both NDS and Lombard speech (Bobb et al., 2019; Kangatharan, 2015b). Lombard speech is a register produced by native speakers to address another native listener in adverse listening conditions and not oriented to L2 learners. Thus, contrary to FDS (G. Piazza, Martin, et al., 2022), Lombard speech (as well as NDS) probably lacks a didactic function (both purpose and impact).

These rating findings indicate that the perception of FDS and the enhancement of clarity differ between native and L2 learners, suggesting that there may be differences in how helpful FDS may be for the two populations (Rothermich et al., 2019). However, direct evidence showing that FDS supports spoken word learning, recognition, or pronunciation in L2 learners is still missing. Few experiments tested the efficacy of clear speech registers for word learning in adults. For instance, Golinkoff & Alioto (1995) and Ma et al. (2020) found that Chinese Infant Directed Speech (IDS) helps non-native adult participants to learn words. IDS shares various acoustic features and didactic function with FDS, although these registers have different intended addressees. Thus, one could expect that FDS is particularly suited to support adults' L2 learning. To test this assumption, we need to explore how L2 learners acquire perception and pronunciation of L2 words and phonemes when exposed to FDS. Learning an L2 entails acquiring several skills and managing various elements of that language. In the following section we introduce the most relevant aspects of L2 learning and the difficulties that novice L2 learners can face.

3.1.2. Aspects of L2 learning

Perception and assimilation. L2 learning is primarily mediated by perception of novel phonemes (Escudero, 2005, 2006; Flege, 1995; Melnik-Leroy et al., 2022; Van Leussen & Escudero, 2015). Initially, naïve L2 learners have difficulties in discriminating phonetic contrasts that are not present in their L1 (both vowels and consonants). The relative difficulty to distinguish L2 phonemes depends on the perceptual assimilation to the listener's L1 phonology (Best & Tyler, 2007; Flege, 1995; Mora et al., 2022). According to the Perceptual Assimilation Model for L2 (PAM-L2, Best, 1991; Best & Tyler, 2007), the most difficult situation for L2 perception is when the two L2 phonemes map onto a single native category (see also Flege, 1995; Flege & Bohn, 2021; Kramsch, 2007 for alternative framework). In this case, the two L2 phonemes can either map equally poorly to that category (Single Category), or one phoneme can be a better fit than the other (Category Goodness), making the perception slightly easier than the former category. For Spanish learners of English, an example of Single Category is the vowel contrast /æ - ʌ/ (contained in words like *cap/cup*), comprised by two vowels that are not present in the Spanish phonemic inventory. In this case the pair of L2 vowels fall within the perceptual space of a single L1 vowel category (/a/), which makes it difficult to perceive the phonetic differences between the vowels (Baigorri et al., 2019; Escudero, 2001; Rallo Fabra & Romero, 2012). Conversely, an example of Category Goodness for Spanish listeners is the vowel contrast /i - ɪ/ (contained in words like *ship/sheep*), which is relatively easier than the previous category because the /i/ of *sheep* is a better instance of the Spanish /i/ than /ɪ/ (which is not present in the Spanish phonemic inventory). To test this hypothesis, Baigorri et al. (2019) investigated Spanish-English bilinguals on the categorical perception of English contrasts. The authors found that /æ - ɑ/, /ʌ - ɑ/, and /ʌ - æ/ contrasts are particularly difficult to recognize for late Spanish-English bilinguals, whereas discrimination accuracy was higher for /ɪ - ɛ/ and /i - ɪ/ (see also

Boomershine, 2013 for similar findings on perceptual ratings). Although accuracy was higher for /i - ɪ/, it was also found that late bilinguals tend to perceive /i - ɪ/ vowels in a less categorical way than native listeners (Escudero, 2001; Peng et al., 2010).

There is broad consensus on the assumption that experience (re)shapes L2 learners' phoneme perception (Aoyama & Flege, 2011; Escudero, 2006; Flege et al., 1997; Flege & Liu, 2001). In fact, Flege et al. (1997) tested experienced and inexperienced L2 learners of various languages on synthetic /i - ɪ/ and /æ - ɛ/ continua and reported that the experienced group was more accurate than the inexperienced one at both perceiving and producing the vowel contrasts. This suggests that the perceptual system adapts to learning novel vowel contrasts and the perception of those can be changed with training (Iverson et al., 2008; Logan et al., 1991; Tremblay et al., 1998; Wang, 1999). However, it is not clear how much training is needed to observe such a change in the perception of phonological boundaries in L2 learners. Some studies found that perceptual change happens only in mid-to-high proficient experienced L2 learners (Reinisch et al., 2013), whereas others found changes in low proficient L2 learners within the duration of an experimental session (O. A. Drozdova et al., 2015; P. Drozdova et al., 2016). Nevertheless, it is currently unknown whether such a perceptual change occurs in L2 learners when exposed to FDS. L2 learners' perceptual change of L2 phonemes is likely an important step in the learning process. Testing learning in the context of FDS will shed new light also on the phonological boundaries' adaptation after short training in the L2.

In both the above discussed types of phonetic assimilation, problems with correct mapping of L2 phonemes hinders L2 learners from creating two distinct vowel categories. This determines the difficulty to perceive and produce these vowels in a distinct manner (Flege, 1995; Mora et al., 2022). So far, there is little evidence on the effectiveness of phonetic training for improving such mappings and phonological representations (Lee et al., 2015), and there is no research at all on the effectiveness of FDS on improving L2 perception. For this reason, in this study we focused on the learning process of both perception and production of

L2 vowels and words in FDS. We focused on the FDS didactic impact for learning the two types of assimilation categories, Single Category and Category Goodness respectively, including the /ʌ - æ/ and /i - ɪ/ English vowel contrasts. By doing this we aim to provide a well-rounded research approach for the study of the didactic impact of FDS with a simulation of L2 learning of English.

Production. All the above reported studies focus on L2 phoneme perception. However, this is just one, although fundamental, aspect of learning an L2. The learning process entails acquiring various aspects of the L2, which include production too (Flege, 1995). L2 learners must also deal with the challenge of correctly pronouncing novel words. Most adult L2 learners cannot reach native-like pronunciation so a non-native accent, which depends on their L1, is common (Derwing & Munro, 2009; Flege, 1984). As well as non-native pronunciation is expected, most naïve learners also have issues in distinguishing the pronunciation of L2 vowel contrasts (O'Brien, 2021; e.g., Category Goodness). This makes the two vowels difficult to distinguish, it lowers intelligibility, and possibly leads to miscommunication. Thus, to accurately pronounce L2 vowels, phonetic differences between vowel categories must be learned. For this reason, we are also interested in investigating the advantage of learning to pronounce words and vowels derived from exposure to FDS.

3.2. The present study

L2 learners perceive FDS to be clearer than NDS (Bobb et al., 2019), but to date, research assessing the impact of FDS on L2 learning is not available (Piazza et al., 2022). To disclose the didactic impact of FDS in the second language acquisition process, there is need for research on the effect of exposure to FDS on learning, perceiving, and producing L2 words and vowels. For this purpose, we recruited Spanish native listeners who were novice learners of English to participate in an online experiment. Participants were randomly assigned to a register group (NDS, FDS) and asked to learn a set of 24 English pseudowords. All participants learned three types of novel words: (1) minimal pairs containing the /ʌ - æ/

contrast (like *guck*/*gack*), (2) minimal pairs containing the /i - ɪ/ contrast (like *deest/dist*), (3) non-minimal pair novel words containing the /a/ and /u/ vowels as fillers (like *parg/phoon*), which were included to increase item variability in the experiment. After the learning phase, participants were tested on word learning, word production and vowel perception. We did this in three tasks:

1) *Recognition task*. Participants were auditorily taught the associated label for each object in either FDS or NDS. In the testing phase, participants were asked to recognize the association between the labels and novel objects. Accuracy and response times across blocks (*Block* factor) were compared between the FDS and NDS groups.

2) *Production task*. Participants were presented with the previously learned objects, one by one, and were asked to pronounce their name. Accuracy and response latencies across blocks (*Block* factor) were compared between the FDS and NDS groups. Also, we computed phonetic accuracy as means of a) Euclidean distance between the participants' vowel production and the vowels in the stimuli. In addition, we computed the Euclidean distance b) within each vowel contrast (/i - ɪ/ and /ʌ - æ/). With this latter measure, we investigated phonetic accuracy but from the perspective of vowel distinctiveness.

3) *Continuum discrimination task*. Participants were administered two continuum categorical perception tests of the /i - ɪ/ and the /ʌ - æ/ contrasts (separately). We tested participants before the learning phase (pre-test) and after they completed both the Recognition and the Pronunciation tasks (post-test). That is, we investigated potential changes in participants' ability to discriminate these vowels as a result of exposure to a specific speech register.

Using these tasks, we are interested in answering the following questions:

1) Does FDS enhance word learning as compared to NDS?

2) Does exposure to FDS improve L2 vowel pronunciation accuracy and distinction as compared to NDS?

3) Does exposure to FDS as compared to NDS shape L2 vowel perception?

The Recognition and Production tasks aimed to answer the first question. In line with the assumption that FDS yields a didactic impact on the process of L2 learning, we expected the FDS group to learn better than the NDS group. This would be revealed by a steeper learning curve across blocks and faster responses in the Recognition task. FDS is also assumed to deliver articulatory information by providing L2 learners with exaggerated phonetic contrasts, which is not the case for NDS. Thus, in the Production task, we expected higher accuracy (i.e., production closer to target reference) and faster responses with a steeper learning curve in the FDS group compared to the NDS group. In addition, for Spanish participants, the /ʌ - æ/ contrast (Single Category, henceforth Single) is expected to be more difficult to produce than the /i - ɪ/ contrast (Category Goodness, henceforth Goodness) (Best & Tyler, 2007; Escudero, 2006; Flege et al., 1997). Thus, we expected participants to have lower accuracy and higher response times, in both tasks, for the Single than Goodness contrast. FDS is expected to exaggerate features of all the contrasts, which makes the two contrasts equally difficult in the two registers, but overall easier to perceive and pronounce for the FDS group.

The Production task also aimed to answer the second research question. If FDS provides enhanced articulatory information, the FDS group was predicted to produce vowels that are closer to what they heard as compared to the NDS group. This should be reflected by a reduced Euclidean distance between participants' production and the target reference – and steeper improvement across blocks – in the FDS as compared to the NDS group. We also expected FDS participants to pronounce vowel contrasts (Single and Goodness) in a more distinctive way than the NDS group, reflected by greater Euclidean distance between vowels of both contrasts.

Lastly, the Continuum discrimination task aimed to answer the third research question. If FDS enhances vowel discrimination, this effect may not be limited to the words that participants learned during the task. Native perception of vowels is quasi-categorical (Altmann et al., 2014), but non-native perception is not. Thus, both NDS and FDS participants were not expected to show a clear perceptual boundary between the two target vowels in the pre-test. Instead, in the post-test, we expected only the FDS group to show a more native-like perception of the two contrasts. This would suggest that FDS induces adaptation in the listener’s L2 perceptual system after short training.

3.3. Method

3.3.1. Participants

We recruited 50 native Spanish participants with a mid-low level of English, aged 18-40. The participants were randomly assigned to one of two groups (25 participants each), exposed to either NDS or FDS (NDS group: M_{age} = 26.76 y.o., SD = 6.55; FDS group: M_{age} = 27.36 y.o., SD = 6.48). All participants were tested for their English level in an individual interview with an expert linguist, who assigned marks from 1.0 to 5.0 (1.0 = low; 5.0 = native-like). In the interview, fluency, vocabulary, grammar, and pronunciation are evaluated, and altogether concur in the overall mark. We only recruited participants who obtained an overall mark between 1.0 and 3.0 (NDS group: M_{age} = 1.8, SD = 0.45, FDS group: M_{age} = 1.9, SD = 0.32). In addition, at the end of the experimental session, participants were asked to carry out a Raven matrices test and a pseudoword repetition task in Spanish, used as indices of participants’ non-verbal IQ and phonological memory (Clark et al., 2012; Kaufman & Kaufman, 2014; see Appendix 3.4 for a description of these tasks).

Bayesian analyses were conducted to assess that the two participant groups did not differ in age, English proficiency, non-verbal IQ, and phonological memory. Two-tailed analyses with Cauchy prior distribution (scale of $\gamma = 0.707$) revealed that age, proficiency, IQ, and phonological memory of the two groups were respectively (Bayes factors, BF_{01}) 3.39, 2.25, 3.05, and 3.53 times more likely under the null than the alternative hypothesis. This evidence suggests that the two groups did not significantly differ in age, English proficiency, non-verbal IQ, and phonological memory.

3.3.2. Material

Empirical evidence on other vowels than /a/,/i/,/u/ (e.g., /ɪ/, /ʌ/, /æ/) in FDS literature is limited. For this reason, before the actual experiment, we ran a pilot study to assess matrices of FDS adaptation on /ɪ/, /ʌ/, /æ/ vowels. We report the results and description of this preliminary study in the Appendix 3.1. Below, the material of the three tasks is described.

Recognition task and Production task. For the present study, we created 16 novel words containing the /i - ɪ/ (e.g., [diːst - dɪst]) and /ʌ - æ/ contrasts (e.g., [gʌk – gæk]). The novel words for both vowel contrasts were built as minimal pairs, so that participants had to rely on the target vowels to distinguish the words. To increase item variability, we also created 8 novel words containing the /a/ and /u/ vowels (not forming minimal pairs) that served as fillers (e.g., [p^haːg – fuːn]; see Appendix 3.1 for the full list of experimental stimuli). A set of 24 novel objects was selected to match the 16 target novel words and 8 filler words. The images were taken from the Horst & Hout (2016) novel object database and represented unknown objects and unfamiliar tools. To create the object-word pairings while avoiding any effects derived from specific relations between words and objects in our stimuli, we created 3 lists of pseudo-random associations, and the presentation of these word-object lists was counterbalanced across speakers.

The stimuli were recorded by a female native speaker of British English. This speaker was chosen from the 5 speakers who participated in the pilot study as best representing the observed preliminary results (see Appendix 3.1; wider vocalic area, longer sentence duration, larger /ʌ - æ/ Euclidean distance and /i - ɪ/ duration difference). This speaker produced novel words in FDS with wider vocalic area (+187%), longer sentence duration (M_{FDS}= 3640ms,

M_{NDS}= 3561ms), greater /ʌ - æ/ Euclidean distance (M_{FDS}= 358.10 Hz^2, M_{NDS}= 161.96 Hz^2), and larger /i - ɪ/ duration difference (FDS: 15ms, NDS: 4ms) than in NDS[3]. Conversely, she produced smaller /i - ɪ/ Euclidean distance in FDS than NDS (M_{FDS}= 933.88 Hz^2, M_{NDS}= 1169.25 Hz^2). All stimuli were normalised for intensity and used in both the Recognition and the Production task.

Continuum discrimination task. A female native speaker of British English, who did not record the stimuli for the other tasks, was recorded while producing the words *sheep*, *ship*, *cup*, *cap*. These recordings were used to create two seven-step continua. The sheep-ship continuum was created by gradually changing the formants and the length of the target vowels. The cup-cap continuum was created by solely changing the formants of the target vowels as this contrast is not marked by vowel duration (Escudero, 2001, 2006). Based on the continua, we created 7 isolated instances of words from sheep to ship and from cup to cap that were used in this task.

3.3.3. Procedure

The experiment was administered online via PennController for Ibex (Zehr & Schwarz, 2018), which is a JavaScript-based platform. During the session, participants remained connected with the experimenter via Zoom™, but video streaming was disabled throughout the experiment. This allowed the experimenter to verify that participants' microphone worked properly and that they stayed focused on the task. We asked participants to wear headphones and a head-mounted microphone if available, but any type of microphone with acceptable quality was allowed. Before the start of the experiment, participants recorded and played back their own voice to self-check audio quality. . Participants' compliance was confirmed using a screening test (Woods et al., 2017). After that, the experimental session followed this order:

---

[3] Statistical analysis cannot be run here. Stimuli contain 24 words and each Euclidean distance or duration difference need comparison between 2 words. So, we have 12 observations, 6 for each contrast: very few to test significance.

Continuum discrimination task (pre-test), Recognition task, Production task, Continuum discrimination task (post-test), Raven matrices test, Pseudoword repetition task.

Continuum discrimination. The task began by displaying two images on the screen, one at a time (either a sheep and a ship or a cup and a cap, in counterbalanced order across participants). For each image, participants were presented with an auditory recording of the image's name pronounced in NDS. Then, the task started, and participants used the mouse to click a button on the center of the screen to listen to the stimuli. They were presented with one sound of the continuum at a time (in a random order). The two pictures previously displayed (a sheep and a ship or a cup and a cap) were presented on the screen and participants were asked to click on the picture corresponding to the word they heard. Each endpoint and mid-step word (7 in total) were repeated 6 times (42 trials per contrast). After completing the block corresponding to the first two images (e.g., sheep and ship), the same procedure was followed for the remaining minimal pair (e.g., cup and cap). Both pre-test and post-test followed the exact same procedure.

Recognition task. The object-word pairs were presented once during a familiarisation phase, where participants passively learned the associations. During familiarisation, participants were exposed to novel objects presented together with the auditory version of their name, embedded in a carrier phrase (e.g., “this is a *deest*”). Each sentence was pronounced in either FDS or NDS depending on the participant' group. The recognition task began after each object-word pair was presented once (in the familiarisation phase). In the recognition phase, participants saw images of 4 objects on the screen and heard a sentence identical to the familiarisation phase (e.g., “this is a *deest*”). The 4 objects comprised the target object (e.g., deest), a competitor (e.g., dist) and two distractors (e.g., *gack* and *phoon)*. Participants used the mouse to click a button on the centre of the screen to listen to the cue-sentence. Then, the objects were displayed on the screen until participants provided a response by clicking on one of the 4 objects. As soon as they did so, all the objects disappeared and the correct one was displayed on the centre of the screen for 2500ms. This way, participants were provided with feedback on whether their answer was correct. The stimuli were presented in a randomised order within each block. Each block included 24 trials (16 experimental trials + 8 fillers) and participants were exposed to 6 blocks in a row (total number: 96 experimental trails + 48 fillers). Stimuli were presented in a pseudorandomized fashion to prevent the same target vowel appearing more than twice in a row.

Production task. Participants were presented with the same 24 objects from the recognition task. The objects were displayed one at a time on the screen and the participants were asked to name each of them. As soon as an object was displayed on the screen, the browser started recording participants' oral responses. The object remained on the screen until participants clicked the button 'Send your response'. The microphone continued recording for 500ms after the response was sent to avoid early button press to trim that off. After sending their response, participants heard the novel word in the carrier phrase as in the recognition task, which served as feedback. Then, the next trial began by displaying a new object on the screen. This procedure was repeated until all the object-word pairs were (randomly) presented and repeated in 6 consecutive blocks (96 trials + 48 fillers in total).

3.3.4. Measures and statistical analysis

Recognition task. For this task we extracted 1) response accuracy across the 6 blocks. Offline, 0 and 1 point were assigned respectively to incorrect and correct responses. We also measured 2) response latencies across blocks. Latencies were measured from the moment the cue-sentence finished playing to the moment participants provided an answer. Only correct answers were included in the latency analysis.

Production task. We measured 1) response accuracy across blocks. As a proxy of production accuracy, we computed the Aline distance between each participant's phonetic production and the target pronunciation of the stimuli with the *alineR* package (Downey et al., 2008, 2017) in R (R Core Team, 2021). This measure was the result of feature-weighted linguistic distance calculations. Scores resulted in finite values from 0 to 1, where 0 represented perfect production-target match and 1 represented answers containing sound categories unrelated to the target. We also calculated 2) response latencies across blocks, measured from the object presentation until participants orally responded. Furthermore, based on the values of the first (F1) and second (F2) formants, we computed 3) the Euclidean distance between participants' vowel productions and the stimulus speaker's pronunciation of the vowel stimuli (henceforth just vowel stimuli), and 4) the Euclidean distance within participants' vowel contrast productions (/ʌ - æ/ and /i - ɪ/). All incorrect responses or that – despite some similarity with the target – clearly pointed at a distractor were excluded from the analyses of 2), 3), and 4). For example, if a participant said [piːfəl] for the object associated with the novel word [piːv], they would have obtained an Aline score of 0.31, but their response was considered incorrect because it pointed at the distractor [biːfəl]. The excluded trials represented 39.58% of the total responses. The remaining valid trials were 1559 in FDS and 1341 in NDS (BF_{01} = 1.89, anecdotal evidence for H_0), for a total of 2900. In addition, in order to test the two Euclidean distance variables, formant values of participants' vowel productions (and vowel stimuli) were normalised by employing the Lobanov method (Lobanov, 1971). This method uses a log-mean method to normalise the formant values and computes a single grand mean for all participants. This was done to prevent participants' physiological differences from driving the observed effects.

The dependent variables of the Recognition and Production tasks were independently analysed by using growth curve analysis (GCA) models (Mirman, Dixon, et al., 2008; Mirman, Magnuson, et al., 2008) fitted in R (*lme4* package; Bates et al., 2015; see Appendix 3.3 for a list of the models). This technique is explicitly designed to assess changes over time at group and individual levels. GCA allowed us to add to the models the linear and quadratic polynomial terms to account for the overall slope change and the curvature of the observed effects. The linear term reflects the overall slope, and the quadratic term reflects the curvature (i.e., change in slope across learning blocks). Thus, the 6 blocks were added to the model as *Block* factor, including linear and/or quadratic terms depending on the best model fit. The *Register* (FDS and NDS) and *Contrast* (Single and Goodness contrasts) factors, together with the *Block* factor, were added to the models as fixed effects (unless differently specified). Subject and novel words were included as random effects. Other predictors were considered only if they improved the model fit (see Appendix 3.3 for a list of the final models). Starting with the minimal structure, various models were created; the final models were chosen according to the best fit indicated by the *Performance* package in R (Lüdecke et al., 2021). For all models, we set a priori sum contrasts so that within *Register,* -0.5 was assigned to NDS and +0.5 to FDS, whereas within the *Contrast* factor, -0.5 was assigned to Category Goodness and +0.5 to Single Category (Schad et al., 2020). Response latencies were transformed using the Box-Cox method (Box & Cox, 1964). Conversely, accuracy of the Recognition task was tested by fitting GCA with generalised linear (*glmer*) mixed-effects models and binomial family. Given the distribution of the Aline distance, results were fitted in a *glmer* model of the beta family. Both measures of Euclidean distance were tested in two separate models (one for each contrast: Single and Goodness).

Continuum discrimination. For this task, we used a generalised linear mixed effect model (binomial family) to compare vowel discrimination between the pre-test and the post-test (*Exposure* factor) and between the two speech register groups. We did not include polynomial terms because GCA did not apply for this variable. Ship/sheep and cup/cap continua were tested in separate models.

For all tasks, model significance was tested with the *lmerTest* Package (Kuznetsova et al., 2017) and interactions between main effects were explored by running post-hoc analyses in the *emmeans* package (Lenth et al., 2019) with Tukey HSD correction for multiple comparisons. Given the number of interactions tested in each model, below we report only significant interactions; all results, including non-significant results are reported in the Supplementary material.

3.4. Results

3.4.1. Recognition task

Accuracy. This task aimed to investigate whether FDS promotes L2 novel word learning. *Accuracy.* The final model indicated a significant effect of the *Block* factor's linear term ($\beta = 0.561$, $z = 4.019$, $p < .001$) but not quadratic term ($\beta = -0.0655$, $z = -0.809$, $p = .419$). Participants linearly improved accuracy from 56.75% on average in the 1st block to 71.63% in the 6th block. The main effects of *Register* ($\beta = -0.277$, $z = -0.785$, $p = .432$) and *Contrast* ($\beta = -0.261$, $z = -1.866$, $p = .062$) were not significant but their interaction was ($\beta = 0.533$, $z = 4.007$, $p < .001$). The FDS and NDS groups did not differ for the Single ($z = 0.028$, $p = 1$) and the Goodness accuracy ($z = 0.502$, $p = 0.436$). However, within contrasts, FDS participants were more accurate in recognizing novel words containing the Goodness contrast than the Single contrast ($z = 3.373$, $p = .004$; see Figure 11). Conversely in the NDS group this difference was not significant ($z = -0.034$, $p = 1$). All other interactions were not significant (see Supplementary material).

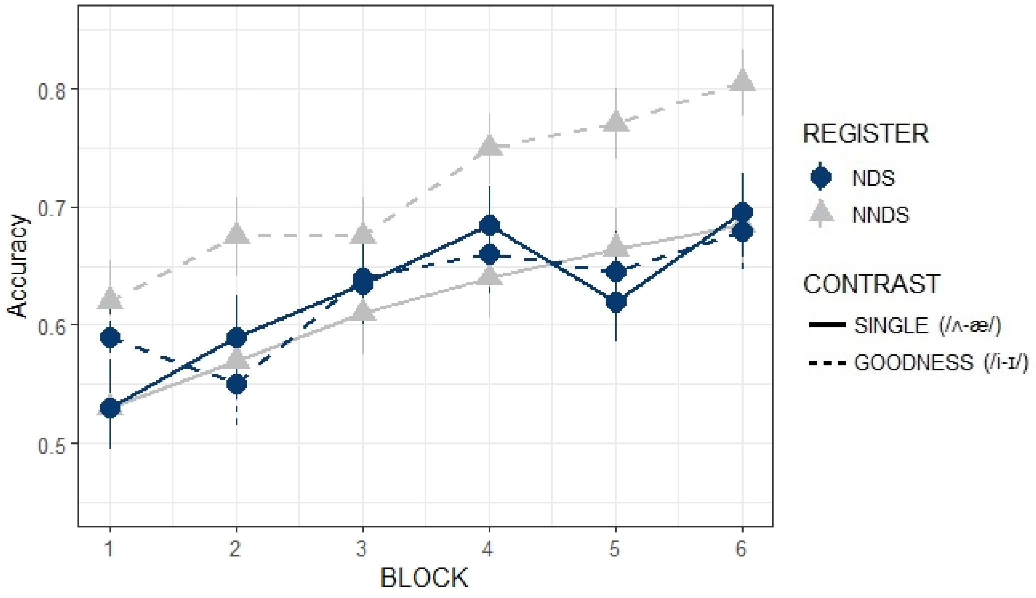


Figure 11. *Recognition task.* Accuracy across blocks, by *Register* (FDS = Foreigner Directed Speech, NDS = Native Directed Speech) and *Contrast* (SINGLE = novel words with the /ʌ - æ/ contrast; GOODNESS = novel words with the /i - ɪ/ contrast). Bars indicate SE.

Response latencies. The final model showed significant effects of linear (β = -0.077, t = -13.754, p < .001) and quadratic terms (β = 0.033, t = 6.854, p < .001). This is because participants got faster, from 6338ms on average in the 1st block to 4834ms in the 5th block, and then reached plateau performance in the 6th block (4892ms on average). Also, the effect of *Register* was significant (β = 0.051, t = 7.562, p < .001) because the FDS group (3792ms) responded overall faster than the NDS group (4551ms; see Figure 12). Conversely the effect of *Contrast* (β = -1.525e-04, t = -0.019, p = .984) and any interaction did not reach significance.

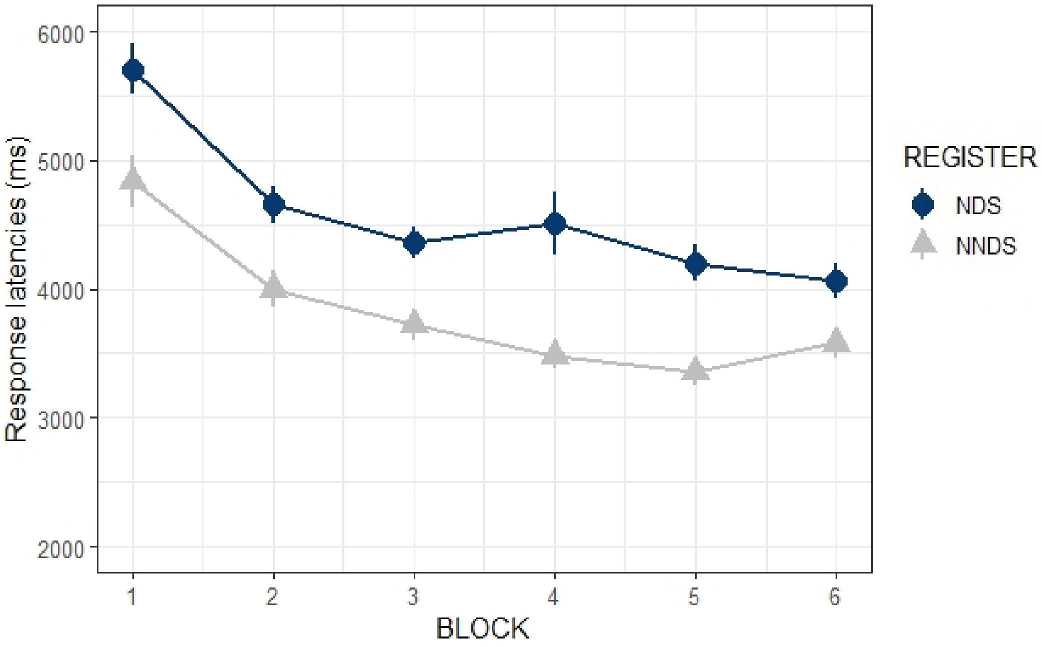


Figure 12. *Recognition task.* Response latencies across blocks by *Register* (FDS = Foreigner Directed Speech, NDS = Native Directed Speech) with responses collapsed across the two contrast types. Bars indicate SE.

3.4.2. Production task

This task investigated whether FDS promotes learning of novel words for production. One participant was excluded from the analyses due to very low production accuracy (~8%).

Accuracy. Results indicated significant improvement in model fit on the linear term (β = -0.469, z = -2.775, p = .005), but not the quadratic term (β = 0.086, z = 1.043, p = .297). This is reflected in the Aline distance scores, which, on average, went from 0.332 in the 1st block to 0.235 in the 6th block. Effects of *Register* (β = 0.231, z = 1.128, p = .259) and *Contrast* (β = 0.074, z = 0.446, p = 0.656) did not reach significance, but their interaction did (β = -0.315, z = -2.298, p = .022). This interaction was likely driven by the NDS group's difference between Single and Goodness contrasts, but post-hoc analyses failed to reveal any differences between Single and Goodness in both the NDS (t = 1.436, p = .477) and FDS groups (t = - 0.446, p = .971; see Figure 13). All other interactions were not significant.

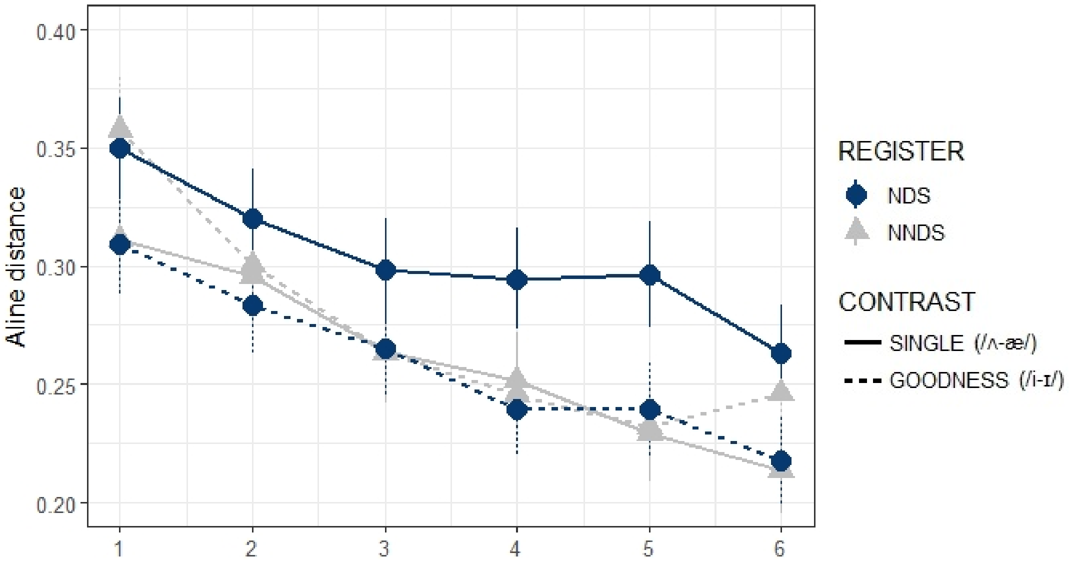


Figure 13. *Production task.* Aline distance between participants' production and the target pronunciation across blocks, by *Register* (FDS = Foreigner Directed Speech, NDS = Native Directed Speech) and *Contrast* (SINGLE = novel words with the /ʌ - æ/ contrast; GOODNESS = novel words with the /i - ɪ/ contrast). Bars indicate SE.

Response latencies. The final model yielded a significant quadratic term (β = 0.048, t = 2.844, p = .004) but not linear term (β = 0.014, z = 0.817, p = .414), indicating that participants' response latencies across blocks best fitted a parabola shape. The *Contrast* factor (β = -0.049, t = -2.093, p = .037) showed a significant effect because the Goodness contrast was produced with shorter latencies (1946ms) than the Single contrast (2408ms; see Figure 14) especially starting from the third repetition. Conversely, the *Register* factor (β = -0.029, t = -0.390, p = .698) and all interactions were not significant.

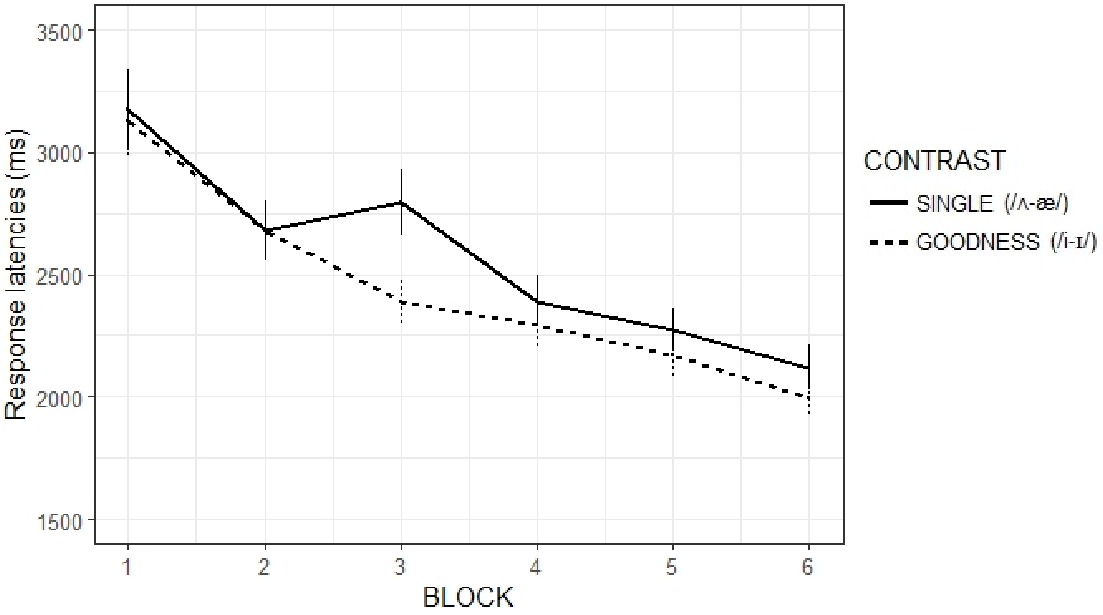


Figure 14. *Production task.* Response latencies across blocks by *Contrast* (SINGLE = novel words with the /ʌ - æ/ contrast; GOODNESS = novel words with the /i - ɪ/ contrast). Bars indicate SE. These data are collapsed across registers.

Euclidean distances. The analysis of Euclidean distances assessed whether the exposure to FDS improves participants' vowel production accuracy and category distinction as compared to NDS. For this purpose, we computed two measures of Euclidean distance:

a) *Euclidean distance between participants' vowel production and vowel.* For this variable, the Goodness (/i - ɪ/) and Single (/ʌ - æ/) contrasts were analysed in two separate models that each included the factors *Vowel* (i and ɪ or ʌ and æ) and *Register*. The results for Goodness model did not reveal significant improvement in model fit on the linear (β = 0.0375, t = 0.481, p = .630) or quadratic terms (β = 0.381, t = 0.504, p = .614). The main effects of *Register* (β = 0.387, t = 5.448, p < .001), *Vowel* (β = 0.417, t = 2.461, p = .038) and their interaction were significant (β = 0.219, t = 3.320, p < .001). Post-hoc analyses indicated that the NDS group produced /ɪ/ further from the target than /i/ (t = 3.269, p = .031), whereas this was not true for the FDS group (z = 2.144, p = .198). Also, NDS participants pronounced /ɪ/ with greater distance from the target vowel than the FDS participants (t = -5.302, p < .001; see Figure 15). This was not true for the /i/ vowel (t = -2.219, p = .127). This result suggests lower phonetic accuracy for the NDS than the FDS group for this vowel. The model for the Single contrast did not reveal any significant effects (see Figure A of the Appendix 3.2) (*Register* β = -0.257, t = -0.455, p = .650, *Vowel* β = -0.012, t = -0.146, p = .887, *Block* linear t.: β = -0.031, t = -0.538, p = .591; quadratic t.: β = -0.057, t = -0.980, p = .327) or interactions.

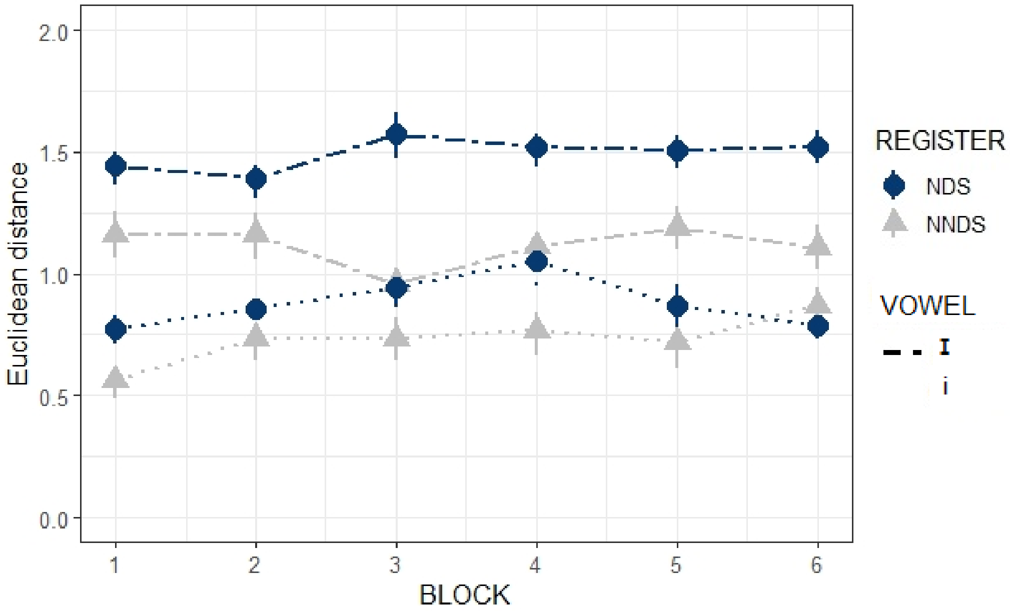


Figure 15. *Production task.* Goodness contrast*:* (normalised) Euclidean distance participants' production – vowel stimuli, across blocks, by *Register* (FDS = Foreigner Directed Speech, NDS = Native Directed Speech) and *Vowel* (i and ɪ). Bars indicate SE.

b) Euclidean distance within vowel contrasts. The two contrasts Goodness (/i - ɪ/) and Single (/ʌ - æ/) were separately investigated in two models, which included the *Block* and *Register* factors. The final model for the Goodness contrast indicated a main effect of *Register* (β = 0.166, t = -2.520, p = .016) but no effect of linear (β = -0.061, t = -1.396, p = .163) or quadratic terms (β = -0.005, t = -0.114, p = .909). The FDS group produced the vowels in this contrast more distinctly than the NDS group (Euclidean distance FDS = 0.732 Hz^2; NDS =

0.557 Hz^2), without substantial changes across the 6 blocks (see Figure 16). Instead, the model for the Single contrast did not show any significant effects or interactions (*Register*: β = 0.021, t = 0.518, p = .607; *linear term:* β = - 0.035, t = -1.178, p = .239*; quadratic term:* β = -0.026, t = -0.853, p = .394).

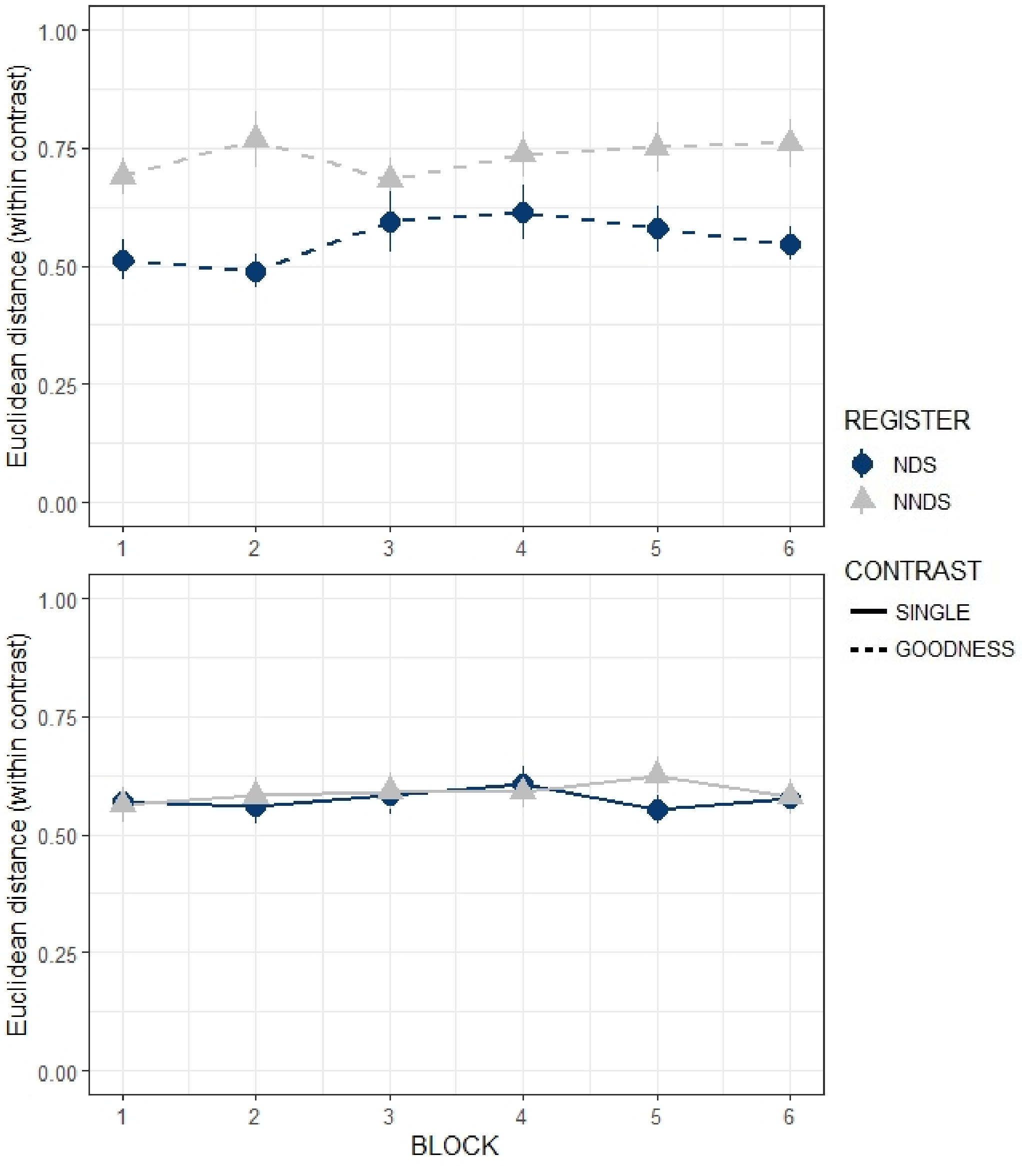


Figure 16. *Production task.* (normalised) Euclidean distance by *Register* (FDS = Foreigner Directed Speech, NDS = Native Directed Speech) and *Contrast* (SINGLE = novel words with the /ʌ - æ/ contrast; GOODNESS = novel words with the /i - ɪ/ contrast). Bars indicate SE.

3.4.3. Continuum discrimination task

This task assessed whether the exposure to FDS or NDS for a short period of time induces changes in the participants' L2 perceptual system, becoming more categorical in the FDS group. In the final model for the sheep-ship continuum, the *Register* (β = 0.419, z = 1.287, p = .198) and *Exposure* factors (β = 0.105, z = 0.800, p = .424) and their interaction were not significant. This was also the case for the cup-cap continuum (Register, β = 0.264, z = 0.942, p = .346; Exposure, β = -0.085, z = -0.924, p = .355, and the interaction was not significant) (see Figure 17).

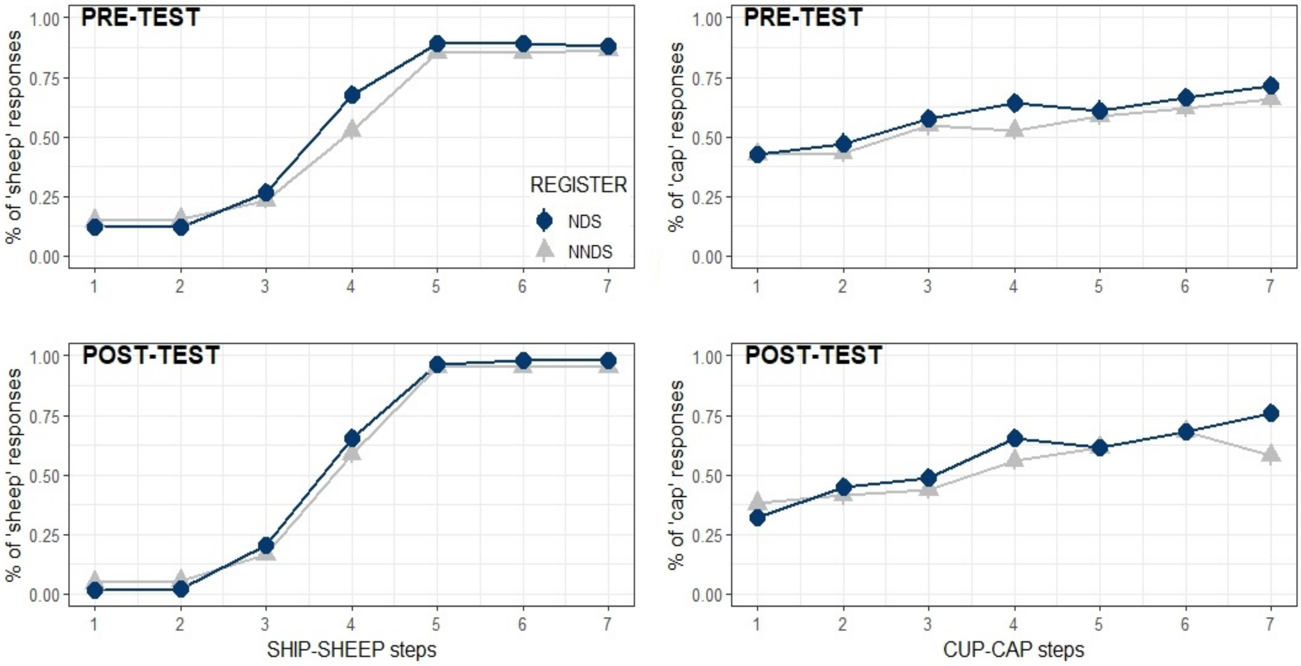


Figure 17. *Continuum discrimination*. Average percentage of 'sheep' choices (on the left) and of 'cap' choices (on the right) across the seven-step continuum by *Register* (FDS = Foreigner Directed Speech, NDS = Native Directed Speech) and by *Exposure* (pre-test and post-test). Bars indicate SE.

3.5. Discussion

Previous literature assumed that FDS is endowed with a didactic purpose – reflected on the acoustic features of FDS – and a didactic impact (Piazza et al., in preparation, 2022;

Scarborough et al., 2007; Uther et al., 2007). Such a didactic impact would support L2 learners both in comprehension and production. However, so far, whether L2 learners' perceptual and production learning is promoted after the exposure to FDS remains unknown. We addressed these questions by conducting an online experiment where two groups of L2 learners of English (Spanish L1) learned the association between novel objects and novel English words pronounced in either FDS or NDS. Perception and learning of English vowel contrasts (/i - ɪ/ = Goodness, /ʌ - æ/ = Single), which are absent in Spanish, was assessed. In order to investigate FDS learning benefits to produce novel words and vowels, participants' production accuracy and latency were measured as well. We expected to find that the group exposed to FDS learned to perceive and pronounce novel words and vowel contrasts more successfully and faster than the NDS group.

FDS Benefits

The present study provides evidence for the FDS benefit on L2 learning for the first time. That is, FDS participants were better at perceiving L2 novel words as compared to NDS participants. Such a benefit was mainly shown in the Recognition task results, which indicated that the FDS group responded faster than the NDS group to recognize novel words (both vowel contrasts). This represents evidence in support of the didactic function hypothesis of FDS and speech accommodation theories (Bobb et al., 2019; Giles, 2016; Lindblom, 1990; Piazza et al., 2022; Uther et al., 2007).

FDS benefit depends on properties of the speech contrasts to be learned

Our results also provide evidence that FDS effects are qualified by the properties of the speech contrasts to be learned, and how they relate to listeners' L1 (Best & Tyler, 2007). That is, FDS benefits were particularly pronounced for the Goodness contrast (/i - ɪ/). For example, Recognition accuracy of the FDS group was higher for the Goodness than the Single contrast, whereas this did not happen in the NDS group. This suggests that even though the FDS group did not show overall better accuracy than the NDS group, their exposure to FDS promoted recognition of words including the Goodness contrast. On the other hand, Production results showed that FDS delivers articulatory information that has an impact on L2 pronunciation learning, but for the Goodness contrast only (Best & Tyler, 2007). The NDS group produced /ɪ/ vowels further away from the target than /i/ vowels, whereas this was not true for the FDS group. More importantly, FDS participants produced /ɪ/ closer to the target than NDS participants. This means that the FDS group was more precise to hit the /ɪ/ target vowel than the NDS group. Moreover, the FDS group was better than the NDS group at separating the vowels of the Goodness contrast (larger /i - ɪ/ distance in their production). These results suggest that FDS provides articulatory information to the listeners, who use such cues to achieve the target pronunciation.

These findings were likely due to the acoustic features of FDS, which enhance the differences between contrast vowels. The FDS novel words containing the Goodness contrast were produced (by our native speaker who recorded the stimuli) with greater /i - ɪ/ duration differences and reduced Euclidean distance than the same novel words pronounced in NDS (see Material and Appendix 3.1). Participants' performance was in line with previous literature that reported Spanish listeners to be particularly sensitive to duration differences between L2 vowels (Escudero, 2001, 2006). This also indicates that FDS duration cues (directed to Spanish listeners) are particularly suited to enhance L2 learners' discrimination of the /i - ɪ/ vowel contrast rather than another contrast signalled by formant value information. Research has hinted that such cues are intuitively produced by native speakers to support communication with L2 learners (Giles, 2016; Piazza et al., 2022, in preparation; Uther et al., 2007; see also Appendix 3.1). Here, we show that these duration cues also support word learning, bearing a didactic impact for L2 learners. We cannot exclude that FDS supports learning of Single contrast vowels as well, but cue adjustments for such a contrast do not seem to be readily effective and more exposure to FDS is probably needed to detect improvement on the Single contrast. Lastly, production latency results indicated participants' faster responses for Goodness than Single contrast words, regardless of the Register group.

This finding does not relate to our focus on differences between FDS and NDS, but it is still interesting because it confirms that the Goodness contrast used here is easier to discriminate than the Single contrast for our participants, as we discuss below.

Theories of Second Language Acquisition that explain the FDS benefit

Such a benefit asymmetry between Goodness and Single contrasts is in line with PAM-L2, which claims that Goodness contrast phonemes are more easily recognized and pronounced than Single contrast phonemes (Best, 1991; Best & Tyler, 2007). Also, there are other second language acquisition theories that account for this learning asymmetry, such as the Speech Learning Model (Flege, 1995; Flege & Bohn, 2021), the Contrastive Analysis Hypothesis (Foley & Flynn, 2013; Kramsch, 2007), and the Input Hypothesis (Krashen, 1989). For example, the Speech Learning Model assumes that the greater the perceived phonetic dissimilarity between an L1 and L2 sounds, the easier it will be to discern them. However, the existing models do not consider the interaction between Register and Contrast that we observed in our results. This interaction refers to the finding that participants showed greater improvement in learning the Goodness contrast than the Single contrast if exposed to FDS rather than NDS. This can be explained by adding a complementary socio-cognitive factor (Atkinson, 2014) to the previous models, which would provide a combined framework that can explain this advantage for the Goodness contrast in FDS (Kuhl, 1997; Piazza et al., 2022; Uther et al., 2007). The socio-cognitive theory of second language acquisition claims that L2 learning is a natural and adaptive process of ecological alignment (Atkinson, 2002, 2014; Atkinson et al., 2007). In fact, our results reveal that learners adapt their perception and production of L2 novel words to the social environment (i.e., learning differs depending on speech adaptation of the speaker/teacher). This hints that FDS is a socially mediated promoter of phoneme category distinction and acquisition.

FDS benefit depends on the modality and task demand

FDS seems to be a suitable tool for teaching a second language, which supports L2 learners' performance, both in recognition and production. However, the present study also revealed that this overall L2 support differs depending on the modality (i.e., word recognition vs. production). We observed better recognition and production performance in the FDS than NDS group (as for the Goodness contrast), but the production benefit was limited to higher accuracy in target vowel pronunciation (Euclidean distance measures). Contrary to our expectation, accuracy (Aline distance) and production latency results did not differ across registers. Even though the model for the Aline distance yielded a significant interaction between Register and Contrast, post-hoc analyses failed to highlight any difference between conditions. We argue that this interaction was driven by the NDS group's higher Aline distance (hence lower phonemic accuracy) in the Single contrast than the Goodness contrast (see Figure 3). However, this remained an anecdotal observation as no significant differences between conditions were observed. In sum, the FDS benefit was visible in faster word recognition but not in faster word production. It could be that FDS is beneficial for word production accuracy and speed as well, but that longer training would be needed to observe those effects on production. This assumption is in line with previous literature reporting that, when learning linguistic elements, comprehension precedes production learning (Childers & Tomasello, 2002; Gershkoff-Stowe & Hahn, 2013; Hendriks & Koster, 2010).

However, it is worth reminding that the Production task was always carried out after the Recognition task. Participants first had to learn to perceive the differences between vowels and novel words, and only after were asked to produce them. We argue that this could be the main cause of the observed disadvantage in the production of /ɪ/ of the NDS participants. During the Recognition task, NDS participants were exposed to novel words containing the /i - ɪ/ contrast in which the duration cue (/ɪ/ shorter than /i/) was reduced as compared to the FDS group. We think this absence of clear duration cues might have impaired accurate perception (and thus learning) of the Goodness contrast. Thus, NDS participants carried over this disadvantage to the Production task, where they could not improve their production learning (Flege, 1995). Nevertheless, FDS participants, who were exposed to reduced /i - ɪ/ Euclidean distance as compared to NDS, produced instead wider /i - ɪ/ Euclidean distance than NDS participants. This leads to two important observations: a) the FDS production benefit does not simply derive from mirroring production of target phonemes; b) for Spanish listeners, duration cues are particularly relevant for learning the /i - ɪ/ contrast, and this affects vowel formant production as well.

An important consideration is that the present study used an online method to collect participants' responses. Several studies have addressed the question of whether online experiments provide reliable results and revealed that chronometric experiments for speech production can be implemented online without information loss (Anwyl-Irvine et al., 2020; Bridges et al., 2020; Fairs & Strijkers, 2021; Piazza, Kartushina et al. 2022; Vogt et al., 2021). Thus, we are confident in sustaining that the differences we found between speech register groups were genuine and not driven by the online setting. However, future research should run similar experiments in a laboratory to dispel any doubts that benefit derived from the exposure to FDS differ online and onsite.

FDS does not induce changes in L2 sound phonetic boundaries (after short training)

Lastly, we found that the effects of FDS exposure do not – at least in this study – change participants' phonetic boundaries of the /i - ɪ/ and /ʌ - æ/ contrasts: phonetic boundaries did not become more native-like despite the improvement in both word recognition and production. In the Continuum discrimination task, we expected to find an adaptation of the phonetic boundaries for both continua (*sheep-ship* and *cup-cap*) in the FDS group's post-test. However, we did not find any difference between the two groups, nor between pre-test and post-test in both vowel continua. This means that the two groups did not significantly differ for initial perception of the two vowel contrasts, and that neither of the two changed their phonetic boundaries in the post-test. Previous research suggests that adaptation of phonetic boundaries can happen within a single experimental session (O. A. Drozdova et al., 2015; P. Drozdova et al., 2016), whereas other research points that longer exposure and experience is needed (Reinisch et al., 2013). Our result aligns with the latter proposal. However, research reported that phonetic adaptation within a single experimental session is visible at the neurophysiological level (Grimaldi et al., 2014). We cannot exclude, therefore, that FDS induces phonetic adaptation after short training, but it is not detectable at the behavioural level, with the particular task and stimuli we used. Thus, further research (both using behavioural and neurophysiological methods) is needed to address this point.

To summarise, this study provides new insights on the process of learning an L2 after exposure to FDS and makes a step forward to understanding the precise mechanisms involved in L2 teaching and learning. We found that FDS has an impact on learning L2 words for recognition and production, but (especially) production improvements depend on the relationships between the phonemes to be learned and learner's L1 phonemic categories. It is important to underline that, in this study, participants were exposed to FDS (or NDS) for a very short period (< 2 hours); hence, it is probable that more benefits would derive from extended exposure to FDS (e.g., classroom teaching). These findings and future research on more prolonged exposure to FDS are fundamental to building models of L2 communication and learning. This research is particularly relevant given that communication between native and non-native speakers is becoming ever more frequent in our increasingly multicultural and multilingual societies.

3.6. Conclusions

The main goal of this chapter was to understand whether the exposure to Foreigner Directed Speech (FDS) promotes L2 learning, both in perception and production, as compared to Native Directed Speech (NDS). This study found that L2 learners (Spanish L1) exposed to English FDS were faster at recognizing novel words than L2 learners exposed to NDS. Also, the FDS group, but not the NDS group, was more accurate at recognizing words containing the /i - ɪ/ contrast than words containing the /ʌ - æ/ contrast. FDS seems to promote L2 listeners' learning and recognition of L2 words. However, a single session was not sufficient to induce phonetic boundary changes in any of the Register groups. In addition, the FDS group was found to be better at pronouncing the /ɪ/ vowel than the /i/ vowel and at producing separate /i - ɪ/ vowels than the NDS group. That suggests that FDS supports L2 listeners' production learning as well, but it depends on the relation between the sounds to be learned and the learner's L1 phonemic inventory.

Some aspects of FDS, which potentially support L2 perception, may be too subtle to be visible in a single experimental session using behavioural technique. One question that needs further investigation after this chapter's experiment is whether L2 listeners' perception and comprehension is supported by exposure to FDS. We will investigate subtle electrophysiological changes (EEG) in non-native listeners' perception in the next chapter.

FDS is used in everyday life interactions between native and L2 learners, and even in L2 classroom environments. Thus, it is fundamental to understand benefits of long-term learning in ecologically valid situations and with long-term exposure (e.g., longitudinal study). For this reason, in the next chapter we aimed to investigate FDS impact on continuous speech perception, with a naturalistic experiment.

Supplementary material

Material, data, experiment script, analysis code, and non-significant results can be found at https://osf.io/xtky5/?view_only=4ec02c26bd084296b088780811ebbb07. Instead, a list of the statistical formula can be found in Appendix 3.3.

Chapter 4 - Does Foreigner Directed Speech support non-native listeners' cortical tracking of speech?

4.1. Introduction

In the previous chapters of this thesis, we showed that FDS is a clear speech register that native speakers use when addressing non-native listeners of their language. We also described the features of this speech register, which are mainly low speech rate and an expanded vocalic space, known as vowel hyperarticulation. Low speech rate and vowel hyperarticulation respectively provide more time to parse speech and enhance detection of vocalic contrasts, which are assumed to promote L2 comprehension and learning (Biersack et al., 2005; Piazza, Martin, et al., n.d.; Uther et al., 2007). This aspect of FDS represents its didactic purpose, which is the speakers' intention to adapt their speech to meet non-native listeners' linguistic needs. On the other hand, previous studies exploring clarity and comprehension ratings have tried to assess FDS didactic impact, which is the actual effect on L2 perception, comprehension, and learning (see Chapter 1 for a definition of didactic impact and didactic purpose). Indeed, those studies suggested that FDS makes perception and comprehension better for non-native listeners, but not for native listeners (Bobb et al., 2019; Cooke & Lecumberri, 2012; Knoll et al., 2009; Uther et al., 2007). These findings supported the didactic impact of FDS for non-native listeners. But the experiment presented in Chapter 3 is – to the best of our knowledge – the first to directly assess FDS impact on non-native listeners' L2 learning.

Such experiments with controlled manipulations do not reflect listeners' naturalistic exposure to L2 speech. In fact, it is unknown so far whether FDS also supports perception and comprehension of continuous speech. No studies have investigated the cognitive processes involved in listening to FDS by employing ecologically valid stimuli. Moreover, most studies that investigated FDS perception (with controlled manipulations) employed behavioural experiments, which can only measure L2 processing when it is already concluded (but see Uther et al., 2012). Instead, using neuroimaging techniques (e.g., EEG), allows the exploration of this process as it unfolds. One way to do this with ecological validity is by using EEG to measure cortical tracking (CT) of continuous speech. In fact, efficient CT is linked to comprehension and clarity of speech (Ahissar et al., 2001; Ding & Simon, 2014; Etard & Reichenbach, 2019; Giraud & Poeppel, 2012). Such study would shed light on whether the phonetic features of FDS improve non-native listeners' CT of speech, and on how non-native listeners process an L2. This research will contribute to the literature on perception of listener-specific speech adaptations and second language acquisition. To elucidate the neural mechanisms that underlie L2 perception, there is a need to assess the effect of speech registers (namely FDS) on neural measures that capture language cortical tracking. In the following section we review the main findings on CT in relation to speech perception and comprehension.

4.1.1 Cortical tracking of speech

When a complex neurocognitive process takes place, millions of cortical and subcortical neurons fire at the same time (Buzsáki & Draguhn, 2004). The results of this activation are what is known as neural oscillations, which are typically studied at various frequency bands (e.g., delta, <4Hz, and theta, 4-8Hz). Neural or cortical tracking happens when neural oscillatory activity synchronises with rhythmic fluctuations of exogenous stimuli (e.g., speech; Giraud & Poeppel, 2012). Speech is a source of acoustic regularities because it contains syllables and word boundaries (which create speech rate patterns). Cortical tracking of speech happens when the oscillations align to the temporal rhythm of, for example, the speech envelope (Obleser & Kayser, 2019). Speech envelope is the low-frequency amplitude modulation of speech that carries acoustic information required for perceptual and linguistic encoding (Attaheri et al., 2022).

CT of speech is considered a marker of the underlying linguistic process during speech perception (Giraud & Poeppel, 2012; Luo & Poeppel, 2007; Meyer, 2018); in fact, some studies found that efficient cortical tracking derives from good comprehension of the linguistic message (Ahissar et al., 2001; Ding & Simon, 2014; Keitel et al., 2018; Peelle et al., 2013; Pérez et al., 2015; Riecke et al., 2018; for opposite finding see also Peña & Melloni, 2012). In addition, some findings suggest that CT of speech at low frequencies bands (delta and theta bands) subserve higher level linguistic information analysis (Meyer, 2018; Molinaro & Lizarazu, 2018), correlates with abstract structure of language processing and its meaning (Ding et al., 2016), reflects word parsing (Kösem et al., 2016) and syllables segmentation (Oever & Sack, 2015). For example, Ding et al. (2016) sustain the existence of hierarchical structure building operations in language comprehension that is reflected on CT. Also, CT of speech has been linked to development of phonological skills by subserving temporal sampling of phrases, words, and syllables (Goswami, 2011; Goswami & Leong, 2013).

Moreover, there are some factors that are known to influence CT of speech. For example, language proficiency, clarity of speech (Etard & Reichenbach, 2019) and speech rate were shown to be reflected on CT in the low frequency bands (Kösem et al., 2018; Verschueren et al., 2022). Etard & Reichenbach (2019) assessed CT of speech clarity as a function of adding/removing background noise to speech. They found that clearer speech induced higher CT as compared to more noisy speech (especially in the theta band). Particularly relevant to our study, Verschueren and colleagues (2022) investigated (native) linguistic speech processing as a function of varying speech rate. Their findings revealed that cortical tracking (CT) decreased as the speech rate increased, indicating a link between the tracking of linguistic representations and the decline in speech comprehension. The above reported results suggest that CT is modulated based on speech processing and changes in quality of speech. So, if this is true, it should be possible to study the impact of FDS on CT. However, CT of FDS has never been investigated and it is unknown whether exposure to FDS is associated with efficient CT of speech. Some analogies with previous studies can be used to further formulate hypotheses for this research. In the next section, we discuss these findings.

4.1.2. Cortical tracking of FDS

While the effects of FDS on CT have not been studied, we do have evidence of efficient CT in response to another speech register, Infant-directed speech (IDS), which is used to address infants and support their language acquisition (Kalashnikova et al., 2017; Kuhl, 1997; Trainor & Desjardins, 2002). IDS has similar acoustic features to FDS, with low speech rate and hyperarticulated vowels, which has been proposed to serve a linguistic function similar to FDS (Burnham et al., 2015; Kalashnikova & Burnham, 2018; see Chapter 1, section 1.6.1 for a comparison between FDS and IDS). Research showed that infants prefer to listen to IDS and show higher CT of IDS as compared to adult directed speech (which is the equivalent of NDS) (Attaheri et al., 2022; Cooper & Aslin, 1990; Fernald, 1985; Kalashnikova et al., 2018; Menn et al., 2022).

While acknowledging the inherent distinctions between adults and infants, the findings from studies exploring CT of IDS seem to indicate that listeners' CT is enhanced when listeners are exposed to speech registers specifically intended for them. These results, along with acoustic features and didactic function analogies drawn between IDS and FDS, suggest that non-native listeners may benefit in perceiving and tracking FDS. In fact, non-native listeners (but not native listeners) reported FDS to be clearer than other clear speech registers (Cooke & Lecumberri, 2012). This allows us to hypothesise more efficient CT of FDS than NDS in non-native listeners, who are the intended addresses of FDS.

4.1.3. Temporal response function

CT of speech can be investigated by employing the (multivariate) temporal response functions (TRFs; (Crosse et al., 2016; Di Liberto et al., 2018, 2021; Di Liberto et al., 2015) of the EEG oscillatory signal in response to continuous stimuli, usually within low-frequency bands (<8Hz; Di Liberto et al., 2015; 2021). By doing so, it is possible to explore the cortical tracking of specific sets of features, such as register-specific acoustic features (Di Liberto et al., 2018). Typically, CT is studied on speech envelope, which reflects speech perception likely with the influence of various factors such as attention, engagement, and comprehension (Ding & Simon, 2014; Keitel et al., 2018). But TRF also allows the investigation of other abstract and non-acoustic speech features, such as semantic information, by building a semantic model of CT (Broderick et al., 2018, 2022; Klimovich-Gray et al., 2023). For instance, Broderick et al. (2018, 2022) modelled TRFs with a computational model of how semantically surprising (and dissimilar) words were to their preceding context. They computed the linear mapping between this regressor and EEG data as the participants listened to narrative speech. The TRF-weighted EEG response showed a prominent centro-parietal negativity around 200-400ms comparable to the classical semantic N400. Instead, they did not find such an N400-TRF response for non-linguistic conditions like reversed speech (see also Broderick et al., 2022 for similar results).

To summarise, it is possible to investigate whether listeners benefit from specific acoustic features for perceiving (envelope tracking) and understanding (semantics tracking) an L2. It is unknown so far whether CT of speech envelope and semantic surprisal in non-native listeners is promoted by the exposure of FDS as compared to other speech registers. For this reason, we decided to run two experiments that are described in the next section.

4.2. The present study

FDS is a clear speech register that likely facilitates non-native listeners’ perception and comprehension as compared to other speech registers, such as NDS. Conversely this is not expected to facilitate speech processing in native listeners, as they have high proficiency, and they are not the intended addresses of FDS. To assess this assumption, in two experiments, we asked participants to listen to continuous speech in the form of storytelling while their EEG activity was recorded. During experiment breaks, participants were asked questions about the content of the stories (comprehension questionnaire) so that we obtained behavioural data on comprehension as well. The stories were presented in three speech registers (FDS, NDS and Slow-NDS). Given that the differences in perception between FDS and NDS might arise from differences in their speech rates (FDS slower than NDS), and that speech rate affects CT, we decided to include the Slow-NDS condition, with the same features of NDS but with speech rate of FDS. This condition allowed us to disentangle the effect of speech rate and other acoustic features of FDS influencing participants' L2 processing (and CT). FDS is assumed to enhance not only speech clarity, but also speech perception and (consequently) comprehension for L2 listeners (Bobb et al., 2019; Piazza et al., 2022), whereas Slow-NDS enhances clarity only because it has slow speech rate, but no other acoustic features directed to support non-native listeners' perception or comprehension.

In experiment 1, the participant group was formed by non-native listeners of English (Spanish listeners, henceforth SL), with mid-low level of proficiency. In experiment 2, participants were native English listeners (henceforth EL). In both experiments, participants listened to three stories in babble noise, which simulated natural listening conditions for both native and non-native listeners, without compromising understanding. In addition to EEG recordings, we collected behavioural data through the comprehension questionnaire to a) assess the comprehension benefit derived from the exposure to each speech registers, b) check that none of the two experiments reached ceiling/floor effects. Hence, with experiment 1 we aimed to answer the following questions: does FDS promote non-native listeners' cortical tracking of L2 speech? Does FDS promote non-native listeners' L2 comprehension? With experiment 2, we aimed to answer this question: Does FDS also promote native listeners' speech perception and comprehension?

Previous literature showed that FDS has specific acoustic features tailored to non-native listeners' linguistic needs. Thus, in experiment 1, we expected the SL group to respond with higher accuracy to questions about FDS stories than NDS. SL were also predicted to show more efficient cortical tracking, of both speech envelope and semantics, when exposed to FDS than NDS. This would be in line with the didactic impact assumption that FDS promotes L2 learning, including perception and comprehension (Piazza et al., 2022). Slow-NDS was expected to show increased accuracy and CT compared to NDS because of its low speech rate that provides participants with more time to process speech and that was found to affect speech envelope CT (Kösem et al., 2018; Verschueren et al., 2022). However, Slow-NDS was expected to show reduced accuracy and CT compared to FDS because it is not a register adapted to non-native listeners and it lacks the acoustic features of FDS (except for low speech rate).

Conversely, in experiment 2, the EL group was expected to show increased envelope CT in both FDS and Slow-NDS as compared to NDS, which would reflect the effect of low speech rate (only) on CT of EL. Participants were not expected to benefit from any speech register in their comprehension, due to their native language proficiency. Thus, we did not expect any difference between speech registers in the questionnaire accuracy. Also, if speech rate does not affect CT of semantics too, semantic TRF for EL should not yield any difference between speech registers.

4.3. Experiment 1

4.3.1. Method

In experiment 1, we tested native Spanish listeners on comprehension accuracy and CT of speech.

4.3.1.2. Participants

A total of 28 participants, aged between 18-35, were recruited to take part in experiment 1. They were non-native listeners of English (native speakers of Spanish), with mid-low proficiency in English (Spanish Listeners, henceforth SL; M_{age} = 22.8 y.o., SD = 3.42, Female = 21). SL participants were tested for their English level in an individual interview with an expert linguist, who assigned marks from 1.0 to 5.0 (1.0 = low; 5.0 = native-like). In the interview, fluency, vocabulary, grammar, and pronunciation were evaluated, and altogether concurred in the overall mark. We only recruited participants who obtained an overall mark between 1.0 and 3.0 (M = 2.96, SD = 0.34). Of the original SL sample, 2 participants were excluded due to technical problems and 1 due to very low comprehension score (4% of correct responses), leaving the final cohort to 25 non-native listeners. The experiment was carried out at the Basque Center on Cognition, Brain and Language (Spain). Before starting the experimental procedure, all participants signed an informed consent form approved by the BCBL Ethics Committee. Participants were paid 20 euros for taking part in the study.

4.3.1.2. Material

Ecologically valid stimuli, represented by continuous speech streams, were employed in this study (see Supplementary material for stimuli and data availability). Speech streams were recorded by a female native speaker of British English in the form of storytelling. Each story was recorded in the two speech registers (NDS and FDS) and then adapted to create the Slow-NDS register. In line with previous literature (Piazza et al., under review, Uther et al., 2007), FDS stories were pronounced with wider vocalic area (-30%) and lower speech rate (-30%) than the NDS stories. The Slow-NDS register was created by using dynamic time warping (Müller, 2007), which kept the acoustic features of the NDS stimuli constant but matched speech rate of the NDS stories to the speech rate of the FDS stories. This technique aims to find the optimal alignment between two time-dependent sequences, which are warped in a nonlinear fashion to match each other. Duration of FDS and Slow-NDS stories was ~15 minutes, conversely duration of NDS stories was ~11 minutes, due to higher speech rate. This option was adopted to maintain the same content for the three stories. English multi-talker babble noise (Krishnamurthy & Hansen, 2009) was added to all the stories (+16 dB SNR) to simulate natural listening conditions. Babble noise level was chosen after an online pilot study that identified this level of noise as adapt to avoid floor effect for SL in experiment 1 (and ceiling effect for EL in experiment 2). Babble noise was created in MATLAB 2014b with a custom script by mixing continuous speech streams of 8 British English speakers (4 females).

Amplitude Modulation Spectrum analysis (Goswami et al., 2002, 2010; Pérez-Navarro et al., 2022) revealed that FDS was characterised by local temporal organisation with amplitude modulation bands aligning mostly along the delta frequency band at ~3 Hz (see Figure 18). Slow-NDS was pronounced with similar amplitude modulation at ~3 Hz, whereas NDS was mostly aligned ~4.2 Hz. All speech registers rapidly decreased their power (dB below 0) at higher frequency-bands than theta.

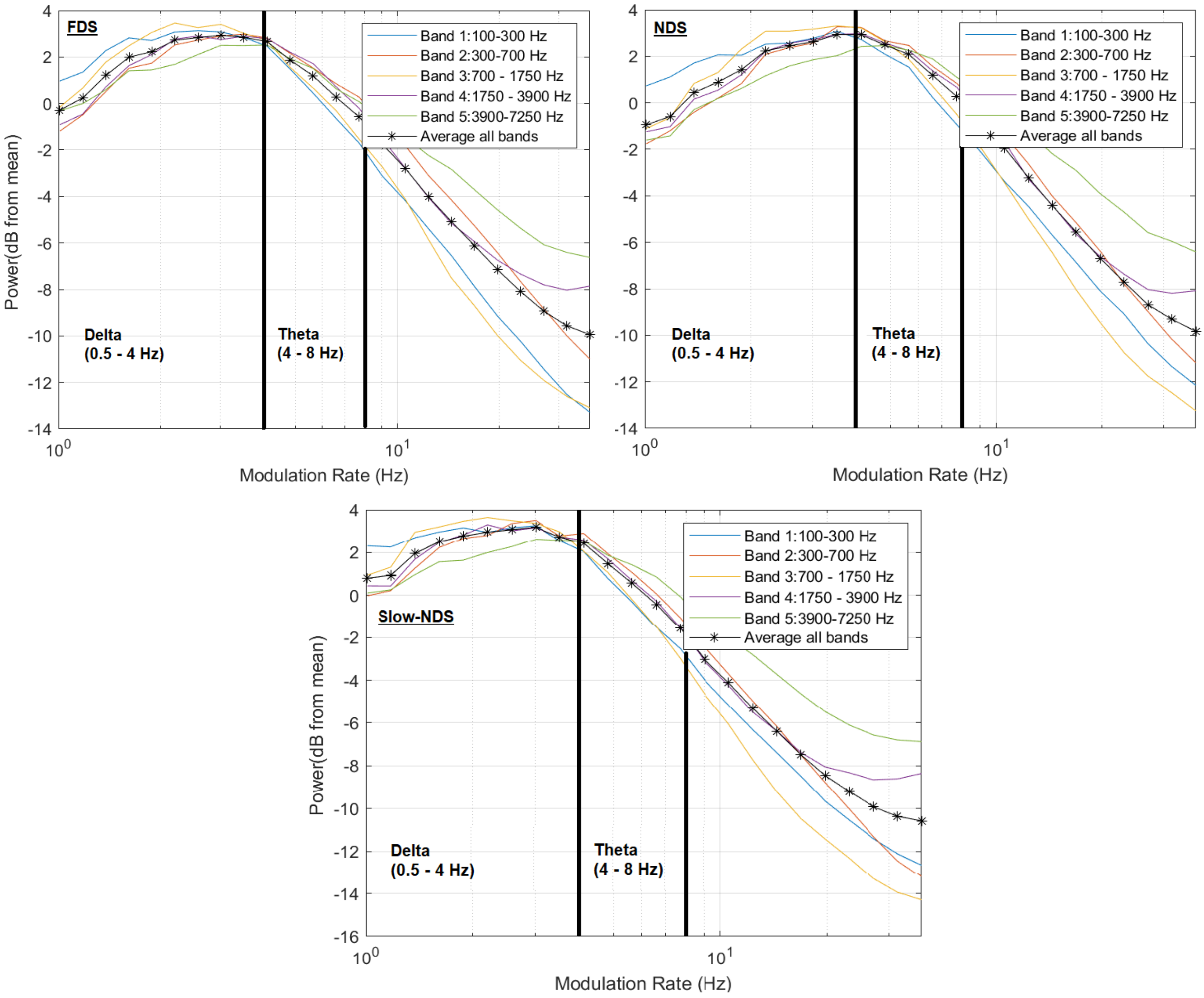


Figure 18. Amplitude Modulation Spectrum of the experimental stimuli by the three speech registers (FDS = Foreigner Directed Speech, NDS = Native Directed Speech, Slow-NDS = Slow-Native Directed Speech).

4.3.1.3. Equipment

Electroencephalography (EEG) data were recorded using a 64 Ag-AgCl electrodes standard setting (two actiCAP 64-channel systems, Brain Products GmbH, Germany). Electrode impedance was always kept below 15 kΩ, being below 10 kΩ in the large majority of the electrodes across participants. The EEG signal was amplified (BrainAmp DC, Brain Products GmbH, Germany), bandpass filtered between 0.05 and 500 Hz, digitised using a sampling rate of 1000 Hz and online referenced to the left earlobe. Psychopy 2021 (2.3; Peirce et al., 2019) was employed to present the stimuli and send triggers. Triggers were sent to indicate the start of each trial with contingent stimulus presentation and ensure synchronisation with EEG recordings.

4.3.1.4. Procedure

All EEG data were collected in a dimly lit and sound-proof booth. Stimuli were presented at a sampling rate of 44100 Hz, monophonically, and at a comfortable volume from Xiaomi Hybrid Mi In-Ear Pro HD headphones. Participants were asked to listen attentively to three stories while EEG signal was recorded. They were asked to sit calmly and upright while looking at a fixation cross, which was presented on the centre of a computer screen right in front of them (at ~80 cm of distance from their eyes). During the experimental session, participants were presented with one story per speech register, with counterbalanced order across participants. To avoid any effects derived from specific relations between stories and speech registers (e.g., a certain story is more interesting/easier to understand), each story was presented in all the speech registers across participants (with counterbalanced story-register association). The continuous narration of each story was divided into five shorter blocks of ~3 minutes each. At the end of each block, participants were asked 5 comprehension questions (15 questions per story, 45 questions in total). Experimental sessions lasted ~2 hours including preparation and testing.

4.3.2. Analysis

4.3.2.1. Behavioural data

Behavioural data were analysed to identify and discard those participants with very low accuracy, who did not pay a sustained level of attention throughout the experiment or who had very low English proficiency (they could not understand most of the stories). In addition, accuracy based on responses to the questionnaire was used as a proxy of participants' comprehension. Each question could be scored a finite number ranging between 0 to 1, which respectively represented wrong and correct answers. Most questions required to list multiple answers, which together summed 1 (see Supplementary material for a complete question list). If participants could recall only part of the possible answers (e.g., 1 out 4 elements) for a completely correct answer, they got a fraction score of 1 (e.g., 0.25 points; since 0.25 x 4 = 1).

4.3.2.2. EEG pre-processing

EEG signal analyses were performed on MATLAB (MathWorks, 2021b), using custom scripts, Fieldtrip toolbox functions (Oostenveld et al., 2011) and the mTRF toolbox (Crosse et al., 2016). Offline, the data were resampled to 100 Hz and band-pass filtered between 1 and 8 Hz with a Butterworth zero-phase filter (order 2+2). Channels with variance 3 times larger than the channels median variance were rejected by using the mTRF-toolbox in MATLAB (Crosse et al., 2016). Channels contaminated by noise were recalculated by spline interpolating the surrounding clean channels in EEGLAB (Delorme & Makeig, 2004). We had planned to discard from the analysis participants with more than 30% of rejected data or more than 4 contaminated electrodes, but no participants were discarded for these reasons.

4.3.2.3. EEG analysis

We investigated CT of the speech registers by using a filter that describes the linear transformation of the ongoing stimulus to the ongoing neural response, known as temporal response function (TRF; Figure 19). A given stimulus input at time 0 affects the neural signal in the following time windows. TRF measures the linear relationship between such an input and the neural response. In this case, TRFs were fitted to describe the linear mapping between speech features and EEG oscillatory activity within a given time window of interest for each EEG channel (Lalor et al., 2009). For describing how speech features in each speech register are tracked in EEG signal, we used a (m)TRF model that calculated the linear mapping between the stimulus features and EEG response vectors. This was done by means of regularised linear regression (Crosse et al., 2016), which is used to estimate a filter that predicts neural response from the stimulus features (forward – or encoding – model). Such a regression includes time-shifted versions of the stimulus features to test various lags at the same time. A ridge regularisation was included in the TRF model estimation to find the best regularisation parameter (λ). Ridge regularisation is a method to estimate the coefficients of multiple-regression models in dependent variables suffering from multicollinearity (Crosse et al., 2016).

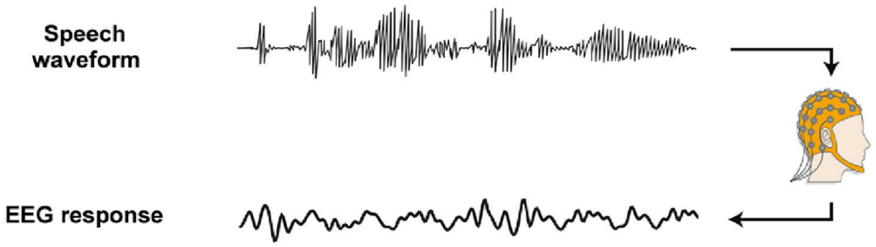


Figure 19. Taken from Di Liberto et al. (2021). Investigating the cortical encoding of speech registers with the temporal response function (TRF) analysis framework. Multichannel EEG signals were recorded as participants listened to audio stories in the three speech registers.

In order to train and validate the model, continuous EEG signal was split into folds of equal length. The leave-one-out cross-validation procedure was employed to verify the reliability of the TRF models. The encoding model was trained in n-1 folds and tested in the remaining one, repeating this process for 30 iterations per fold. This method quantified the EEG prediction correlation on unseen data based on the remaining EEG fitted data (Crosse et al., 2016). Each iteration provided a prediction correlation coefficient (r-value) between each feature and the EEG response (per channel). The prediction correlation coefficient is the estimate of how good an EEG signal encodes or tracks the stimulus speech feature added in the model as a regressor (e.g., speech envelope or semantic surprisal). R-values of 1 would represent perfect correspondence between EEG signal and TRF features, whereas r-values of 0 indicate no correlation whatsoever. It is important to stress that the prediction correlation values (Pearson's *r*) were extracted from the EEG signal, which is inherently noisy. That is, prediction correlation values have low values that are around ~ 0.1 yet being significant and informative (Di Liberto et al., 2015, 2021).

4.3.2.4. Tracking of speech features

By means of separate TRF models, we tested two speech representations: speech envelope and semantic surprisal in each story were used as regressors. This process provided us with speech register-level estimations of EEG prediction correlations. These reflected the brain oscillatory response to such features for every participant in the dataset.

For the speech envelope feature, we used a TRF model that included the amplitude envelope of speech between 1–8 Hz that was then extracted (Hilbert transform). In this case, the time window used to fit the TRF model was -200 – 600ms because previous research considered this lag to contain most relevant EEG responses to speech and its acoustic features (Broderick et al., 2018; Klimovich-Gray et al., 2023).

Conversely, for investigating tracking of semantic surprisal, we first calculated its values as the negative logarithmic probabilities extracted from the Generative Pre-trained Transformer 2 (GPT-2). GPT-2 calculated the probability of the upcoming words of each sentence of the stories, given the previous context. GPT-2 is a model pre-trained on a large corpus of English data in a self-supervised fashion. Surprisal values were then added in a time-distributed vector in correspondence of each word onset (t = 0) of all the stories (see Figure 20). The same alignment was done for word onset regressor, by putting ones (one hot encoding) in correspondence of word t = 0. We fitted a multivariate TRF that included semantic surprisal values and word onset information and a TRF that included word information only. Then, we subtracted prediction correlations of the univariate model to the multivariate model. This was done to get rid of the confounding factor derived from word acoustic boundaries and isolate the impact of the semantic surprisal on the EEG response. After that, we ran a one-sample t-test to assess whether the remaining r-values were significantly greater than zero (see the results section). We then used these remaining values to investigate the semantic tracking differences across the three speech registers.

The time window considered for these models was -200 – 700ms, because previous literature showed that semantic processing and surprisal (especially in L2 listeners) emerges with long latencies TRFs (Broderick et al., 2018; Di Liberto et al., 2021; Klimovich-Gray et al., 2023). Although TRF was modelled on the wider window, in previous research a 300-700ms lag was used to detect the N400-like response (Klimovich-Gray et al., 2023), which is associated with semantic processing. In the present study, the semantic mTRF was built on the whole -200 – 700ms window but the prediction correlation was cross-validated only in the 300-700ms lag to further get rid of the response to acoustic boundaries of words. Indeed, semantic surprisal vectors were aligned with word onsets, so that prediction correlation calculated on the whole window would include early acoustic response, not only semantic processing.

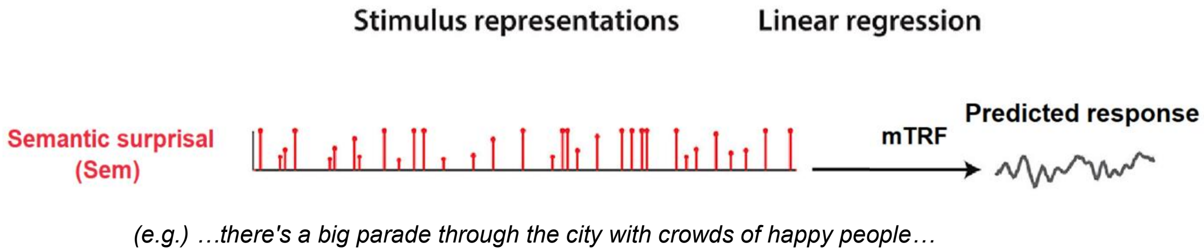


Figure 20. Semantic surprisal information was extracted from the stimulus and encapsulated into data vectors and matrices. Regularised linear regression was used to identify a linear fit that optimally predicted the EEG signal from features.

4.4. Statistical analysis

To assess the effect of Speech register, statistical analyses were performed using linear mixed (*lme*) effect models including effect of Speech register and participants (and electrodes for EEG data) as random effect (see Appendix 4.1 for a list of statistical models). For behavioural data (comprehension scores) the effect of Speech register was tested by fitting generalised linear mixed effects (*glme*) models with logit family. To determine significance of the models we used the type II Wald chi-square tests included in the CAR package (Fox, 2015; Fox & Weisberg, 2019). For post-hoc analyses, we used the *emmeans* package with Tukey HSD correction for multiple comparisons.

4.5. Results

Due to the features of FDS, SL were expected to show increased CT of FDS as compared to the other two speech registers, and higher in Slow-NDS than NDS. This was predicted to be reflected in higher comprehension scores and prediction correlations (both envelope and semantic surprisal TRFs) in FDS than Slow-NDS, followed by NDS.

4.5.1. Comprehension scores

The *glme* final model revealed a significant effect of Speech register on SL's comprehension accuracy (χ^2 = 34. 685, p < .001). Post-hoc analysis indicated that SL had higher comprehension scores in FDS than NDS (z = 4.793, p < .001) and Slow-NDS (z = 5.318, p < .001), whereas the latter two did not significantly differ (z = 0.530, p = .857; see Figure 21).

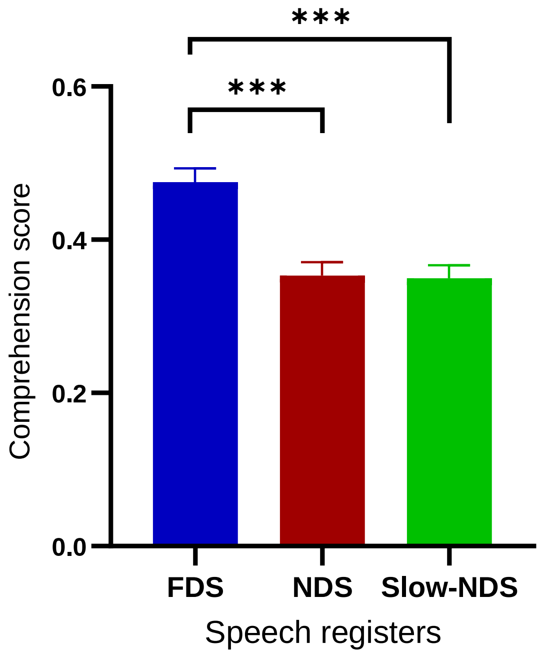


Figure 21. ***Spanish listeners group.*** Comprehension score by Speech Register (FDS = Foreigner-Directed Speech, NDS = Native-Directed Speech, Slow-NDS = Slow–Native-Directed Speech). Bars indicate SEM. Asterisks indicate significant differences (*p <.05, **p < .01, ***p < .001).

4.5.2. TRF results: Envelope model

The final model for the Envelope regressor yielded a significant effect of Speech register ($\chi^2 = 276.630$, $p < .001$). Post-hoc analysis revealed that EEG prediction correlations were higher when SL listened to FDS as compared to NDS ($z = 9.482$, $p < .001$; see Figure 22) and Slow-NDS ($z = 16.575$, $p < .001$), with NDS yielding higher r-values than Slow-NDS ($z = 7.092$, $p < .001$).

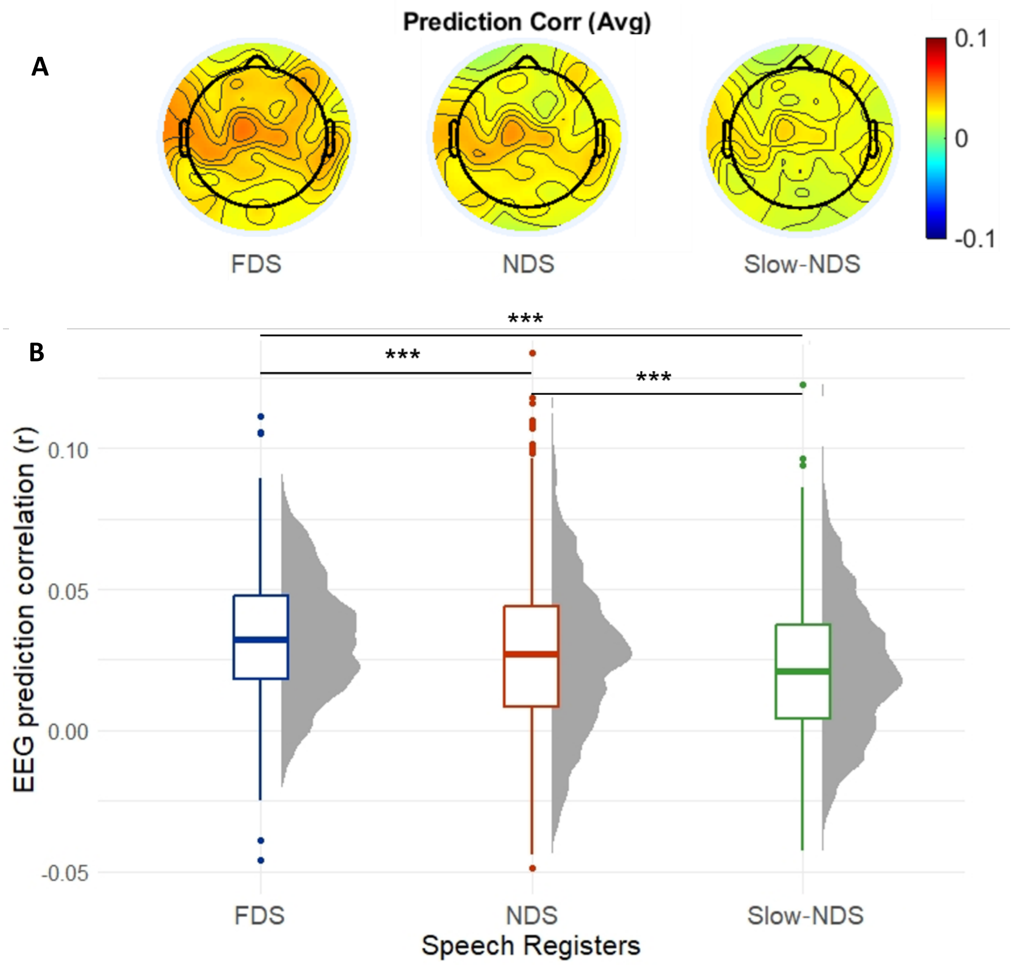


Figure 22. ***Spanish listeners group: Envelope.*** EEG prediction correlation scores (Pearson's r) of the Envelope model, by Speech Register (FDS = Foreigner-Directed Speech, NDS = Native-Directed Speech, Slow-NDS = Slow–Native-Directed Speech). (A) Topographical representation of the prediction correlations (averaged across participants). (B) Box & raincloud plots of the r-values distribution. Lines in the boxplots represent means. Asterisks indicate significant differences (*p <.05, **p < .01, ***p < .001).

4.5.3. mTRF results: Semantic surprisal model.

We tested the r-values difference between the multivariate model (semantic surprisal + word onset) and univariate model (word onset) against 0 and we found that the remaining r-values were significantly greater than 0 ($t = 2.103$, $df = 4799$, $p < 0.018$). This suggests that semantics was tracked to some extent by the EEG response. Such r-values were used in the LME analysis to test the differences between speech registers.

The final LME model for Semantic surprisal did not yield a significant effect of Speech register ($\chi^2 = 0.152$, $p = .927$), suggesting that there was no significant semantic tracking improvement when participants listened to any speech register (see Figure 23).

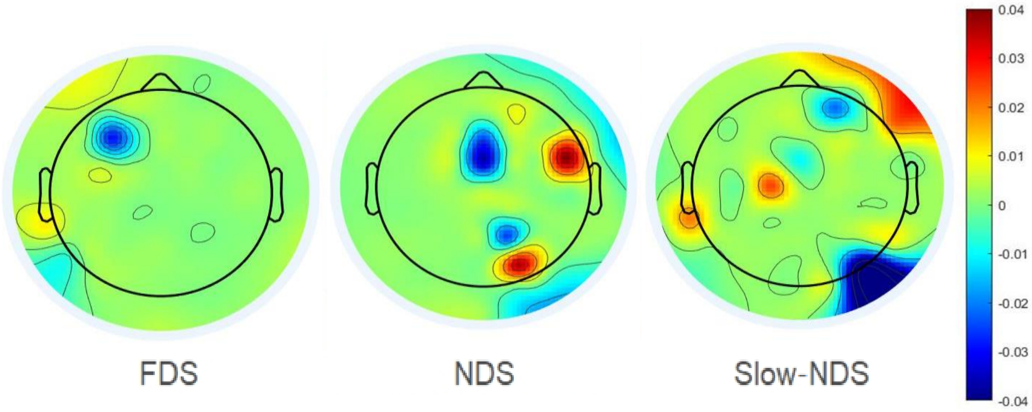


Figure 23. ***Spanish listers group: Semantic surprisal.*** Mean r-value differences between the multivariate TRF (word onset + semantic surprisal) and univariate TRF (word onset) models by Speech registers.

4.6. Discussion of Experiment 1

In this experiment we aimed to understand whether FDS promotes L2 perception and comprehension as compared to other speech registers. Here, Spanish listeners, novice learners of English, listened to three stories pronounced in FDS, NDS, and Slow-NDS, while their EEG activity was recorded. They also answered questions about the content of the stories as a proxy of language comprehension. We discuss the results for the three dependent variables we analysed in the following paragraphs.

We employed the (multivariate) temporal response function technique to measure cortical tracking of two speech features - envelope and semantics - in the three speech registers. In line with our hypotheses, participants showed higher EEG prediction correlations when they listened to FDS than NDS. This suggests that FDS promotes L2 processing for non-native listeners as compared to NDS. Conversely, we expected more efficient CT in Slow-NDS than NDS due to lower speech rate, which previous research found to affect CT (Kösem et al., 2018; Verschueren et al., 2022), but we actually observed the opposite effect. Such an unexpected result pattern could be due to the fact that low speech rate, if not accompanied by other acoustic features tailored to the listener (as in FDS), is detrimental for non-native listeners' perception.

Conversely, we did not find straightforward and objective evidence that FDS promotes L2 comprehension. In fact, in line with our hypothesis, participants responded with higher accuracy to questions about the FDS stories than stories in the other two registers. This suggests that non-native listeners benefitted from the exposure to FDS in their comprehension. However, questionnaire results can be affected by factors such as attention, engagement, and memory (Boyle, 1984; Hamouda, 2013). With this behavioural measure it was not possible to disentangle the comprehension advantage that led to higher accuracy scores from other factors. For this reason, we had decided to run an mTRF to measure the CT of semantic surprisal. Although we found evidence of semantic tracking (multivariate - univariate model difference was significantly greater than 0), we failed to detect any modulation of this across speech registers. This suggests that there was no comprehension improvement in any register. Thus, the accuracy improvement observed in the FDS condition is likely linked to other factors than pure comprehension benefit. For instance, perception was found to be boosted by FDS, this may have induced FDS stories to be more engaging, so that participants paid more attention to them.

We still cannot exclude that those results derive from the overall clarity of FDS, which would support both native and non-native listeners' speech perception, not just non-native listeners'. We argue that the fact that FDS is oriented to non-native listeners is the key point in this observed benefit, pointing at a link between FDS and non-native listeners. To rule this out, in experiment 2 we investigated whether such an FDS comprehension benefit is present in native listeners as well.

4.7. Experiment 2

In experiment 2 we tested whether the advantage originated from the exposure to FDS observed in the first experiment extended to native listeners.

4.7.1. Methods

4.7.1.2. Participants

Nineteen native speakers of English (English Listeners, henceforth EL; M_{age} = 21.40, y.o., SD = 2.50, Female = 4) were recruited to take part in the Experiment 2 and were tested at Trinity College Dublin (Ireland). One participant was excluded due to a technical error, leaving the final cohort to 18 participants. Of these, 16 were native listeners of Irish English and 2 of American English. Before starting the experimental procedure, all participants signed an informed consent form approved by the Trinity College Dublin Ethics Committee. Participants were paid 20 euros for taking part in the study.

4.7.1.3. Material

The stimuli employed for experiment 1 were also used for experiment 2. It is worth noting that the stories were recorded by a British English native speaker to make them more understandable to the SL (tested in experiment 1). This meant that English native listers (EL group) in experiment 2 were exposed to regional-accented speech. However, their understanding was likely not affected by that (see Results section).

4.7.1.4. Equipment

Data were acquired with a 64-channel device at a rate of 1024 Hz using an ActiveTwo system (BioSemi B.V., Netherlands). An additional external electrode was placed on participants' left earlobe for offline referencing. As for experiment 1, Psychopy 2021 (2.3) was employed to present the stimuli and send triggers.

4.7.1.5. Procedure

The same procedure employed for experiment 1 was followed for experiment 2.

4.8. Analysis

The exact same pre-processing and analysis of experiment 1, both for behavioural and EEG data, were conducted on the EL data of experiment 2. Also, the same TRF models and speech feature representations as in experiment 1 were employed to investigate the effect of FDS on native listeners' speech perception. Even though all these steps overlapped between experiment 1 and 2, two separate analyses were conducted because data were collected in two different laboratories and with different EEG recording systems (Brainvision and Biosemi).

4.9. Results

EL participants were expected to show close-to-ceiling performance in understanding all stories, that is, they were not expected to benefit from the exposure of any speech register in their comprehension accuracy and tracking of semantic surprisal. Also, EL participants were predicted to show increased CT of FDS and Slow-NDS envelope as compared to NDS envelope due to lower speech rate[4] (Kösem et al., 2018; Verschueren et al., 2022).

4.9.1. Comprehension scores

The *glme* final model did not reveal a significant effect of Speech register (χ^2 = 1. 589, p = .452), suggesting that EL's comprehension did not benefit from exposure of any speech register (see Figure 24).

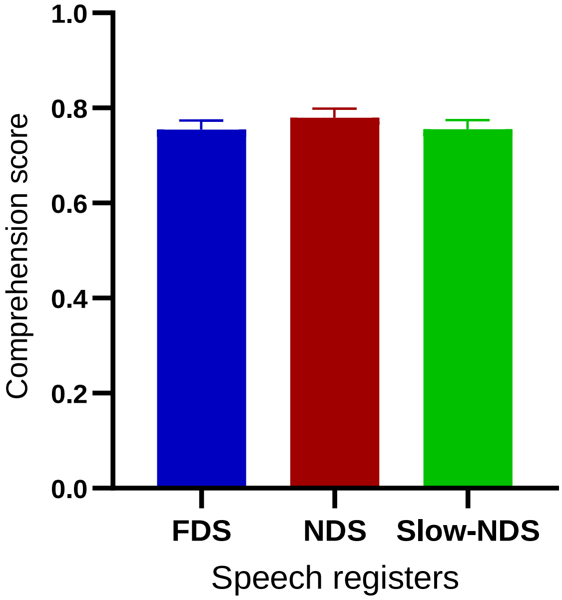


Figure 24. ***English listeners group.*** Comprehension score by Speech Register (FDS = Foreigner-Directed Speech, NDS = Native-Directed Speech, Slow-NDS = Slow–Native-Directed Speech). Bars indicate SEM.

4.9.2. TRF results: Envelope model

The final model for the Envelope regressor yielded a significant effect of Speech register (χ^2 = 12.040, p = .002). Post-hoc analysis indicated that prediction correlations were higher when EL where exposed to NDS than FDS (t = 2.793, p = .027) and Slow-NDS (t =

---

[4] Those were the original hypotheses before running experiment 1.

3.117, p = .004), and no difference between the latter two (t = 0.384, p = .922; see Figure 25).

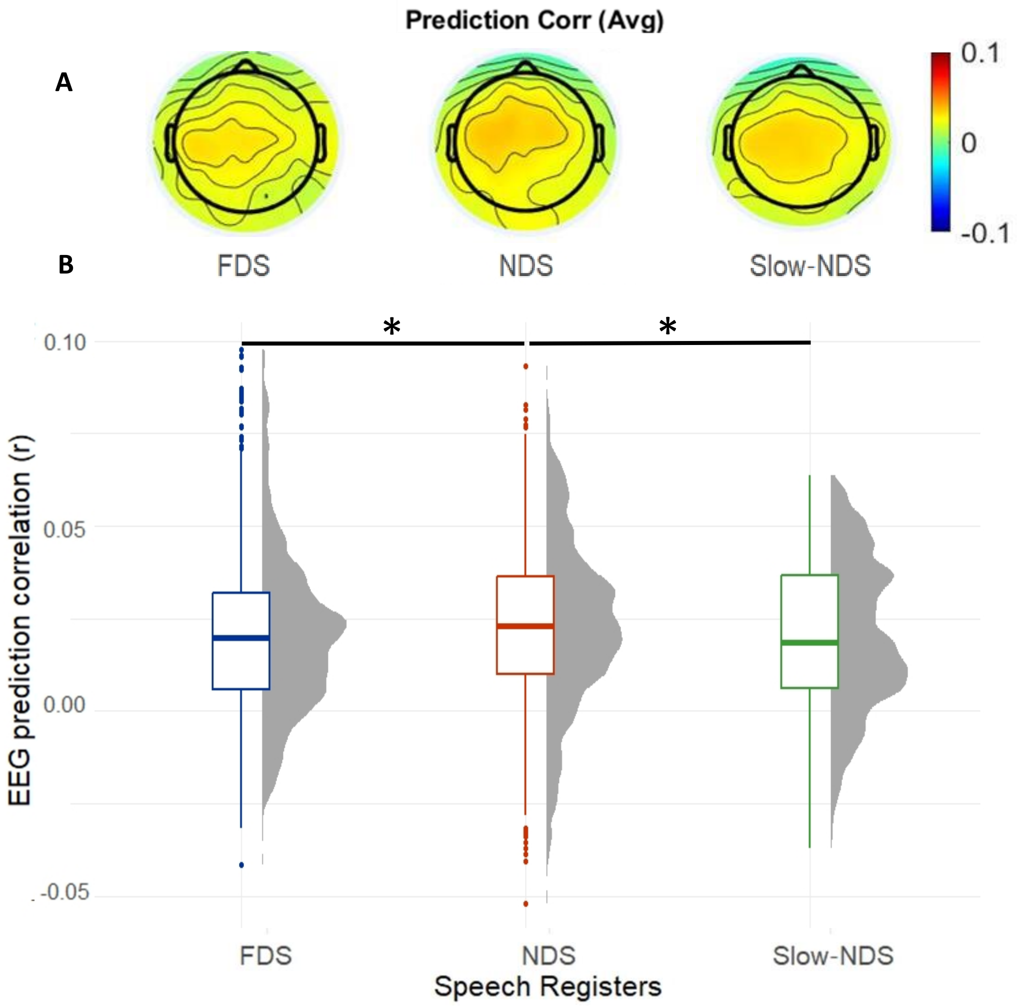


Figure 25. ***English listeners group: Envelope.*** EEG prediction correlation (Pearson's r) score of the Envelope model by Speech Register (FDS = Foreigner-Directed Speech, NDS = Native-Directed Speech, Slow-NDS = Slow–Native-Directed Speech). (A) Topographical representation of the prediction correlations (averaged across participants). (B) Box & raincloud plots of the r-values distribution. Lines in the boxplots represent means. Asterisks indicate significant differences (*p <.05, **p < .01, ***p < .001).

4.9.3. mTRF results: Semantic surprisal model.

We tested the r-values difference between the multivariate model (semantic surprisal + word onset) and univariate model (word onset) against 0 and we found that the remaining r-values were significantly greater than 0 (t = 4.996, df = 3263, p < .001). This suggests that semantics was tracked to some extent by the EEG response. Such r-values were used in the LME analysis. For EL, the model for the Semantic surprisal regressor did not yield a significant effect of Speech register (χ^2 = 4.108, p = .128), suggesting no differences across conditions (see Figure 26).

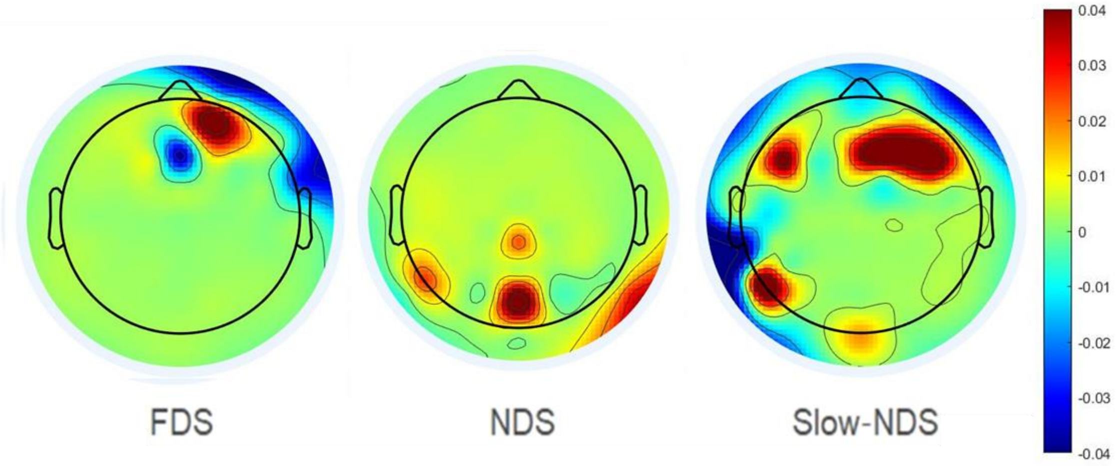


Figure 26. ***English listers group: Semantic surprisal.*** Mean r-value differences between the multivariate TRF (word onset + semantic surprisal) and univariate TRF (word onset) models by Speech registers.

4.10. Discussion of Experiment 2

With this experiment we aimed at understanding whether the benefit in processing and understanding FDS observed in the English learner group was also present in native English listeners. This would suggest that FDS is just a clear speech register that has no special status in supporting specifically non-native listeners' perception and comprehension.

We investigated participants' speech processing by calculating an envelope TRF model, which yielded higher prediction correlations when participants were exposed to NDS than FDS and Slow-NDS. This was in contrast with our hypotheses because we expected an effect of low speech rate, with more efficient CT of FDS and Slow-NDS (with no difference) than NDS. Instead, the opposite pattern emerged, hinting again (as in experiment 1) that speech rate did not affect native listeners' CT. Those results are also in line with experiment 1 because higher prediction correlations were associated to NDS, the speech register that a) ELs are more exposed to on a daily basis and b) for which ELs are the intended addressees. Also, this confirms that native listeners do not rely on FDS features to boost perception.

As for the SL group (experiment 1), here we also investigated whether FDS induces better understanding as compared to the other speech registers. To answer this question, we used two comprehension measures: a comprehension questionnaire and mTRF in response to semantic surprisal. In line with our predictions, ELs did not show benefit in any of the three speech registers for comprehension accuracy, this null effect not being due to ceiling effects. Congruently, prediction correlations did not show differences across speech registers either, which again suggests no comprehension benefit in any register. This is in line with their native proficiency in English.

4.11. General Discussion on Experiment 1 and 2

As reported in the previous chapters, in comparison with NDS, FDS is assumed to support various aspects of L2 acquisition, such as improving L2 perception and comprehension. However, to date it was unknown whether FDS promotes non-native listeners' cortical tracking of L2 speech, suggesting processing benefits, and whether FDS promotes L2 comprehension as well. Also, research on L2 processing has never assessed non-native listeners' perception in any register but NDS, which is not directed to low proficient L2 listeners. This represents a limitation to the generalisation of how L2 is processed in a naturalistic context. To shed light on whether FDS promotes non-native listeners' perception and comprehension, we conducted two EEG experiments: one with non-native listeners and one with native listeners of English. In the following, we describe the implications of these findings and provide a general discussion.

Remarkably, for speech perception, we observed different patterns of results in the two groups of participants. For SL (experiment 1), we found a more efficient envelope CT of FDS than NDS, whereas EL (experiment 2) showed a more efficient envelope CT of NDS as compared to FDS. This result has twofold implications. First of all, we found that non-native listeners' perception was indeed promoted by the exposure to continuous streams of FDS. This is the first time that the benefit of FDS in the form of continuous and naturalistic speech is assessed. Importantly, such an advantage is not extended to native listeners' perception. This leads to the second implication: here, we provide compelling evidence that listeners' envelope CT is enhanced when exposed to speech registers specifically intended for them. Thus, this finding indicates that speech accommodation affects the intended listeners' neural encoding and perception of speech. This highlights the importance of considering the relationship between speech register and target audience when investigating L1 and L2 processing and building models of speech communication.

On the other hand, we expected to observe an effect of speech rate in both experiments, with Slow-NDS showing increased CT as compared to NDS. We added this condition to disentangle the effect of speech rate from the features tailored to the intended addressees. Regarding SL, Slow-NDS yielded lower prediction correlations than NDS and FDS. It is possible that Slow-NDS did not reflect natural exposure to L2, as non-native listeners are usually exposed to either NDS (e.g., TV shows) or FDS (e.g., live interactions), but not to Slow-NDS. This could have induced higher CT in NDS than Slow-NDS in the first experiment. However, just by looking at this result, one could argue that this effect is due to Slow-NDS sounding unnatural. If this was the case, this same result should hold true for native listeners too. However, the pattern of experiment 2 does not point in that direction: ELs did not show significant differences between FDS and Slow-NDS. Even though ELs are never exposed to either FDS or Slow-NDS, they are likely able to adapt to slow speech due to their native proficiency, without hindering their speech perception. Conversely, as discussed above, non-native listeners may have difficulty at processing low speech rate if not supported by other features adjusted to their L2 proficiency. At the same time, our results probably do not reflect a mere amount of exposure in one particular speech register. In fact, nowadays non-native listeners’ exposure to English NDS is more common (e.g., TV shows) than exposure to FDS, which is limited to few occasions (e.g., classroom environment), because native listeners often do not employ FDS to address them (see Rothermich et al., 2021 for a description of this phenomenon in the USA). Also, some research found that higher amounts of exposure negatively affect envelope CT (lower CT; Perez-Navarro et al., in preparation). Thus, if anything we should have observed lower CT in the more familiar condition; but this was not the case. This does not exclude that amount of exposure plays some roles in speech perception, but it is unlikely that it drove the effect we observed.

To summarise, the special relation between speech register and intended listeners is fundamental for boosting speech perception, and amount of exposure does not seem to explain such an effect. This finding is important as a starting point for future research on communication tailored to interlocutors. Communication is not fixed, it is not the same for all listeners, and language communication is a dynamic relationship between speaker and listener that cooperate to keep communication successful.

We also aimed to understand whether FDS enhances comprehension as compared to NDS, and whether this happens only for non-native listeners. We found that SLs had a higher accuracy score in FDS as compared to the other registers, whereas ELs did not show accuracy differences across registers. The comprehension score indicates an FDS comprehension benefit for non-native listeners only. However, as already discussed, this could be driven by perception (envelope tracking), or engagement and attention, instead of reflecting simple comprehension benefits (Boyle, 1984; Hamouda, 2013). To further deepen this aspect, we investigated CT of semantic surprisal. With this analysis, in both experiments, we failed to find a modulation of encoding of semantic information in any register. One could ask why there was an improvement in non-native listeners' FDS accuracy score then. It is possible that the FDS benefit was only perceptual/attentional. That is, we argue that perception improvement (efficient CT) in the SL group led to higher accuracy scores because they relied on perceptual cues given their low L2 proficiency. Conversely, although EL perception was boosted by NDS, their ability to respond correctly to content questions was not affected due to their native proficiency.

4.12. Limitations and future directions

The experiment 2 employed native speakers of other English varieties but the British English accent. Even though they are likely less frequently exposed to this accent than their own, they showed more efficient CT of NDS than the other registers. We do not think that this negatively affected our results. In fact, ELs' accuracy score was high (~80%) and, given the proximity between Dublin and England, contact with that accent is quite frequent. However, future research should investigate native listeners' perception of NDS in their own accent.

Another aspect of our study warrants a side note. It is worth highlighting that our approach to investigating semantic tracking may have hindered the actual speech registers' effect because of being too conservative. In fact, TRF of word onset contains some semantic information that is subtracted from the multivariate model (word onset and semantic surprisal). Hence, future analysis should use another way to isolate the effect of semantic encoding. For instance, it is possible to use the EEG response weighted on the multivariate model, which allows separate responses to word onset and semantic surprisal. This will permit the measurement of the N400-mTRF response across speech registers, which is associated with semantic tracking (Broderick et al., 2018, 2022). At the moment we are working on extracting this measure from our data.4.13. Conclusion

This study on cortical tracking of speech showed that FDS supports non-native listeners’ speech perception, whereas natives’ perception was boosted by NDS. This study highlights the importance of considering the speech register's target audience and demonstrates the differential effects of FDS and NDS on language processing in both non-native and native listeners. Overall, this research sheds light on the role of FDS in supporting speech perception in non-native listeners. This study indicates that the speech register employed significantly impacts the degree to which listeners engage and process the speech information, highlighting the importance of considering the intended addressees when selecting a speech register. The findings have implications for language learning and teaching, emphasising the significance of tailoring language input to the intended audience.

Supplementary material

Material, data, experiment script and analysis code can be found at https://osf.io/ba3p4/?view_only=960986158dd94b92b3b31cca1839b58f. Instead, a list of the statistical formula can be found in Appendix 4.1.

Chapter 5 – General discussion

The focus of this dissertation is Foreigner Directed Speech (FDS), the speech register directed to non-native listeners. We explored FDS from the perspective of both production and perception, from speakers' speech adaptation to listeners' cortical tracking of FDS. With this work, we aimed at providing a well-rounded investigation on this understudied phenomenon, which has potential implications for several research fields related to language. The first and more straightforward implication of this research is its contribution to providing new methods and instruction for teaching a second language. This dissertation also provides useful insight for building new and more appropriate models of speech communication and production, and of L2 processing. In fact, in the previous chapters, the importance of the relationship between interlocutors and their collaboration for successful communication became evident. This aspect is likely a constant in everyday life communication, and for this reason this should be accounted for in communication models. In addition, most research on L2 perception focuses on standard registers (like Native Directed Speech), forgetting about the social factor and communicative intention of language interactions.

To study all these aspects related to FDS, in the first chapter, we started by defining the features of FDS and open questions on the production and perception sides. On the production side, we developed a production experiment to study how FDS is adjusted to communicative goals, which was described in the second chapter. As for the perception perspective, we developed 3 experiments to understand whether non-native listeners benefit from exposure to FDS in learning L2 novel words (both in production and recognition), in perceiving and comprehending an L2. These questions were investigated in chapters 3 and 4. In this section, we will first summarise the findings of the 4 experiments presented in this dissertation, then we will connect the implications of the findings with each other and relate them to the theory. We will then draw some conclusions indicating possible future directions.

5.1. Summary of the findings

In chapter 2, we investigated whether speakers adjust FDS and its acoustic features depending on their communicative goals and the listeners' profiles (i.e., native vs. non-native). We recorded native Spanish speakers naming novel objects to aid listeners' performance in comprehension, pronunciation, and writing tasks. Each speaker interacted with a native (NDS condition) and a non-native Spanish listener (FDS condition). Across the various acoustic features we considered in this experiment, we found that speakers dynamically adjust their speech to adapt to the communicative goals and their listeners' linguistic profile. We also found some evidence in support of differential adaptation of phonetic cues, which deliver linguistic information, and acoustic cues, which enhance salience of relevant parts of speech, without providing linguistic information. That is, speakers seem to intuitively adjust phonetic cues and acoustic cues to guarantee successful communication. These results support the didactic function assumption of FDS and extend its use to Spanish speakers.

In chapter 3, we investigated whether the exposure to FDS promotes L2 learning, both in perception and production, as compared to NDS. Spanish participants were asked to learn novel English words in an online experiment. Participants were divided in two groups: one being exposed to FDS and one to NDS. They carried out three tasks: a) Recognition task, b) Production task, and c) Continuum discrimination task. In a) participants were presented with novel objects and their auditory labels (e.g., "*This is a deest*") produced in FDS or NDS. Growth curve analyses showed that the FDS group recognized novel words faster and learned words containing a certain vowel contrast (/i - ɪ/) better than the NDS group. Also, FDS group's production for that same vowel contrast was more accurate than the NDS group. These findings support the didactic assumption of FDS, which should be considered in L2 teaching models.

In the appendix of this chapter, we also reported the results of a preliminary study that served as stimuli for the main experiment (Appendix 3.1). In this study, we demonstrated that speakers adjusted FDS by stressing the duration difference between vowels /I/ and /i/, whereas they reduced cues related to formants (formant Euclidean distance between vowels). On the contrary, formant distance was enhanced when addressing native listeners. This suggests that English FDS directed towards Spanish listeners was adjusted in a way that accommodated listeners' sensitivity and enhanced recognition of certain acoustic cues, namely durational cues. In fact, previous literature found Spanish listeners to be more sensitive to duration than spectral cues (Escudero, 2001).

In chapter 4, we conducted two experiments to investigate whether FDS promotes non-native listener's cortical tracking (CT) of speech. In the first experiment, we recorded the EEG activity of Spanish participants listening to continuous English speech in FDS, NDS, and Slow-NDS. We investigated CT of speech by using a filter that describes the linear transformation of the ongoing stimulus to the ongoing neural response, known as temporal response function (TRF). Here we found higher CT of FDS as compared to NDS followed by Slow-NDS, which suggested that non-native listeners' perception benefits from exposure to FDS. The second experiment employed the same procedure and material but tested native listeners to explore whether this FDS benefit expanded to native listeners. In this case we found higher CT in NDS as compared to the other conditions. Thus, FDS supports non-native listeners' speech perception, whereas native listeners' perception was boosted by NDS. Such results highlight the importance of considering the speech register's target audience and demonstrates the differential effects of FDS and NDS on language processing in both non-native and native listeners. Overall, this research sheds light on the role of FDS in supporting speech perception in non-native listeners.

5.2. Theoretical implications of the results

Production and communication

In this dissertation, we measured a large set of acoustic features to understand whether those can be adjusted to the listener's linguistic profile and communicative goals. In chapter 2, we indeed found that speech is adjusted in a way that accommodates listeners' linguistic profiles and communicative goals of the interaction. In the preliminary study reported in the appendix of chapter 3, we observed that different cues are provided depending on the acoustic sensitivity of the listener (e.g., Spanish listeners are more sensitive to durational than spectral cues in English). This supports the notion that speech accommodation is a dynamic and complex phenomenon influenced by various factors. Speech production seems to involve many ways to adjust acoustic features (and many acoustic features to adjust depending on speakers' L1) and each one is picked depending on the contextual advantage or needs. The evidence that different feature accommodations are implemented depending on various purposes, such as teaching writing, pronunciation, or simply facilitating communication, provides compelling evidence that accommodation is constantly at play and is an integral component in every speech interaction. Also, we shed some light on the role of phonetic and acoustic cues in speech accommodation. Speakers naturally use phonetic and acoustic cues without receiving overt instruction to do so. This further sustains the assumption that accommodation is a fundamental aspect of speech production, and it is imperative to account for that in models of production*.*

Indeed, this has consequences for various psycholinguistic models of speech production and communication. For instance, Bock & Levelt's sequential model (1994) of speech production consists of 4 levels of processing (but see also Levelt, 1989). This posits a sequence of processing stages involved in speech production, including conceptualization, functional processing (where lexical selection happens), positional processing (where grammatical inflection happens), and phonological encoding (where sound units, acoustic features are assembled). Bock and Levelt (1994) provide insights into the internal processes of speech production and the organisation of language production systems, but their model does not explicitly incorporate the concept of speech accommodation. Such adaptive processes are, so far, not well accounted for in sequential models of speech production.

Conversely, other influential psycholinguistic models explain speech production as parallel processing. For instance, Dell (1986) and Dell & O'Seaghdha's (1992) spreading activation model focuses on the activation and spreading of lexical and phonological representations during speech production. This model is known as Connectionist Model of speech production and, compared to sequential models, it claims that speech is produced by connected nodes representing units of speech (e.g., meaning, phonemes etc.). Those nodes are hypothesised to interact with each other in any direction, from the Semantic level, through the Lexical level, and the Phonological level. Although such models do not directly address FDS, our results (and assumptions) on the continuous accommodation to various contextual factors fit bidirectional communication between this model's nodes. Dell & Jacobs (2016) tackle this issue on speech accommodation by describing it as a phenomenon of short-term speaker tuning, as a type of implicit learning. Thus, Dell & O'Seaghdha's (1992) Connectionist Model is potentially able to incorporate contextual factors that modulate the spreading of activation, including the accommodation of acoustic features. This enables the model to capture how the activation of specific lexical and phonological representations is influenced by the listener's characteristics and communicative goals, leading to the adjustment of acoustic properties during speech production.

To fully account for speech accommodation, psycholinguistic models like these would benefit from explicitly incorporating mechanisms to account for the dynamic adjustment of acoustic features. This would allow the model to capture the flexible nature of accommodation, which is also based on continuous feedback, and its influence on the formulation stage, where linguistic information is transformed into an articulatory plan. That is, when addressing a non-native listener, speakers select one word among several others to vehiculate a certain meaning, they probably pick high frequency words (lexical selection) (lexical entrainment or alignment Brennan & Hanna, 2009; Clark & Wilkes-Gibbs, 1986; Costa et al., 2008), then plan articulation in a way that speech is clear to their listeners (Pardo, 2006 on phonetic convergence). Our research indeed shows that articulatory planning (phonological encoding) can result in adaptation of various acoustic features, without any changes in the vocabulary being used. That is, for example, a certain word is pronounced with wider or reduced vowel space, depending on communicative goals. In line with this, accommodation theories (e.g., Hyperarticulation & Hypoarticulation theory, Lindblom, 1990; see chapter 1) and the Interactive Alignment account (Pickering & Garrod, 2004) posit that during communication, individuals strive to establish a shared understanding and keep the communication successful through a process known as *accommodation* or *alignment* (depending on the theories under consideration; Costa et al., 2008; Giles; 2016; Lindblom, 1990; Pickering & Garrod, 2004, 2006, 2014). The accommodation perspective focuses on the speakers' adaptive process to meet listeners' needs, whereas the alignment account rather points at the continuous interaction between interlocutors (and their linguistic alignment). However, the underlying assumptions are common to both frameworks and our findings are in line with both models. Alignment (or accommodation) is assumed to occur at multiple levels, including linguistic, conceptual, and perceptual level (Costa et al., 2008), although in our studies we focused on the last one only. Alignment plays a crucial role in successful communication and collaboration. It allows individuals to understand each other's intentions and knowledge, leading to effective cooperation and coordination. The Interactive Alignment account suggests that alignment is an ongoing and dynamic process, which is line with our findings. As communication progresses, individuals continually adjust their linguistic, conceptual, and perceptual states to maintain alignment. This framework (and the accommodation theories) accounts for acoustic adjustments too. For instance, the Interactive Alignment account recognizes that successful communication involves not only aligning linguistic and conceptual aspects but also adjusting perceptual elements, such as acoustic features. When people engage in a conversation, they naturally adapt their speech patterns, including their pitch and rhythm, to align with their interaction partner. By adjusting their phonetic features, speakers enhance their ability to be understood and, as we reported, meet communicative goals.

In summary, psycholinguistic models of speech production should consider the dynamic adjustment of acoustic features observed in the dissertation. At present, accommodation models and Interactive alignment account are able to account for FDS accommodation results across communicative goals. By incorporating mechanisms that consider the listener's linguistic profile, communicative goals, and contextual factors, these models simulate the cognitive processes involved in accommodating acoustic features during speech production and communication.

Second language acquisition

In chapters 3 and 4, we addressed the FDS didactic impact and investigated whether FDS promotes L2 acquisition. We discovered that exposure to FDS facilitates L2 learning in both perception and production. L2 learners exposed to English FDS showed faster recognition of novel words and were more accurate in distinguishing specific vowel contrasts compared to NDS. FDS also aided non-native speakers’ production learning, although effectiveness depended on the relationship between target sounds and the learners' native language phonemic inventory (see Category Goodness vs Single in Chapter 3).

These results indicate that even a single session of exposure to FDS can yield benefits in L2 learning, suggesting that longer exposure may enhance L2 acquisition further. Notably, we did not only find evidence for FDS impact on word learning, but also on continuous speech perception, in a naturalistic listening paradigm. We found that FDS promoted non-native listeners' L2 perception, as the CT results suggest. Altogether, such findings reveal that FDS is a suitable tool for teaching an L2.

Our research has implications for models of second language acquisition. We found evidence supporting the asymmetry between phoneme contrast types (Goodness /i - ɪ/ and Single /æ - ʌ/ contrasts), aligning with Perceptual Assimilation Model-L2 (Best & Tyler, 2007) and other models, such as the Speech Learning Model (Flege, 1995; Flege & Bohn, 2021). For instance, PAM-L2 suggests that learners assimilate L2 sounds into their existing native language categories based on perceptual similarity. Our study's evidence for the phoneme contrast asymmetry supports the model predictions regarding how L2 learners perceive and categorise these specific contrasts.

In addition, other approaches to theories of second language learning exist. For instance, the Behaviourist theory (Bloomfield, 1944; Delprato & Midgley, 1992; Skinner, 1950) focuses on imitation, reinforcement, and practice as the primary mechanisms of language learning. Similarly, the Information Processing theory of second language acquisition (McLaughlin et al., 1983; Nyikos & Oxford, 1993), which was born to address the cognitive process involved in language learning, relates L2 learning with the process of encoding information into memory. The Information Processing theory posits that humans actively process information, rather than being passively exposed to linguistic stimuli. This approach suggests that L2 learners engage in cognitive processes such as attention, perception, memory, and problem-solving to acquire a second language. These models, together with those described above, do explain some of the mechanisms involved in L2 acquisition, which can be observed in our results too. However, such theories fail to capture the impact of social and interactive contexts on second language acquisition. They do not fully consider the sociocultural aspects of language learning that was highlighted in our research on FDS. In contrast, the socio-cognitive approach highlights the importance of social interaction, cultural context, and adaptive processes in language acquisition. In fact, our results can be explained by turning to the socio-cognitive theory of second language acquisition (Atkinson, 2002, 2014; Atkinson et al., 2007). This theory proposes that L2 learning involves adaptive processes influenced by social and interactive contexts (see also Chapter 3). This claims that L2 learning is a dynamic process influenced by both the learner's cognitive abilities and the social context in which language is acquired. L2 acquisition is not solely a result of individual cognitive processes or proximity to L1 phoneme inventory but is also shaped by the sociocultural and interpersonal aspects of language use. According to this approach, language learners actively engage in social interactions and negotiate meaning with others in order to develop their linguistic competence. The socio-cognitive approach emphasises the importance of social factors such as communication and collaboration during L2 acquisition. In line with this view, our results reveal that L2 learners adapt their perception (and production) of L2 novel words and speech to the social environment, indicating that FDS serves as a socially mediated promoter of phoneme and speech perception, and thus of L2 acquisition. Hence, we argue that the above reported models should consider a socio-cognitive factor, as speakers adapt linguistic elements to make understanding easier in social interactions.

Auditory speech perception

In chapter 4, our study on cortical tracking of speech showed that FDS supports non-native listeners' perception of continuous speech, whereas natives' perception was boosted by NDS. This finding aligns with the study in chapter 3, which focused on the acquisition of L2 vowel contrasts and their perception. Here, we showed that the speech register employed during communication significantly impacts the degree to which listeners engage in and process speech information. As it seems, speech accommodation affects the intended listeners' neural encoding and perception of speech. The findings have implications for models of auditory processing, emphasising the significance of tailoring language input to the intended audience. There is a need for models that take this aspect into account, especially models of L2 auditory processing. The reason is that accommodation seems to be particularly relevant for non-native listeners' perception, given their low L2 proficiency. Indeed, we found that non-native listeners' comprehension scores in chapter 4 were boosted only by FDS, whereas native listeners did not show such an advantage, not even in NDS, the speech register for which they had more efficient cortical tracking. Auditory speech perception models should consider the importance of how speech is pronounced, by building models that integrate various aspects of speech perception including facilitation derived from speech accommodation.

One of the most relevant models of speech perception is the so-called Motor Theory of speech perception (Liberman, et al., 1967; Liberman & Mattingly, 1985). This model suggests that speech perception is based on the activation of motor representations involved in speech production (Galantucci et al., 2006). In other words, listeners perceive speech by mapping the acoustic signal onto corresponding articulatory gestures. This suggests that listeners rely on their knowledge of speech production, specifically the motor representations involved in producing speech sounds, to perceive and understand speech. The Motor Theory of speech perception offers a relevant framework to understand our findings reported in chapter 3 and 4. In the context of FDS, the acoustic features are modified to facilitate perception. These adjustments likely enhance the mapping process between the acoustic signal and the corresponding articulatory gestures, which supports non-native listeners' perception and interpretation of speech. In line with this, in Chapter 3, we found that non-native speakers recognized novel words containing the Goodness contrast with higher accuracy when exposed to FDS than NDS. Importantly, they also pronounced the same novel words better (still containing the Goodness contrast; in line with Flege's (1995) assumption that production is connected to perception). Thus, our results provide empirical support for the Motor Theory of speech perception. When non-native listeners are exposed to FDS, both their L2 speech perception and production improve, suggesting a connection between perception of acoustic features and motor representations. Although such evidence is far from being conclusive, these findings contribute to a deeper understanding of the mechanisms underlying speech perception and highlight the role of motor processes in the context of second language perception.

5.3. Future directions

Although this dissertation provided some insight on various aspects of production and perception of FDS, there are many other questions that need responses. Future research should aim to answer those and contribute to speech production and L2 acquisition literature.

FDS production. To confirm our hypothesis that speech accommodation is active in any speech interaction, there is a need to investigate whether FDS can be adjusted to other factors than those we already assessed. For instance, we would hypothesise that different accommodation strategies may be at play depending on whether speakers have some experience with the listeners' L1. For instance, English speakers who live in Spain may adjust acoustic features depending on whether they address non-native listeners of English with Spanish L1 as compared to (e.g.) French L1. Also, further research on FDS production should aim to understand whether FDS is adjusted to listeners' proficiency and whether FDS is a continuum that goes from no adaptation (with native-like L2 listeners) to the maximum grade of speech modifications with recent L2 learners. If future studies will provide evidence for this, the assumption that accommodation is always at play in any instance of production will be further confirmed.

FDS perception. As we explained above, our study only looked at the impact of FDS after short exposure. FDS is used in everyday life interactions between natives and L2 learners, and even in L2 classroom environments. Thus, it is fundamental to understand benefits of long-term exposure and learning in ecologically valid situations (e.g., longitudinal study). As explained above, further learning benefit is highly probable after long exposure to FDS. However, it has also been hypothesised that L2 learners can reach high proficiency only if immersed in a native environment, where they are mainly exposed to NDS (Margić, 2017). Thus, future research should implement a longitudinal study with the objective of comparing exposure to FDS and NDS in both classroom environment and in naturalistic interactions.

5.4. Conclusion

We expanded the current understanding of how speech is produced and perceived in interactions between native and non-native speakers. We provided evidence for the didactic purpose and impact of FDS, for the dynamicity of speech production and its accommodation, and the importance of the relationship between speakers and listener during L2 perception.

In particular, our research revealed that speakers adapt their speech characteristics to aid the acquisition of specific second language (L2) skills. This shows that FDS is endowed with a didactic purpose. This adjustment was found to facilitate the learning of new words in the L2. Additionally, our findings indicated that FDS also enhances the perception of connected speech, as indicated by cortical tracking results. The findings have implications for language learning and teaching, emphasising the significance of tailoring language input to the intended audience.

References

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America*, *141*(5), 3070–3078. https://doi.org/10.1121/1.4982246

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13367–13372. https://doi.org/10.1073/pnas.201400998

Al-Tamami, J., & Ferragne, E. (2005). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. *INTERSPEECH-2005*, 2465–2468.

Anderson, N., Meagher, K., Welder, A., & Graham, S. A. (2018). Animacy cues facilitate 10-month-olds' categorization of novel objects with similar insides. *PLOS ONE*, *13*(11), e0207800. https://doi.org/10.1371/journal.pone.0207800

Andruski, J. E., Kuhl, P. K., & Hayashi, A. (1999). Point vowels in Japanese mothers' speech to infants and adults. *The Journal of the Acoustical Society of America*, *105*(2), 1095–1096. https://doi.org/10.1121/1.425135

Antón, E., Thierry, G., Dimitropoulou, M., & Duñabeitia, J. A. (2020). Similar Conceptual Mapping of Novel Objects in Mixed- and Single-Language Contexts in Fluent Basque-Spanish Bilinguals. *Language Learning*, *70*(S2), 150–170. https://doi.org/10.1111/lang.12397

Antón, E., Thierry, G., & Duñabeitia, J. A. (2015). Mixing Languages during Learning? Testing the One Subject—One Language Rule. *PLOS ONE*, *10*(6), e0130069. https://doi.org/10.1371/journal.pone.0130069

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2020). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01501-5

Aoyama, K., & Flege, J. E. (2011). Effects of L2 experience on perception of English /r/ and /l/ by native Japanese speakers. *Journal of the Phonetic Society of Japan, 15 (3)*, 5–13. https://doi.org/10.24467/ONSEIKENKYU.15.3_5

Arthur, B., Wemer, R., Culver, M., Lee, Y. J., & Thomas, D. (1980). The Register of Impersonal Discourse to Foreigners: Verbal Adjustments to a Foreign Accent. *Discourse Analysis in Second Language Research*, 111–124.

Atkinson, D. (2002). Toward a Sociocognitive Approach to Second Language Acquisition. *The Modern Language Journal*, *86*(4), 525–545.

Atkinson, D. (2014). Language Learning in Mindbodyworld: A Sociocognitive Approach to Second Language Acquisition. *Language Teaching*, *47*(4), 467–483. https://doi.org/10.1017/S0261444813000153

Atkinson, D., Churchill, E., Nishino, T., & Okada, H. (2007). Alignment and Interaction in a Sociocognitive Approach to Second Language Acquisition. *The Modern Language Journal*, *91*(2), 169–188. https://doi.org/10.1111/j.1540-4781.2007.00539.x

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., & Goswami, U. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, *247*, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698

Audibert, N., & Falk, S. (2018). Vowel space and f0 characteristics of infant-directed singing and speech. *9th International Conference on Speech Prosody 2018*, 153–157. https://doi.org/10.21437/SpeechProsody.2018-31

Austerlitz, R. (1956). Gilyak Nursery Words. *WORD*, *12*(2), 260–279. https://doi.org/10.1080/00437956.1956.11659604

Baigorri, M., Campanelli, L., & Levy, E. S. (2019). Perception of American–English Vowels by Early and Late Spanish–English Bilinguals. *Language and Speech*, *62*(4), 681–700. https://doi.org/10.1177/0023830918806933

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication*, *46*(3–4), 252–267. https://doi.org/10.1016/j.specom.2005.02.016

Barry, T. J., Griffith, J. W., De Rossi, S., & Hermans, D. (2014). Meet the Fribbles: Novel stimuli for use within behavioural research. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00103

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), Article 1. https://doi.org/10.18637/jss.v067.i01

Beebe, L. M., & Giles, H. (1984). Speech-accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language*, *46*, 5–32. https://doi.org/10.1515/ijsl.1984.46.5

Benders, T. (2013). Mommy is only happy! Dutch mothers’ realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior & Development*, *36*(4), 847–862. https://doi.org/10.1016/j.infbeh.2013.09.001

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*(8), 706–712. https://doi.org/10.1111/j.1467-9280.2007.01964.x

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege.* https://doi.org/10.1121/1.1332378.

Biersack, S., Kempe, V., & Knapton, L. (2005). Fine-tuning speech registers: A comparison of the prosodic features of child-directed and foreigner-directed speech. *INTERSPEECH 2005*. https://doi.org/10.13140/2.1.2133.5049

Blaauw, E. (1992). Phonetic differences between read and spontaneous speech. *ICSLP-1992*, 751–754.

Bloomfield, L. (1933). *Language*. Holt, Rinehart &Winston.

Bloomfield, L. (1944). Secondary and Tertiary Responses to Language. *Language*, *20*(2), 45–55. https://doi.org/10.2307/409893

Bobb, S. C., Mello, K., Turco, E., Lemes, L., Fernandez, E., & Rothermich, K. (2019). Second Language Learners' Listener Impressions of Foreigner-Directed Speech. *Journal of Speech, Language, and Hearing Research: JSLHR*, *62*(9), 3135–3148. https://doi.org/10.1044/2019_JSLHR-S-18-0392

Bock, K., & Griffin, Z. M. (2000). Producing Words: How Mind Meets Mouth. *Aspects of Language Production*, 7–47. Pychology Press.

Bock, K., & Levelt, W. J. M. (1994). Language Production: Grammatical Encoding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics,* pp. 945-984. San Diego, CA: Academic Press.

Boersma, P., & Weenink, D. (2001). PRAAT, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Boomershine, A. (2013). The Perception of English Vowels by Monolingual, Bilingual, and Heritage Speakers of Spanish and English. In C. Howe, S. E. Blackwell, & M. L. Quesada (Eds.), *Selected Proceedings of the 15th Hispanic Linguistics Symposium* (pp. 103–118). Cascadilla Proceedings Project. http://www.lingref.com/cpp/hls/15/abstract2879.html

Bourdin, B., & Fayol, M. (1994). Is Written Language Production more Difficult than Oral Language Production? A Working Memory Approach. *International Journal of Psychology*, *29*(5), 591–620. https://doi.org/10.1080/00207599408248175

Bourdin, B., & Fayol, M. (2002). Even in adults, written production is still more costly than oral production. *International Journal of Psychology*, *37*(4), 219–227. https://doi.org/10.1080/00207590244000070

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *Vol. 26*(N° 2), 211–252.

Boyle, J. P. (1984). Problems in Testing Listening Comprehension. *World Englishes*, *3*(1), 44–48. https://doi.org/10.1111/j.1467-971X.1984.tb00570.x

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, *97*(3), 1916–1924. https://doi.org/10.1121/1.412064

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.*, *112*(1), 13.

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research: JSLHR*, *46*(1), Article 1. https://doi.org/10.1044/1092-4388(2003/007)

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Current Biology: CB*, *28*(5), 803-809.e3. https://doi.org/10.1016/j.cub.2018.01.080

Broderick, M. P., Zuk, N. J., Anderson, A. J., & Lalor, E. C. (2022). More than words: Neurophysiological correlates of semantic dissimilarity depend on comprehension of the speech narrative. *European Journal of Neuroscience*, *56*(8), 5201–5214. https://doi.org/10.1111/ejn.15805

Burnham, D., Joeffry, S., & Rice, L. (2010). “D-o-e-s-Not-C-o-m-p-u-t-e”: Vowel Hyperarticulation in Speech to an Auditory-Visual Avatar. *Anthrozoos A Multidisciplinary Journal of The Interactions of People & Animals*, *26*(3), 373–380. https://doi.org/10.2752/175303713x13697429463592

Burnham, D., Kim, J., Davis, C., Ciocca, V., Schoknecht, C., Kasisopa, B., & Luksaneeyanawin, S. (2011). Are tones phones? *Journal of Experimental Child Psychology*, *108*(4), 693–712. https://doi.org/10.1016/j.jecp.2010.07.008

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What’s New, Pussycat? On Talking to Babies and Animals. *Science*, *296*(5572), 1435–1435. https://doi.org/10.1126/science.1069587

Burnham, E. B., Wieland, E. A., Kondaurova, M. V., McAuley, J. D., Bergeson, T. R., & Dilley, L. C. (2015). Phonetic modification of vowel space in storybook speech to infants up to 2 years of age. *Journal of Speech, Language, and Hearing Research: JSLHR*, *58*(2), 241–253. https://doi.org/10.1044/2015_JSLHR-S-13-0205

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers’ subsequent pronunciations. *Journal of Memory and Language*, *89*, 68–86. https://doi.org/10.1016/j.jml.2015.12.009

Buzsáki, G., & Draguhn, A. (2004). Neuronal Oscillations in Cortical Networks. *Science*, *304*(5679), 1926–1929. https://doi.org/10.1126/science.1099745

Canale, M., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, *I*. https://doi.org/10.1093/applin/I.1.1

Caporael, L. R. (1981). The paralanguage of caregiving: Baby talk to the institutionalized aged. *Journal of Personality and Social Psychology*, *40*(5), 876–884. https://doi.org/10.1037/0022-3514.40.5.876

Chan, H., Verspoor, M., & Vahtrick, L. (2015). Dynamic Development in Speaking Versus Writing in Identical Twins. *Language Learning*, *65*(2), 298–325. https://doi.org/10.1111/lang.12107

Chaudron, C. (1978). Foreigner talk in the classroom—An aid to learning. *Classroom Oriented Research in Second Language Acquisition*, 127–143.

Chaudron, C. (1979). Complexity of teacher speech and vocabulary explanation/elaboration. *In 13th Annual TESOL Convention*.

Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology With Inclusive and Accurate Measures. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.715843

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*(6), 967–978. https://doi.org/10.1037/0012-1649.38.6.967

Clark, N. B., McRoberts, G. W., Van Dyke, J. A., Shankweiler, D. P., & Braze, D. (2012). Immediate memory for pseudowords and phonological awareness are associated in adults and pre-reading children. *Clinical Linguistics & Phonetics*, *26*(7), 577–596. https://doi.org/10.3109/02699206.2012.673045

Clyne, M. G. (1981). ‘Second Generation’ Foreigner Talk in Australia. *International Journal of the Sociology of Language*, *1981*(28), 69–80. https://doi.org/10.1515/ijsl.1981.28.69

Cooke, M., & Lecumberri, M. L. G. (2012). The intelligibility of Lombard speech for non-native listeners. *The Journal of the Acoustical Society of America*, *132*(2), 1120–1129. https://doi.org/10.1121/1.4732062

Cooper, R., P. ,., & Aslin, R., N. (1990). Preference for Infant Directed Speech in the First Month after Birth. *Child Development*, *61*, 1584–1595. https://doi.org/0093920/90/6105-0011$01.00

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes*, *23*(4), Article 4. https://doi.org/10.1080/01690960801920545

Coupland, N., Coupland, J., Giles, H., & Henwood, K. (1988). Accommodating the elderly: Invoking and extending a theory. *Language in Society*, *17*(1), 1–41. https://doi.org/10.1017/S0047404500012574

Cristia, A. (2013). Input to Language: The Phonetics and Perception of Infant-Directed Speech: The Phonetics and Perception of Infant-Directed Speech. *Language and Linguistics Compass*, *7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913–934. https://doi.org/10.1017/S0305000912000669

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnhum.2016.00604

Dahl, D. A. (1981). The role of experience in speech modifications for second language learners. *Minn. Pap. Linguist. Philos. Lang.*, *7*(2), 78–93.

de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1), 287–314. https://doi.org/10.1016/0010-0277(92)90046-K

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Delprato, D., & Midgley, B. D. (1992). Some fundamentals of B. F. Skinner's behaviorism. *American Psychologist*, *47*, 1507–1520. https://doi.org/10.1037/0003-066X.47.11.1507

Depaulo, B. M., & Coleman, L. M. (1981). Evidence for the Specialness of the "Baby Talk" Register. *Language and Speech*, *24*(3), 223–231. https://doi.org/10.1177/002383098102400302

Depaulo, B. M., & Coleman, L. M. (1987). Verbal and nonverbal communication of warmth to children, foreigners, and retarded adults. *Journal of Nonverbal Behavior*, *11*(2), 75–88. https://doi.org/10.1007/BF00990959

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476–490. https://doi.org/10.1017/S026144480800551X

Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Cortical Measures of Phoneme-Level Speech Encoding Correlate with the Perceived Clarity of Natural Speech. *ENeuro*, *5*(2). https://doi.org/10.1523/ENEURO.0084-18.2018

Di Liberto, G. M., Nie, J., Yeaton, J., Khalighinejad, B., Shamma, S. A., & Mesgarani, N. (2021). Neural representation of linguistic feature hierarchy reflects second-language proficiency. *NeuroImage*, *227*, 117586. https://doi.org/10.1016/j.neuroimage.2020.117586

Diehl, R. L., & Lindblom, B. (2004). Explaining the Structure of Feature and Phoneme Inventories: The Role of Auditory Distinctiveness. In S. Greenberg, W. A. Ainsworth,

A. N. Popper, & R. R. Fay (Eds.), *Speech Processing in the Auditory System* (pp. 101–162). Springer. https://doi.org/10.1007/0-387-21575-1_3

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Current Biology*, *25*(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), Article 1. https://doi.org/10.1038/nn.4186

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*. https://www.frontiersin.org/article/10.3389/fnhum.2014.00311

Dodane, C., & Al-Tamimi, J. (2007). An acoustic comparison of vowel systems in adult-directed-speech and child-directed speech. *16th International Congress of Phonetics Sciences (ICPhS), Saarbrücken, 6-10 Août 2007*. https://hal.archives-ouvertes.fr/hal-03068849

Downey, S. S., Hallmark, B., Cox, M. P., Norquest, P., & Lansing, J. S. (2008). Computational Feature-Sensitive Reconstruction of Language Relationships: Developing the ALINE Distance for Comparative Historical Linguistic Reconstruction. *Journal of Quantitative Linguistics*, *15*(4), 340–369. https://doi.org/10.1080/09296170802326681

Downey, S. S., Sun, G., & Norquest, P. (2017). alineR: An R Package for Optimizing Feature-Weighted Alignments and Linguistic Distances. *The R Journal*, *9*(1), 138–152.

Drozdova, O. A., Zamyatina, E. V., Volodina, D. N., Zakharova, E. O., Ruchina, A. V., & Nepryakhin, A. F. (2015). Situational Communication in Teaching Russian as a Foreign Language to Beginner Learners. *Procedia - Social and Behavioral Sciences*, *215*, 118–126. https://doi.org/10.1016/j.sbspro.2015.11.584

Drozdova, P., Hout, R. V., & Scharenborg, O. (2016). Lexically-guided perceptual learning in non-native listening*. *Bilingualism: Language and Cognition*, *19*(5), 914–920. https://doi.org/10.1017/S136672891600002X

Englund, K. T., & Behne, D. M. (2005). Infant Directed Speech in Natural Interaction—Norwegian Vowel Quantity and Quality. *Journal of Psycholinguistic Research*, *34*(3), 259–280. https://doi.org/10.1007/s10936-005-3640-7

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. LOT.

Escudero, P. (2006). The phonological and phonetic development of new vowel contrasts in Spanish learners of English. *English with a Latin Beat*, 41–55.

Escudero, P. R. (2001). The Role of the Input in the Development of L1 and L2 Sound Contrasts: Language-Specific Cue Weighting for Vowels. In *Proceedings of the 25th Annual Boston University Conference on Language Development, Vols. 1-2* (pp. 250–261). Cascadilla Press.

Etard, O., & Reichenbach, T. (2019). Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *39*(29), 5750–5759. https://doi.org/10.1523/JNEUROSCI.1828-18.2019

Fairs, A., & Strijkers, K. (2021). Can we use the internet to study speech production? Yes we can! Evidence contrasting online versus laboratory naming latencies and errors. *PLOS ONE*, *16*(10), e0258908. https://doi.org/10.1371/journal.pone.0258908

Farley, S. D., Hughes, S. M., & LaFayette, J. N. (2013). People Will Know We Are in Love: Evidence of Differences Between Vocal Samples Directed Toward Lovers and Friends. *Journal of Nonverbal Behavior*, *37*(3), Article 3. https://doi.org/10.1007/s10919-013-0151-3

Ferguson, C. A. (1975). Toward a Characterization of English Foreigner Talk. *Anthropological Linguistics*, *17*(1), Article 1.

Ferguson, C. A. (1981). 'Foreigner Talk' as the Name of a Simplified Register. *International Journal of the Sociology of Language*, *1981*(28), 9–18. https://doi.org/10.1515/ijsl.1981.28.9

Ferguson, C. A. (1971). Absence of Copula and the Notion of Simplicity. A Study of Normal Speech, Baby Talk, Foreigner Talk and Pidgins. *Cambridge University Press*, 141–150.

Ferguson, H. S., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *135*(6), 3570–3584. https://doi.org/10.1121/1.4874596

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, *112*(1), 13.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development*, *60*(6), 1497–1510. https://doi.org/10.2307/1130938

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development*, *10*(3), 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/s0305000900010679

Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, *76*(3), 692–707. https://doi.org/10.1121/1.391256

Flege, J. E. (1995). Second Language Speech Learning. Theory, Findings, and Problems. *In Winifred Strange (CditOf), Speech Perception and Linguistic Experience: Issues in Cross-Language Research. Timonium*.

Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model. *Second Language Speech Learning: Theoretical and Empirical Progress*. https://doi.org/10(9781108886901.002).

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., & Liu, S. (2001). The Effect of Experience on Adults' Acquisition of a Second Language. *Studies in Second Language Acquisition*, *23*(4), 527–552. https://doi.org/10.1017/S0272263101004041

Foley, C., & Flynn, S. (2013). The role of the native language. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge Handbook of Second Language Acquisition* (pp. 97–113). Cambridge University Press. https://doi.org/10.1017/CBO9781139051729.008

Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models* (3° edizione). SAGE Publications, Inc.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed., Vol. 1–1). Sage. https://us.sagepub.com/en-us/nam/an-r-companion-to-applied-regression/book246125

Gaies, S. J. (1979). Linguistic Imput in Formal Second Language Learning: The Issues of Syntactic Gradation and Readability in ESL Materials. *TESOL Quarterly*, *13*(1), 41–50. https://doi.org/10.2307/3585974

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377.

Gallego, J. C. (1990). The Intelligibility of Three Nonnative English-Speaking Teaching Assistants: An Analysis of Student-Reported Communication Breakdowns. *Issues in Applied Linguistics*, *1*(2), 219–250.

García-Sierra, A., Ramírez-Esparza, N., Wig, N., & Robertson, D. (2021). Language learning as a function of infant directed speech (IDS) in Spanish: Testing neural commitment using the positive-MMR. *Brain and Language*, *212*, 104890. https://doi.org/10.1016/j.bandl.2020.104890

Garnier, M., Bailly, L., Dohen, M., Welby, P., & Loevenbruck, H. (2006). An acoustic and articulatory study of Lombard speech: Global effects on the utterance. *INTERSPEECH*.

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language*, *28*(2), 580–597. https://doi.org/10.1016/j.csl.2013.07.005

Garnier, M., Ménard, L., & Alexandre, B. (2018). Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues? *The Journal of the Acoustical Society of America*, *144*(2), 1059–1074. https://doi.org/10.1121/1.5051321

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory (Hove, England)*, *2*(2), 103–127. https://doi.org/10.1080/09658219408258940

Gershkoff-Stowe, L., & Hahn, E. R. (2013). Word comprehension and production asymmetries in children and adults. *Journal of Experimental Child Psychology*, *114*(4), 489–509. https://doi.org/10.1016/j.jecp.2012.11.005

Giles, H. (2016). Communication Accommodation Theory. In *The International Encyclopedia of Communication Theory and Philosophy* (pp. 1–7). American Cancer Society. https://doi.org/10.1002/9781118766804.wbiect056

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Editions de la Maison des Sciences de l’Homme. https://doi.org/10.1017/CBO9780511663673.001

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Golinkoff, R. M., & Alioto, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition*. *Journal of Child Language*, *22*(3), 703–726. https://doi.org/10.1017/S0305000900010011

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. https://doi.org/10.1016/j.tics.2010.10.001

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal*, *23*, 995–1019. https://doi.org/10.1007/s11145-009-9186-6

Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, *4*(1), 67–92. https://doi.org/10.1515/lp-2013-0004

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(16), 10911–10916. https://doi.org/10.1073/pnas.122368599

Grabowski, J. (2005). Replicating the Bourdin and Fayol results. *Proceedings Writing2004*.

Graf Estes, K., & Hurley, K. (2013). Infant-Directed Prosody Helps Infants Map Sounds to Meanings. *Infancy*, *18*(5), 797–824. https://doi.org/10.1111/infa.12006

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., & Brattico, E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: A behavioral and ERP study. *Frontiers in Human Neuroscience*, *8*. https://www.frontiersin.org/articles/10.3389/fnhum.2014.00279

Hamouda, A. (2013). *An Investigation of Listening Comprehension Problems Encountered by Saudi Students in the EL Listening Classroom*. https://www.semanticscholar.org/paper/An-Investigation-of-Listening-Comprehension-by-in-Hamouda/b811984d6e30068a62a970b1f75b2e701e0b159e

Han, M., de Jong, N. H., & Kager, R. (2018). Lexical Tones in Mandarin Chinese Infant-Directed Speech: Age-Related Changes in the Second Year of Life. *Frontiers in Psychology*, *9*, 434. https://doi.org/10.3389/fpsyg.2018.00434

Hartman, K. M., Ratner, N. B., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of Child Language*, *44*(5), 1140–1162. https://doi.org/10.1017/S0305000916000520

Hatch, E. (1979). Interaction, input and communication strategies. *First Nordic Symposium on Interlanguage*.

Hatch, E., Shapira, R., & Gough, J. (1978). “Foreigner-Talk” Discourse. *ITL - International Journal of Applied Linguistics*, *39*(1), Article 1. https://doi.org/10.1075/itl.39-40.03hat

Hazan, V., & Baker, R. (2010). Does reading clearly produce the same acoustic-phonetic modifications as spontaneous speech in a clear speaking style? *Proceedings of DiSS-LPSS Joint Workshop 2010*, 7–10.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, *130*(4), 2139–2152. https://doi.org/10.1121/1.3623753

Hazan, V., Uther, M., & Granlund, S. (2015). How does foreigner-directed speech differ from other forms of listener-directed clear speaking styles? *ICPhS 2015*.

Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, *120*(8), 1887–1897. https://doi.org/10.1016/j.lingua.2010.02.002

Henzl, V. M. (1973). Linguistic Register of Foreign Language Instruction. *Language Learning*, *23*(2), Article 2. https://doi.org/10.1111/j.1467-1770.1973.tb00656.x

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Ito, T. A., Thompson, E., & Cacioppo, J. T. (2004). Tracking the Timecourse of Social Perception: The Effects of Racial Cues on Event-Related Brain Potentials. *Personality and Social Psychology Bulletin*, *30*(10), 1267–1280. https://doi.org/10.1177/0146167204264335

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., & Evans, B. (2008). Category and Perceptual Interference in Second-Language Phoneme Learning: An Examination of English /w/-/v / Learning by Sinhala. *Journal of Experimental Psychology. Human Perception and Performance*, *34*, 1305–1316. https://doi.org/10.1037/0096-1523.34.5.1305

Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of Child Language*, *45*(5), 1035–1053. https://doi.org/10.1017/S0305000917000629

Kalashnikova, M., Carignan, C., & Burnham, D. (2017). The origins of babytalk: Smiling, teaching or social convergence? *Royal Society Open Science*, *4*(8), 170306. https://doi.org/10.1098/rsos.170306

Kalashnikova, M., & Carreiras, M. (2021). Input quality and speech perception development in bilingual infants' first year of life. *Child Development*. https://doi.org/10.1111/cdev.13686

Kalashnikova, M., Goswami, U., & Burnham, D. (2018). Mothers speak differently to infants at-risk for dyslexia. *Developmental Science*, *21*(1). https://doi.org/10.1111/desc.12487

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-32150-6

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*, *94*(4), 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Kangatharan, D. J., Uther, P. M., & Gobet, P. F. (2023). The effect of clear speech to foreign-sounding interlocutors on native listeners' perception of intelligibility. *Speech Communication*. https://doi.org/10.1016/j.specom.2023.04.005

Kangatharan, J. (2015). *The role of vowel hyperarticulation in clear speech to foreigners and infants* [Brunel University London]. https://bura.brunel.ac.uk/bitstream/2438/10519/1/FulltextThesis.pdf

Kangatharan, J., Uther, M., Kuhn, L., & Gobet, F. (2012). A-M I S-P-EA-K-I-NG C- L- E- AR-L- Y E-N-OU-GH?: An investigation of the possible role of vowel hyperarticulation in speech communication. In S. F. d’Acoustique (Ed.), *Acoustics 2012*. https://hal.archives-ouvertes.fr/hal-00811042

Kaufman, A. S., & Kaufman, N. L. (2014). Kaufman Brief Intelligence Test, Second Edition. In *Encyclopedia of Special Education*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118660584.ese1325

Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biology*, *16*(3), e2004473. https://doi.org/10.1371/journal.pbio.2004473

Kemper, S., Vandepute, D., Rice, K., Cheung, H., & Gubarchuk, J. (1995). Speech Adjustments to Aging During a Referential Communication Task. *Journal of Language and Social Psychology*, *14*(1–2), 40–59. https://doi.org/10.1177/0261927X95141003

Kendi, A. A., & Khattab, G. (2019). Acoustic Properties of Foreigner Directed Speech. *International Congress of Phonetic Sciences 2019*, 5.

Kirchner, R. M. (1997). An effort-based approach to consonant lenition. *Doctoral Dissertation, University of California*.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother’s speech: Adjustments for age and sex in the first year. *Infancy*, *4*(1), 85–110. https://doi.org/10.1207/S15327078IN0401_5

Klimova, B. (2014). Constraints and Difficulties in the Process of Writing Acquisition. *Procedia - Social and Behavioral Sciences*, *122*. https://doi.org/10.1016/j.sbspro.2014.01.1367

Klimovich-Gray, A., Di Liberto, G., Amoruso, L., Barrena, A., Agirre, E., & Molinaro, N. (2023). Increased top-down semantic processing in natural speech linked to better reading in dyslexia. *NeuroImage*, *273*, 120072. https://doi.org/10.1016/j.neuroimage.2023.120072

Knoll, M. A., & Costall, A. (2015). Characterising F(0) contour shape in infant- and foreigner-directed speech. *Speech Communication*, *66*, 231–243. https://doi.org/10.1016/j.specom.2014.10.007

Knoll, M. A., Johnstone, M., & Blakely, C. (2015). Can you hear me? Acoustic modifications in speech directed to foreigners and hearing-impaired people. *INTERSPEECH 2015*, 2987–2990.

Knoll, M. A., & Scharrer, L. (2007). Acoustic and Affective Comparisons of Natural and Imaginary Infant-, Foreigner- and Adult-Directed Speech. *INTERSPEECH 2007*, 4.

Knoll, M. A., Scharrer, L., & Costall, A. (2009). Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant- and foreigner-directed speech. *Speech Communication*, *51*(3), 296–305. https://doi.org/10.1016/j.specom.2008.10.001

Knoll, M. A., Scharrer, L., & Costall, A. (2011). “Look at the shark”: Evaluation of student- and actress-produced standardised sentences of infant- and foreigner-directed speech. *Speech Communication*, *53*(1), 12–22. https://doi.org/10.1016/j.specom.2010.08.004

Knoll, M. A., Uther, M., & Costall, A. (2009). Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners. *Speech Communication*, *51*(3), 210–216. https://doi.org/10.1016/j.specom.2008.08.001

Knoll, M. A., Uther, M., & Costall, A. (2011). Using the Internet for speech research: An evaluative study examining affect in speech. *Behaviour & Information Technology*, *30*(6), 845–851. https://doi.org/10.1080/0144929X.2011.577192

Knoll, M. A., Walsh, S. A., MacLeod, N., O’Neill, M., & Uther, M. (2007). *Good performers know their audience! Identification and characterisation of pitch contours in infant and foreigner-directed speech.: Vol. Automated taxon recognition in systematics: Theory, approaches and applications,* (CRC Press/The Systematics Association, Boca Raton). MacLeod, N. (Ed.).

Knoll, M., Uther, M., MacLeod, N., O’Neill, M., & Walsh, S. (2006). Emotional, linguistic or just cute? The function of pitch contours in infant- and foreigner-directed speech. *Speech Prosody 2006*, 5.

Kösem, A., Basirat, A., Azizi, L., & van Wassenhove, V. (2016). High-frequency neural activity predicts word parsing in ambiguous speech streams. *Journal of Neurophysiology*, *116*(6), 2497–2512. https://doi.org/10.1152/jn.00074.2016

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural Entrainment Determines the Words We Hear. *Current Biology*, *28*(18), 2867-2875.e3. https://doi.org/10.1016/j.cub.2018.07.023

Kramsch, C. (2007). Re-reading Robert Lado, 1957, Linguistics across Cultures. Applied linguistics for language teachers. *International Journal of Applied Linguistics*, *17*(2), 241–247. https://doi.org/10.1111/j.1473-4192.2007.00149.x

Krashen, S. (1989). We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis. *The Modern Language Journal*, *73*(4), 440–464. https://doi.org/10.1111/j.1540-4781.1989.tb05325.x

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362–378. https://doi.org/10.1121/1.1635842

Krishnamurthy, N., & Hansen, J. H. L. (2009). Babble Noise: Modeling, Analysis, and Applications. *IEEE Transactions on Audio, Speech, and Language Processing*, *17*(7), 1394–1407. https://doi.org/10.1109/TASL.2009.2015084

Kuhl, P. K. (1997). Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science*, *277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*(1), 54–72. https://doi.org/10.3758/s13423-012-0341-8

Kühnert, B., & Kocjančič Antolík, T. (2017). Patterns of articulation rate variation in English/French tandem interactions. In J. Volin & R. Skarnitzl (Eds.), *Pronunciation of English by Speakers of Other Languages*. Cambridge Scholars Publishing. https://halshs.archives-ouvertes.fr/halshs-01505928

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Laan, G. P. M. (1992). Perceptual differences between spontaneous and read aloud speech. *Proceedings of the Institute of Phonetic Sciences Amsterdam*, 16.

Ladefoged, P. (2006). *A Course in Phonetics*. Thomson, Wadsworth. https://books.google.es/books/about/A_Course_in_Phonetics.html?id=jRmoQgAACAAJ&redir_esc=y

Ladefoged, P., & Maddieson, I. (1990). Vowels of the world's languages. *Journal of Phonetics*, *18*(2), 93–122. https://doi.org/10.1016/S0095-4470(19)30396-1

Lalor, E. C., Power, A. J., Reilly, R. B., & Foxe, J. J. (2009). Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. *Journal of Neurophysiology*, *102*(1), 349–359. https://doi.org/10.1152/jn.90896.2008

Lam, C., & Kitamura, C. (2012). Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation: Hearing loss and vowel hyperarticulation. *Developmental Science*, *15*(2), 212–221. https://doi.org/10.1111/j.1467-7687.2011.01118.x

Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of Clear Speech: Effect of Instruction. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1807–1821. https://doi.org/10.1044/1092-4388(2012/11-0154)

Lee, J., Jang, J., & Plonsky, L. (2015). The Effectiveness of Second Language Pronunciation Instruction: A Meta-Analysis. *Applied Linguistics*, *36*(3), 345–366. https://doi.org/10.1093/applin/amu040

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package ‘emmeans’. *Package ‘emmeans’*. https://github.com/rvlenth/emmeans

Levelt, W. J. M. (1989). *Speaking: From intention to articulation* (pp. xiv, 566). The MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461. https://doi.org/10.1037/h0020279

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Liscombe, J., Venditti, J. J., & Hirschberg, J. (2003). Classifying subject ratings of emotional speech using acoustic features. *INTERSPEECH 2003*.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers’ speech clarity and infants’ speech discrimination skills. *Developmental Science*, *6*(3), F1–F10. https://doi.org/10.1111/1467-7687.00275

Liu, H.-M., Tsao, Feng-Ming, & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, *43*(4), 912. https://doi.org/10.1037/0012-1649.43.4.912

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2009). Age-related Changes in Acoustic Modifications of Mandarin Maternal Speech to Preverbal Infants and Five-Year-Old Children: A Longitudinal Study. *Journal of Child Language*, *36*(4), 909. https://doi.org/10.1017/S030500090800929X

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America*, *49:606-08*.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*(3), 550–557. https://doi.org/10.3758/BF03196312

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. https://doi.org/10.1121/1.1894649

Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales Des Maladies de L'Oreille et Du Larynx.*, *37*, 101–119.

Long, M. H. (1981). Questions in foreigner talk disclosure. *Language Learning*, *31*(1), 135–157. https://doi.org/10.1111/j.1467-1770.1981.tb01376.x

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, *4*(2), Article 2. https://doi.org/10.1093/applin/4.2.126

Lorge, I., & Katsos, N. (2019). Listener-adapted speech: Bilinguals adapt in a more sensitive way. *Linguistic Approaches to Bilingualism*, *9*(3), 376–397. https://doi.org/10.1075/lab.16054.lor

Lovcevic, I., Kalashnikova, M., & Burnham, D. (2020). Acoustic features of infant-directed speech to infants with hearing loss. *The Journal of the Acoustical Society of America*, *148*(6), 3399–3416. https://doi.org/10.1121/10.0002641

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Lynch, A. J. (1988). Speaking up or talking down: Foreign learners’ reactions to teacher talk. *ELT Journal*, *42*(2), 109–116. https://doi.org/10.1093/elt/42.2.109

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word Learning in Infant- and Adult-Directed Speech. *Language Learning and Development*, *7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, *125*(6), 3962–3973. https://doi.org/10.1121/1.2990715

Margić, B. (2017). Communication courtesy or condescension? Linguistic accommodation of native to non-native speakers of English. *Journal of English as a Lingua Franca*, *6*. https://doi.org/10.1515/jelf-2017-0006

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers Speak Less Clearly to Infants Than to Adults: A Comprehensive Test of the Hyperarticulation Hypothesis. *Psychological Science*, *26*(3), 341–347. https://doi.org/10.1177/0956797614562453

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition*, *129*(2). https://doi.org/10.1016/j.cognition.2013.07.015

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2022). On the relationship between perception and production of L2 sounds: Evidence from Anglophones’ processing of the French /u/–/y/ contrast. *Second Language Research*, *38*(3), 581–605. https://doi.org/10.1177/0267658320988061

Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, *251*, 118991. https://doi.org/10.1016/j.neuroimage.2022.118991

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609–2621. https://doi.org/10.1111/ejn.13748

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*(1), 271–280. https://doi.org/10.1016/j.cognition.2008.02.003

Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition*, *166*, 84–93. https://doi.org/10.1016/j.cognition.2017.05.003

Molinaro, N., & Lizarazu, M. (2018). Delta(but not theta)-band cortical entrainment involves speech-specific processing. *The European Journal of Neuroscience*, *48*(7), 2642–2650. https://doi.org/10.1111/ejn.13811

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different conditions: The use of non-lexical materials and masking noise. *Phonetica*, *79*(1), 1–43. https://doi.org/10.1515/phon-2022-2018

Müller, M. (Ed.). (2007). Dynamic Time Warping. In *Information Retrieval for Music and Motion* (pp. 69–84). Springer. https://doi.org/10.1007/978-3-540-74048-3_4

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

Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the*

Acoustical Society of America, *139*(3), 1272–1281. https://doi.org/10.1121/1.4944634

Nekvapil, J., & Sherman, T. (2015). An introduction: Language Management Theory in Language Policy and Planning. *International Journal of the Sociology of Language*, *2015*(232), 1–12. https://doi.org/10.1515/ijsl-2014-0039

Nelson, D. K. (1992). *The Foreigner Talk of a Family Physician: An Observational Study*. https://eric.ed.gov/?id=ED353826

Nyikos, M., & Oxford, R. (1993). A Factor Analytic Study of Language-Learning Strategy Use: Interpretations from Information-Processing Theory and Social Psychology. *The Modern Language Journal*, *77*(1), 11–22. https://doi.org/10.2307/329553

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

O'Brien, M. G. (2021). Ease and Difficulty in L2 Pronunciation Teaching: A Mini-Review. *Frontiers in Communication*, *5*. https://www.frontiersin.org/articles/10.3389/fcomm.2020.626985

Oever, S. ten, & Sack, A. T. (2015). Oscillatory phase shapes syllable perception. *Proceedings of the National Academy of Sciences*, *112*(52), 15833–15837. https://doi.org/10.1073/pnas.1517519112

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*(1), Article 1. https://doi.org/10.1159/000261706

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Panneton, R., Kitamura, C., Mattock, K., & Burnham, D. (2006). Slow Speech Enhances Younger But Not Older Infants' Perception of Vocal Emotion. *Research in Human Development*, *3*(1), 7–19. https://doi.org/10.1207/s15427617rhd0301_2

Papoušek, M., & Hwang, S.-F. C. (1991). Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics*, *12*(4), Article 4. https://doi.org/10.1017/S0142716400005889

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex (New York, NY)*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peña, M., & Melloni, L. (2012). Brain Oscillations during Spoken Sentence Processing. *Journal of Cognitive Neuroscience*, *24*(5), 1149–1164. https://doi.org/10.1162/jocn_a_00144

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Pérez, A., Carreiras, M., Gillon Dowens, M., & Duñabeitia, J. A. (2015). Differential oscillatory encoding of foreign speech. *Brain and Language*, *147*, 51–57. https://doi.org/10.1016/j.bandl.2015.05.008

Pérez-Navarro, J., Lallier, M., Clark, C., Flanagan, S., & Goswami, U. (2022). Local Temporal Regularities in Child-Directed Speech in Spanish. *Journal of Speech, Language, and Hearing Research*, *65*(10), 3776–3788. https://doi.org/10.1044/2022_JSLHR-22-00111

Peter, V., Kalashnikova, M., Santos, A., & Burnham, D. (2016). Mature neural responses to Infant-Directed Speech but not Adult-Directed Speech in Pre-Verbal Infants. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep34273

Piazza, E. A., Iordan, M. C., & Lew-Williams, C. (2017). Mothers Consistently Alter Their Unique Vocal Fingerprints When Communicating with Infants. *Current Biology: CB*, *27*(20), 3162-3167.e3. https://doi.org/10.1016/j.cub.2017.08.074

Piazza, G., Kalashnikova, M., & Martin, C. D. (under review). Speakers' communicative intentions lead to acoustic adjustments in native and non-native directed speech. https://doi.org/10.31234/osf.io/kz72c

Piazza, G., Kartushina, N., Flege, J. E., & Martin, C. D. (2022). Comparison of Acoustic Features in Speech Production Studies Run Online and in the Lab*. 27*, 232. *63rd Psychonomic Society Annual Meeting,* Boston, USA. https://cdn.ymaws.com/www.psychonomic.org/resource/resmgr/annual_meeting/2022_meeting/ps22_abstract__book_10.27.22.pdf

Piazza, G., Martin, C. D., & Kalashnikova, M. (2022). The Acoustic Features and Didactic Function of Foreigner-Directed Speech: A Scoping Review. *Journal of Speech, Language, and Hearing Research: JSLHR*, 1–23. https://doi.org/10.1044/2022_JSLHR-21-00609

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*(4), 434–446. https://doi.org/10.1044/jshr.2904.434

Pickering, M., & Garrod, S. (2006). Alignment as the Basis for Successful Communication. *Reseach On Language And Computation*, *4*, 203–228. https://doi.org/10.1007/s11168-006-9004-0

Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, *27*(2), 212–225. https://doi.org/10.1017/S0140525X04450055

Pickering, M. J., & Garrod, S. (2014). Interactive Alignment and Language Use. In T. M. Holtgraves (Ed.), *The Oxford Handbook of Language and Social Psychology* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199838639.013.026

Porcaro, C., Medaglia, M. T., & Krott, A. (2015). Removing speech artifacts from electroencephalographic recordings during overt picture naming. *NeuroImage*, *105*, 171–180. https://doi.org/10.1016/j.neuroimage.2014.10.049

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, *40*(3), 491–508. https://doi.org/10.1016/j.wocn.2012.01.001

Ramamurti, R. (1980). Strategies Involved in Talking to a Foreigner. *PENN Review of Linguistics*, *4*, 84–93.

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 75–86. https://doi.org/10.1037/a0027979

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural Entrainment to Speech Modulates Speech Intelligibility. *Current Biology*, *28*(2), 161-169.e5. https://doi.org/10.1016/j.cub.2017.11.033

Rodriguez-Cuadrado, S., Baus, C., & Costa, A. (2018). Foreign talk through word reduction in native/non-native spoken interaction. *Bilingualism: Language and Cognition*, *21*(2), 419–426. https://doi.org/doi:10.1017/S1366728917000402

Rothermich, K., Baker, R., Falkins, S., Kum, J., Capps, M., & Bobb, S. C. (2021). *Everyday prevalence of foreigner-directed speech*. SocArXiv. https://doi.org/10.31235/osf.io/jwefk

Rothermich, K., Harris, H. L., Sewell, K., & Bobb, S. C. (2019). Listener impressions of foreigner-directed speech: A systematic review. *Speech Communication*, *112*, 22–29. https://doi.org/10.1016/j.specom.2019.07.002

Rubin, D. L. (1992). Nonlanguage Factors Affecting Undergraduates' Judgments of Nonnative English-Speaking Teaching Assistants. *Research in Higher Education*, *33*(4), 511–531.

Ryan, E. B., Giles, H., Bartolucci, G., & Henwood, K. (1986). Psycholinguistic and social psychological components of communication by and with the elderly. *Language & Communication*, *6*(1–2), 1–24. https://doi.org/10.1016/0271-5309(86)90002-9

Ryan, E. B., Hamilton, J. M., & See, S. K. (1994). Patronizing the old: How do younger and older adults respond to baby talk in the nursing home? *International Journal of Aging & Human Development*, *39*(1), 21–32. https://doi.org/10.2190/M52C-M2D2-R6C2-3PBM

Saadah, E. (2011). The production of Arabic vowels by English L2 learners and heritage speakers of Arabic. *Doctoral Dissertation, University of Illinois at Urbana-Champaign*.

Sankowska, J., García Lecumberri, M. L., & Cooke, M. (2011a). Interaction of intrinsic vowel and consonant durational correlates with foreigner directed speech. *Poznań Studies in Contemporary Linguistics*, *47*. https://doi.org/10.2478/psicl-2011-0009

Sankowska, J., García Lecumberri, M. L., & Cooke, M. (2011b). Interaction of intrinsic vowel and consonant durational correlates with foreigner directed speech. *Poznań Studies in Contemporary Linguistics*, *47*. https://doi.org/10.2478/psicl-2011-0009

Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., & Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. *The Journal of the Acoustical Society of America*, *121*(5), 3044–3044. https://doi.org/10.1121/1.4781735

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Scherer, K. R. (1971). Randomized splicing: A note on a simple technique for masking speech content. *Journal of Experimental Research in Personality*, *5*(2), 155–159.

Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research*, *1*(3), 269–285. https://doi.org/10.1007/BF01074443

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*(3), 249–263. https://doi.org/10.1016/j.wocn.2013.03.007

Skinner, B. F. (1950). Are Theories of Learning Necessary? *Psychological Review*, 57, 193-216.

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *118*(3), 1677–1688. https://doi.org/10.1121/1.2000788

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass*, *3*(1), 236–264. https://doi.org/10.1111/j.1749-818X.2008.00112.x

Smith, C. (2007). Prosodic accomodation by French speakers to non-native interlocutor. *ICPhS XVI*.

Smith, N. A., & Trainor, L. J. (2008). Infant-Directed Speech Is Modulated by Infant Feedback. *Infancy*, *13*(4), 410–420. https://doi.org/10.1080/15250000802188719

Snow, C. E., Eeden, R., & Muysken, P. (1981). The Interactional Origins of Foreigner Talk: Municipal Employees and Foreign Workers. *International Journal of the Sociology of Language*, *1981*(28). https://doi.org/10.1515/ijsl.1981.28.81

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, *128*(1), 389–400. https://doi.org/10.1121/1.3419786

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Starkweather, J. A. (1967). Vocal behavior as an information channel of speaker status. *Academic Press, New York*, 253–262.

Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, *50*(3), 163–178. https://doi.org/10.1016/j.specom.2007.07.005

Stewart, Bouchard-Ryan, M. A., Ellen. (1982). Attitudes Toward Younger and Older Adult Speakers: Effects of Varying Speech Rates. *Journal of Language and Social Psychology*, *1*(2), 91–109. https://doi.org/10.1177/0261927X8200100201

Suffill, E., Kutasi, T., Pickering, M., & Branigan, H. (2021). Lexical alignment is affected by addressee but not speaker nativeness. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728921000092

Tarone, E. (1980). Communication strategies, foreign talk, and repair in interlanguage. *Language Learning*, *30*(2), 417–428. https://doi.org/10.1111/j.1467-1770.1980.tb00326.x

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Tjaden Kris, Sussman Joan E., & Wilding Gregory E. (2014). Impact of Clear, Loud, and Slow Speech on Scaled Intelligibility and Speech Severity in Parkinson's Disease and Multiple Sclerosis. *Journal of Speech, Language, and Hearing Research*, *57*(3), 779–792. https://doi.org/10.1044/2014_JSLHR-S-12-0372

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335–340. https://doi.org/10.3758/BF03196290

Tremblay, R. E., Schaal, B., Boulerice, B., Arseneault, L., Soussignan, R. G., Paquette, D., & Laurent, D. (1998). Testosterone, Physical Aggression, Dominance, and Physical Development in Early Adolescence. *International Journal of Behavioral Development*, *22*(4), 753–777. https://doi.org/10.1080/016502598384153

Tursunov, A., Kwon, S., & Pang, H.-S. (2019). Discriminating Emotions in the Valence Dimension from Speech Using Timbre Features. *Applied Sciences*, *9*(12), Article 12. https://doi.org/10.3390/app9122470

Tweissi, A. I. (1990). Foreigner Talk in Arabic: Evidence for the universality of language simplification. *Annual Symposium on Arabic Linguistics*. Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics. Volume II, Salt Lake City, Utah. https://doi.org/10.1075/cilt.72.16twe

Uther, M., Giannakopoulou, A., & Iverson, P. (2012). Hyperarticulation of vowels enhances phonetic change responses in both native and non-native speakers of English: Evidence from an auditory event-related potential study. *Brain Research*, *1470*, 52–58. https://doi.org/10.1016/j.brainres.2012.06.041

Uther, M., Knoll, M. A., & Burnham, D. (2007a). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*(1), Article 1. https://doi.org/10.1016/j.specom.2006.10.003

Uther, M., Knoll, M. A., & Burnham, D. (2007b). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*(1), 2–7. https://doi.org/10.1016/j.specom.2006.10.003

Valdman, A. (1981). Sociolinguistic Aspects of Foreigner Talk. *International Journal of the Sociology of Language*, *1981*(28), 41–52. https://doi.org/10.1515/ijsl.1981.28.41

Van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01000

Verschueren, E., Gillis, M., Decruy, L., Vanthornhout, J., & Francart, T. (2022). Speech Understanding Oppositely Affects Acoustic and Linguistic Neural Tracking in a Speech Rate Manipulation Paradigm. *Journal of Neuroscience*, *42*(39), 7442–7453. https://doi.org/10.1523/JNEUROSCI.0259-22.2022

Vogt, A., Hauber, R., Kuhlen, A. K., & Rahman, R. A. (2021). Internet-based language production research with overt articulation: Proof of concept, challenges, and practical advice. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01686-3

Wang, W. (1999). Age and Second Language Acquisition in Adulthood: The Learning Experiences and Perceptions of Women Immigrants. *TESL Canada Journal*, 01–19. https://doi.org/10.18806/tesl.v16i2.715

Warren-Leubecker, A., & Bohannon, J. N. (1982). The effects of expectation and feedback on speech to foreigners. *Journal of Psycholinguistic Research*, *11*(3), 207–215. https://doi.org/10.1007/BF01067564

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Wooldridge, B. (2001). ‘Foreigner Talk’: An Important Element in Cross-Cultural Management Education and Training. *International Review of Administrative Sciences*, *67*(4), 621–634. https://doi.org/10.1177/0020852301674002

Xu Rattanasone, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2013). Vowel Hyperarticulation in Parrot-, Dog- and Infant-Directed Speech. *Anthrozoös*, *26*(3), 373–380. https://doi.org/10.2752/175303713X13697429463592

Xu Rattanasone, N., Burnham, D., & Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *Journal of Phonetics*, *41*(5), 332–343. https://doi.org/10.1016/j.wocn.2013.06.001

Zangl, R., & Mills, D. L. (2007). Increased Brain Activity to Infant-Directed Speech in 6-and 13-Month-Old Infants. *Infancy*, *11*(1), 31–62. https://doi.org/10.1207/s15327078in1101_2

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832

Zhang, Y. B., & Giles, H. (2017). Communication Accommodation Theory. In *The International Encyclopedia of Intercultural Communication* (pp. 1–14). American Cancer Society. https://doi.org/10.1002/9781118783665.ieicc0156

Zuengler, J. (1991). Accommodation in native-nonnative interactions: Going beyond the “what” to the “why” in second-language research. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 223–244). Cambridge University Press. https://doi.org/10.1017/CBO9780511663673.007

Appendixes

Appendix 1.1

Table 1. Summary of the design and findings of the published studies that have assessed vowel hyperarticulation, speech rate, or pitch correlates of FDS.

Paper	***Speaker sex***	***Addressee's sex and other characteristics***	***Addressee's physical appearance***	***Addressee's language proficiency***	***Speech type /task***	***Language of interaction***	***Vowel hyperart. in FDS***	***Speech rate***	***Pitch measure***	***FDS pitch measure***
Warren-Leubecker and Bohannon (1982)	*Not reported*	Female	Native	High + simulated low	Dialogue with fixed questions	English	*Not reported*	FDS < NDS	*Not reported*	NA
Nelson (1992)	Male (1)	*Not reported*	*Not reported*	*Not reported*	Conversation (on the telephone)	English	*Not reported*	FDS < NDS	Pitch height	FDS = NDS[1]
Papaousek & Hwang (1991)	Female	Imaginary	NA	NA	Standardized sentences	Mandarin Chinese	*Not reported*	*Not reported*	Pitch range and contour	Pitch range: FDS > NDS Pitch contour: FDS = NDS
Knoll et al. (2004)	*Not reported*	*Not reported*	Foreigner	*Not reported*	Simulated free	English	**Yes** IDS = FDS > NDS	*Not reported*	Mean pitch	IDS > FDS = NDS
Knoll et al (2006)	Female	Imaginary	NA	NA	Simulated free	English	*Not reported*	*Not reported*	Pitch contour	FDS = NDS
Smith (2007)	Female	Male (FDS), Female (NDS)	*Not reported*	*Not reported*	Map description (on the telephone)	French	*Not reported*	FDS = NDS	Pitch range	FDS > NDS
Biersack et al. (2005)	Female	Imaginary	NA	NA	Map description	English	*Not reported*	FDS < IDS < NDS	Pitch max and pitch range	IDS > FDS = NDS
Uther et al (2007)	Female	Both	Foreigner	*Not reported*	Free	English	**Yes** IDS = FDS > NDS	*Not reported*	Mean pitch	IDS > FDS = NDS
Knoll & Scharrer (2007)	Female	Imaginary (Both)	Imaginary (Foreigner)	NA	Simulated free	English	**No** FDS = NDS > IDS	*Not reported*	Mean pitch	FDS = IDS > NDS[2]
Knoll et al. (2009a)	Female	Imaginary (Female)	Imaginary (Foreigner)	NA	Simulated free	English	**No**[3] Students: FDS = NDS < IDS Actresses: NDS = FDS = IDS	*Not reported*	Mean pitch	Students: IDS = FDS; FDS = NDS Actresses: IDS > FDS = NDS
Knoll et al. (2011a)	Female	Imaginary (Female)	Imaginary (Foreigner?)	NA	Simulated free and standardized sentences	English	**Yes** IDS = FDS > NDS (no students)	*Not reported*	Mean pitch	IDS > FDS = NDS
Kangatharan et al. (2012)	*Not reported*	*Not reported*	Native and Foreigner	Not reported	Diapix	English	**No** FDS = NDS	*Not reported*	*Not reported*	NA
Kangatharan et al. (2015)	*Not reported*	*Not reported*	Native and Foreigner		Diapix	English	**Yes** FDS > NDS	*Not reported*	*Not reported*	NA
Hazan et al. (2015)	Both	Probably both	Foreigner	Basic user	Diapix	English	**Yes** FDS > NDS	FDS < NDS	Median pitch and pitch range	Median pitch: FDS > NDS; Pitch range: FDS < NDS[4]
Knoll & Costall (2015)	Female	Imaginary (Female) + real women	Imaginary (Foreigner) + real foreigners	*Not reported*	Simulated free and free speech	English	*Not reported*	*Not reported*	Pitch contour	Level shape contour in FDS
Knoll et al. (2015)	Female	Female (1) and Male (2)	*Not reported*	*Not reported*	Diapix	English	**No** HIDS[5] > FDS = NDS	*Not reported*	Mean pitch and pitch range	Mean pitch: FDS = HIDS > NDS; Pitch range: FDS = HIDS = NDS
Kühnert & Antolík (2017)	Female & both	Both	Native	6.4 & 9.2 years of study	Tandem	French and English	*Not reported*	French: FDS < NDS; English FDS = NDS	*Not reported*	NA
Rodriguez-Cuadrado et al. (2018)	Both	Female (NDS) & Male (FDS)	Native	*Not reported*	Map description	Spanish	*Not reported*	FDS < NDS	*Not reported*	NA
Kendi & Khattab (2019)	Female	Female (domestic helper)	Foreigner	0.7-21 years	Diapix	Arabic	**Yes** FDS > NDS	*Not reported*	Pitch midpoint	FDS > NDS
Lorge & Katsos (2019)	Both	Female	Not reported	Not reported	Map description	English	**Yes** but Biling.: FDS > NDS Mono.: FDS < NDS	*Not reported*	Not reported	NA
Bobb et al. (2019)	Both	Imaginary	Imaginary	NA	Read sentences	English	**Yes** FDS > IDS > NDS	FDS < NDS	Median pitch Pitch range	Median pitch: IDS > FDS > NDS; Pitch range: IDS = FDS = NDS

[1] Described as no “notable differences” between FDS and NDS; [2] Pitch height and hyperarticulation are usually associated in IDS. In this study IDS exhibited a smaller vocal space, but higher mean pitch than FDS or NDS. FDS mean pitch was (almost significantly; p = 0.054) higher than NDS; [3] Knoll et al. (2009a) failed to find greater hyperarticulation in FDS than NDS, but no differences were observed in the actresses’ performance with IDS (which was more hyperarticulated than NDS); [4] These results are described as “minimal differences”; [5] Hearing-Impaired Person-Directed Speech (HIDS).

Appendix 2.1

Mixed-effect model formulas

1. Vowel hyperarticulation

```
model <- lmer(VOWEL_SPACE_norm~  REGISTER*CONDITION + SESSION
+(1+SESSION|SUBJ), data=data, REML=F)
```

2. Speech rate

```
model <- lmer(SR_norm ~  REGISTER*CONDITION +
TRIAL_NUMBER+(1|CONDITION:SUBJ)+ (1|SESSION), data=data, REML=F)
```

3. *Vowel intensity*

```
model <- lmer(INTENSITY_norm ~  REGISTER*CONDITION +
(1+SESSION|VOWEL:SUBJ) + (1|WORD:SESSION), data=data, REML=F)
```

4. *Pitch height*

```
model <- lmer(F0_norm ~ REGISTER*CONDITION + (1 + VOWEL |SESSION:SUBJ) +
(1| VOWEL:WORD), data = data, REML = F)
```

5. Vowel duration

```
model <- lmer(DURATION_norm ~  REGISTER*CONDITION +
SEX+(1|SESSION:SUBJ) + (1|VOWEL:WORD), data=data, REML=F)
```

SEX = participants’ reported sex

Appendix 2.2

List of experimental stimuli

Initial target vowels

HASME
HARJO
HAFLA
HALCO
HIRTA

ANFES
ARBOS
ALPRO
AZCRA
IRGOS

ÁFARO
ÁTAMA
ÁREMO
ÁNUTA
ÍJERA

HIRSA	ISBOS	ÍLARGO
HILMA	ISTRA	ÍBAD
HIRRO	IRFED	ÍFANO
HUCRO	USCLA	ÚDELO
HULCRA	ULSO	ÚCED
HUGA	USMO	ÚPODA
HUJO	UFRAS	ÚSPERO

Appendix 3.1

FDS of English vowel contrasts: five English speakers.

Previous studies on FDS suggest that this register is the result of speech adaptation to non-native listeners' learning needs. Yet, empirical evidence is limited to hyperarticulation of the /a/,/i/,/u/ triangle, which also limits generalisation of FDS impact on learning, especially for a language with rich vowel inventory like English (Piazza et al., 2022). Thus, prior the actual experiment, we aimed to assess the Euclidean distance between vowels of the /i - ɪ/ and /ʌ - æ/ contrasts and the vowel duration difference for the /i - ɪ/ contrast as index of FDS speech adaptation. For this purpose, we recruited five native speakers of English (British accent), who were (or had been) teachers of English with Spanish speaking students (M_{age} = 32.3 y.o.). The speakers carried out an online experiment, where they were asked to speak as if they were addressing a native listener of English or a Spanish native L2 learner of English. Addressing these two fake listeners elicited two registers, respectively NDS and FDS. Before the start of each register-block, speakers listened to a short introduction about the listener. Then, they were instructed to pretend they were speaking to that person. The speakers read English words appearing on the screen and were asked to say "this is a [word on the screen]". After each production, they pressed 'Enter' to go to the next trial.

We selected 48 monosyllabic and disyllabic real English words containing the 4 vowels of interest (/i/, /ɪ/, /ʌ, /æ). We selected minimal pairs differing by only the target vowels. We had 8 word-minimal pairs containing the /i - ɪ/ contrast (total 16 words like *sheep-ship*) and 8 word-minimal pairs containing the /ʌ - æ/ contrast (total 16 words like *cup-cap*). To these, we added 16 words containing the /a/ and /u/ vowels (8 each), which did not form minimal pairs. Those words were considered as fillers and were added in order to avoid speakers being conscious that the two vowel contrasts were of special interest for us. Words were presented in pseudorandomized order, in all caps and in the centre of the screen. Speakers repeated each stimulus 6 times (3 addressing each listener) for a total number of 288 trials.

From the resulting production data, we first examined the vocalic space of the /a/, /i/, and /u/ triangle of the two registers. Also, we measured the duration of sentences ("this is a…") speakers produced. These two measures allowed us to assess whether speakers produced FDS vocalic triangle and speech rate in line with previous literature (e.g., Uther et al., 2007; Piazza et al., in preparation). Next, we focused on the minimal pairs to measure the Euclidean distance of the /i - ɪ/ and /ʌ - æ/ contrasts and the duration difference of the /i - ɪ/ contrast (duration of /i/ - duration of /ɪ/). Euclidean distance was calculated by projecting on a cartesian plane the first (F1) and second (F2) vowel formants of the speakers' production. If speakers adapt to the listeners needs and produce FDS to support L2 learning, we expected the speakers to produce 1) wider vocalic triangle 2) longer sentences (lower speech rate) 2) greater /i/ minus/ɪ/ duration difference 3) greater Euclidean distance between /ʌ - æ/ and between /i - ɪ/ in FDS as compared to NDS.

In line with the literature, non-parametric Wilcoxon tests revealed that FDS was produced with wider vocalic triangle (+30.6%)[5] and with longer sentences (hence lower speech rate) as compared to NDS (W = 856426, p < .0001). Also, non-parametric Wilcoxon tests revealed that speakers produced greater /ʌ - æ/ Euclidean distance (W = 7057, p = .016), and /i - ɪ/ duration difference (W = 8213, p = .001) in FDS than NDS. Unexpectedly, the /i - ɪ/ Euclidean distance was greater in NDS than FDS (W = 4235, *p* = .038). These results suggest that speakers emphasise duration difference in between /i - ɪ/, and Euclidean distance in between

---

[5] Statistical analysis for the vocalic space could not be run due to the low number of observations: we extracted only one vocalic space value per each subject and condition.

/ʌ - æ/, when producing FDS. On these results we built the novel word stimuli of the present study. For recording the stimuli, we selected the speaker that produced FDS with more hyperarticulated vowels, more duration difference between /i - ɪ/, greater /ʌ - æ/ Euclidean distance and with smaller /i - ɪ/ Euclidean distance than NDS.

List of experimental stimuli

BEEFUL	BIFUL	CABBON	CUBBON
DEEST	DIST	DAGMET	DUGMET
GHEEDEN	GHIDEN	GACK	GUCK
PEEV	PIV	TASS	TUSS

List of fillers

POOTON
SOOTIC
TOOD
PHOON
TARPEL
FARSIK
SARN
PARG

Appendix 3.2

Supplementary images

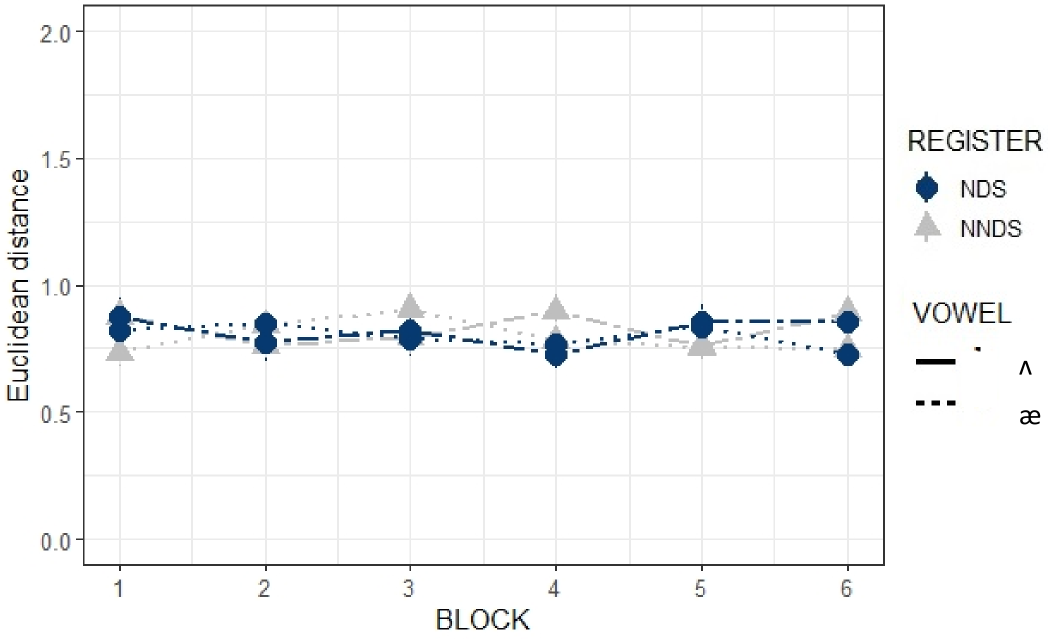


Figure A. *Production task.* Single contrast*:* (normalised) Euclidean distance participants' production – vowel stimuli, across blocks, by *Register* (FDS = Foreigner Directed Speech, NDS = Native Directed Speech) and *Vowel* (ʌ and æ). Bars indicate SEM.

Appendix 3.3

(GCA) Mixed-effect model formulas

1. Recognition accuracy

```
model <- glmer(ACCURACY ~
(ot1+ot2)*REGISTER*CONTRAST+(ot1+ot2|IQ:SUBJ)+(ot1|WORD),family = "binomial",
data = data)
```

2. Recognition RT

```
model <- lmer(RT_bc ~ (ot1+ot2)*REGISTER*CONTRAST+(ot1+ot2|WORD:SUBJ),
data =data,REML = F)
```

3. Aline distance

```
model <- glmmTMB(ALINE
~(ot1+ot2)*REGISTER*CONTRAST+(ot1+ot2|SUBJ)+(ot1|WORD),          ziformula=~0,
data=data, REML = F, betabinomial(link = "logit"))
```

4. Production RT

```
model <- lmer(RT_bc ~
(ot1+ot2)*REGISTER*CONTRAST+(1|SUBJ)+(1|PHON:WORD), data = data, REML = F)
```

5. Euclidean distance A

```
model_Goodness_c<- lmer(EUCLIDEAN ~
(ot1+ot2)*REGISTER*VOWEL+(1|PHON:SUBJ)+(1|WORD), data = Goodness, REML = F)
```

```
model_Single_c <- lmer(EUCLIDEAN ~
(ot1+ot2)*REGISTER*VOWEL+(1|PHON:SUBJ)+(1|WORD), data = Single, REML = F)
```

6. Euclidean distance B (within contrast)

```
model_Goodness_c <-  lmer(EUCLIDEAN ~
(ot1+ot2)*REGISTER+(1|PHON:SUBJ)+(1|CONTRAST:WORD), data = Goodness, REML =
F)
```

```
model_Single_c <- lmer(EUCLIDEAN ~
(ot1+ot2)*REGISTER+(1|PHON:SUBJ)+(1|WORD), data = Single, REML = F)
```

7. Logistic regression (sheep-ship/cup-cap)

```
model <- glmer(CHOICE~ EXPOSURE*REGISTER+(1|SUBJ)+(1|CONT), data = data,
family = "binomial")
```

IQ = Participants' non-verbal IQ scores

PHON = Phonological memory scores

WORD = novel word/item.

Appendix 3.4

Raven matrices test and pseudoword repetition: materials and procedure.

Participants' non-verbal IQ was assessed using the Raven matrices from the Kaufman Brief Intelligence Test (KBIT; Kaufman & Kaufman, 2014). Participants completed as many sequences as they could in 6 minutes. Resulting standardised scores (M = 100, Kaufman & Kaufman, 2014) were calculated. Phonological memory skills were assessed via a pseudoword repetition task (Gathercole et al., 1994). Participants listened to 28 Spanish pseudowords, 4-18 phonemes long, in random order and repeated them one by one as fast and accurately as possible. After each repetition, they clicked on the button 'Send your response' to proceed to the next trial. Production of each pseudoword was checked offline for accuracy: 1 point score was assigned to each correctly pronounced target and a 0.25-point penalty per each incorrect or missed phoneme (min score: 0). The final phonological memory score was computed as total accuracy considering the penalties.

Appendix 4.1

Mixed-effect model formulas

EXP.1

1. *Comprehension score*

```
model <- glmer(SCORE ~ REGISTER + (1|SUBJ) + (1|Part:Story), family = "binomial", data = data)
```

2. *Prediction correlations – Envelope TRF*

```
model <- lmer(R ~ REGISTER + (1|SUBJ)+(1|ELECTRODE), data = data, REML = F)
```

3. *Prediction correlations – Semantic mTRF*

```
model <- lmer(R_sem ~ REGISTER + (1| SUBJ)+(1| ELECTRODE), data = data, REML = F)
```

EXP.2

4. *Comprehension score*

```
model <- glmer(SCORE ~ REGISTER + (1| SUBJ) + (1|Part:Story), family = "binomial", data = data)
```

5. Prediction correlations – Envelope TRF

```
model <- lmer(R ~ REGISTER + (1| SUBJ)+(1| ELECTRODE), data = data, REML = F)
```

6. Prediction correlations – Semantic mTRF

```
model <- lmer(R_sem ~ REGISTER + (1| SUBJ), data = data, REML = F)
```

Part = part of story (1 to 5)

Story = Love, Guerrilla, or Hole: the three stories that participants listened to (regardless the speech register).