The Role of Metalinguistic Awareness in Stress "Deafness" in French-L1 Learners of English

Clare Jamieson 15131483

ACLC, University of Amsterdam

Master's Thesis: Linguistics & Communication

Supervisors: Dr. Kimberley Mulder & Dr. Silke Hamann

July 9th, 2025

1. Background

How people perceive speech sounds is shaped by their linguistic background. Polivanov first brought this idea of native-language-specific perception to linguistics in 1931 through his observations of how Japanese and Korean speakers perceived Russian words, and since then, studies have operationalized and supported his arguments, mostly in relation to perception of phonemes (e.g. Dupoux et al., 1999). There is also evidence that perception of suprasegmentals, such as prosodic features, is shaped by the native language. Over the last three decades, a phenomenon called "stress deafness" has been repeatedly found in first-language (L1) speakers of French (Bijeljac-Babic, 2012; Carpenter, 2015; Dupoux et al., 1997; Dupoux et al., 2001; Dupoux et al., 2010; Frost, 2010; Peperkamp et al., 2010). This so-called "deafness" consists in French-L1 speakers having difficulty perceiving lexical stress as well as L1 speakers of other languages like Spanish. This study aims to explore the possibility that such stress deafness can be modified in the short-term by teaching.

1.1 Stress in English and French

The source of stress deafness seems to lie in the way in which French functions differently from many other languages. Lexical stress is relative and can be defined as the perceived prominence of one syllable as opposed to others in a word. Phonetically, stress can be correlated with increases in fundamental frequency (F0), in duration, and in intensity (loudness), though the relationship is neither direct nor simple. In English, lexical stress is a key characteristic of pronunciation (see Ladefoged & Disner, 2012). Each multisyllabic word has a primary stressed syllable, and though there are some patterns in how stress is placed, they do not explain all of stress placement. English also uses lexical stress contrastively, for example for distinguishing between nouns and verbs in word pairs like /ˈɛk.stɹækt/ (the noun *extract*) and /ɛkˈstɹækt/ (the verb *extract*). Word stress is also key to understanding spoken English. In an analysis of how nineteen different measures affect how English-L1 speakers rate comprehensibility (listeners' ratings of how easily they understand speech) and accentedness of French-L1 speakers, errors in word stress were significantly negatively correlated with both ratings of comprehensibility (-.76) and of accentedness (-.78), and word stress was the only measure included in the best model for both ratings[1] (Trofimovich & Isaacs, 2012).

In contrast to the central role of lexical stress in English, stress is not contrastive or variable in French. Scholars divide into two camps on French lexical stress: some argue that lexical stress does not exist and analyze stress in French only at a higher phrasal level (e.g. Avanzi & Dubosson, 2013; Sichel-Bazin, 2016), and others argue that lexical stress in French is invariable, falling always on the final (full) vowel of a word (e.g. Andreassen & Eychenne, 2013; Kabak et al., 2010).[2] Whatever the inexistence or invariability of lexical stress, French does not use lexical stress contrastively, meaning that there are no minimal pairs that depend on stress placement. There are also no lexical exceptions, such as in loanwords, to its lack of stress or stress rule. Therefore, stress placement in French does not provide any distinctive information about a word, and French speakers do not need to process stress in order to understand and use their language. Thus, following the perception of speech based on one's first language, it is logical that French-L1 speakers would have difficulty perceiving English lexical stress.

1.2 Stress Deafness

Stress deafness in French speakers has been shown through various methods and in comparison to speakers of various other languages. Stress deafness was first identified and described in Dupoux and colleagues' 1997 paper through ABX tasks involving trisyllabic pseudowords that differed only in stress placement, with French-L1 speakers performing less accurately than Spanish-L1 speakers in distinguishing such items. Participants subsequently heard three words and had to decide if the third word (X) was identical to the first word (A) or the second word (B) by pressing a corresponding button. In 2001, Dupoux and other colleagues found similar evidence of stress deafness by comparing the same populations and using a new paradigm with disyllabic pseudowords, in which participants are trained to

---

[1] The model for comprehensibility also included "type frequency", i.e. variation in vocabulary, and "grammatical accuracy," and the model for accentedness also included "rhythm," which was operationalized as vowel reduction.
[2] In southern French (often called *français du Midi* in French) a word-final, unstressed schwa is typically pronounced (often referred to in French as the *e caduc*). For details see Andreassen & Eychenne (2013) and Sichel-Bazin (2016). See Peperkamp et al. (2010) for application of this difference to research on stress deafness.

associate each pseudoword to a key (1 or 2) and then recall a sequence of pseudowords by typing the corresponding keys (e.g. 1211 for a sequence of four pseudowords). This sequence recall paradigm, when used with stimuli that presented sufficient phonetic variability and memory load (see experiments 3 and 4), led to such different performance between French and Spanish participants that the distributions of scores for each L1 group did not overlap, with the French-L1 speakers performing consistently worse than the Spanish-L1 speakers on recalling sequences of pseudowords that only differed in their stress placement. Evidence of stress deafness in French speakers also appears in babies and in relation to languages besides Spanish, as demonstrated by a comparison of ten-month-olds learning only French to those learning both French and one of various stress-contrastive languages in a head-turn preference procedure (Bijeljac-Babic et al., 2012). Through a sequence recall paradigm, Peperkamp and colleagues (2010) also showed stress-deafness in French speakers and speakers of Hungarian and Finnish (languages which have stress always on the initial syllable of a word and do not have any lexical exceptions to this rule), who exhibited greater stress deafness in comparison to both Spanish speakers and speakers of Polish, a language which has a fixed (penultimate) stress pattern but includes lexical exceptions to this stress pattern.

Stress deafness is not absolute; French-L1 speakers seem to process the acoustic correlates of stress but not to process stress abstractly at the phonological level. Even in experiments using the robust sequence recall paradigm, French speakers still perform above chance, suggesting they may encode stress at a phonetic level (Dupoux et al., 2001). In simplified situations where there is less acoustic variability and demand on memory, stress deafness does not appear: Dupoux and colleagues (1997) found that French-L1 speakers do not exhibit evidence of stress deafness in a simple AX paradigm test with tokens from only one speaker, and Dupoux and colleagues (2001) later showed that French-L1 speakers do not exhibit evidence of stress deafness when a single token is used for each pseudoword in a sequence recall paradigm experiment (see experiment 5). Similarly, Bijeljac-Babic and colleagues (2012) showed through a head turn preference procedure experiment that monolingual French-learning 10-month-olds perceive stress better when familiarized with a pseudoword for two minutes as opposed to only one minute, and attributed the poorer performance at one minute to the babies only being able to perceive stress at a phonetic rather than phonological level.

The origins of stress deafness appear to lie in early exposure to language. In Bijeljac-Babic and colleagues' experiment (2012), ten-month-olds learning two languages—a stress-contrastive language and French—were able to discriminate more easily than the French-only babies in the difficult, one-minute familiarization condition *only* if they were exposed to the stress-contrastive language much more than French. Dupoux and colleagues (2008) found that French-L1 speakers who learned Spanish after the age of ten, including those at an advanced level and living in Barcelona and thus in a Spanish-language environment,[3] demonstrated stress deafness in comparison to Spanish-L1 controls in both a sequence recall task and a lexical decision task. In the sequence recall task, there was no significant difference between the late learners and the French monolingual controls. Later, Dupoux and colleagues (2010) compared French-Spanish simultaneous bilinguals to native speakers of Spanish and French-L1 late learners of Spanish from the 2008 experiment, again using the same sequence recall paradigm. The performance of the simultaneous bilinguals was best modeled by a bimodal distribution, with some performing more like the Spanish-L1 speakers and some more like the French-L1 late learners of Spanish. The only significant correlates for degree of stress deafness in this group were country of residency at 0-2 years of age and 2-4 years of age, suggesting a crucial impact of the linguistic environment at a young age.

1.3 Societal Relevance

The phenomenon of stress deafness is theoretically interesting in that it can help us learn how speech perception occurs at a suprasegmental level and how perception is related to the characteristics of a speaker's first language. Beyond these implications for theoretical understandings of human language, stress deafness reflects a broader issue in foreign language learning for French-L1 speakers. In France, all students are required to study at least two foreign or regional languages, starting with one in primary

[3] Only four of the thirty-three living in Barcelona reported knowledge of Catalan.

school, adding a second in middle school (collège) and continuing to study foreign languages through the end of high school (MENJ, 2022) and usually in university (e.g. Ministère de l'enseignement supérieur, de la recherche et de l'innovation, 2021). 96% of French students study English in primary school and by secondary school, 99% study English (MENJ, 2022, Dec 12). The most common other languages studied are Spanish, German, and Italian (Onisep, n.d.), which also all present contrastive lexical stress (for examples and explanation see, respectively, Dupoux, 1997; Sippel & Martin, 2023; Giraudo & Montermini, 2010). Therefore, it is important that students learn to perceive and produce lexical stress for effective communication in the languages they are already learning.

French language policy states that practice of oral skills is a priority (MENJ, 2022), and in accordance with this, a phonology section was added to the national exam for becoming a teacher of English, known as the CAPES,[4] in 2022 (Vaudin, 2024). However, in practice, pronunciation teaching and learning is often a weak point. Teaching of pronunciation is often neglected or avoided in France, at least partly due to teachers' hesitation in teaching the topic (Frost & Henderson, 2013), and pronunciation and lexical stress are deemed difficult even by those students who choose to specialize in English in university (Moore-Mauroux, 2010; Capliez, 2011). These observations are reflected in more recent reports on the national exams for prospective teachers since the addition of phonology, which have noted that some of the most frequent errors in candidates' pronunciation are those of misplaced word stress and unrealized vowel reduction (Turin-Bartier, 2022; Vaudin, 2024). The reports note that prospective teachers often neglected to address phonology in their lessons, and when phonology was addressed, often only addressed segmental features and not suprasegmental features such as lexical stress (Turin-Bartier, 2022). In 2024, out of all questions, candidates performed worst on the phonology question, which was about lexical stress (Vaudin, 2024). The evidence from the CAPES records shows that francophone people, even studying to teach English, struggle with phonology, especially lexical stress. As those teachers taking the

[4] The *CAPES (Certificat d'Aptitude au Professorat de l'Enseignement du Second degré)*, once passed, allows teachers to be permanent state employees *(titulaires)*. One in ten teachers in the secondary classroom however, have not yet passed this exam, as there is a teacher shortage in France and teachers are increasingly now hired on contract (as *contractuels*) with lower requirements and minimal training (Cours des Comptes, 2024).

CAPES before 2022 were not required to focus on phonology, and phonology may not have been mastered even by those who passed the exam after 2022—for example, candidates could pass in 2024 with 8.2 out of 20 points (Vaudin, 2024)—it remains difficult to ensure that French students are taught properly about lexical stress. Furthermore, this necessitates that ways to help francophones learn to perceive, recognize, and produce lexical stress—both in the training of teachers and in their teaching of school students—be effective and scientifically validated.

1.4 Application to Education

Heretofore, only a handful of studies have examined a method to reduce stress deafness in francophones, and none of these has to my knowledge used real English words in either training or testing. Carpenter (2015) tested French and English speakers on a stress perception ABX task based on that of Dupoux and colleagues (1997), with half of participants in each language group receiving training on stress perception immediately before the ABX task. The training used the perceptual fading technique: it consisted of a series of artificially manipulated nonsense words with exaggerated pitch, intensity, and duration, and the degree of exaggerated duration decreased over the course of the training to make the stress contrast gradually less clear. Participants heard each token and had to identify which of the three syllables was stressed, repeating if necessary until they identified the stress correctly. Such training proved to be effective in decreasing the stress deafness of French-L1 speakers at least in the short term: while the untrained French-L1 participants performed significantly worse than untrained English participants on the ABX task, showing a stress deafness for the francophones, trained French-L1 participants performed significantly more accurately on the ABX test than the untrained French-L1 participants.

In a working paper, Schwab and Llisterri (2014) describe a study in which a French-L1 control group was not trained, and another French-L1 group was trained by learning to match Spanish pseudowords that were distinguishable only by stress placement with different arbitrary shapes. Participants did both a pre-test and post-test in which they listened to trisyllabic Spanish pseudowords and indicated on which syllable the stress fell by pressing a button. In the training, French-L1 participants' accuracy increased significantly over subsequent blocks, showing an ability to learn to perceive the stress placement. In regard to the pre-test and post-test, however, this experiment produced unexpected results, in which the trained French-L1 group showed no significant improvement between a pre-test and post-test, while the untrained French-L1 control group did show a statistically significant improvement. However, the interaction comparing the post-test to pre-test difference in both groups did not reach significance, demonstrating that such a difference between the groups could not be interpreted. The authors attributed this unexpected finding to a dependence not simply on training but on the participants' learning ability during training.

Building later on the previous study, Schwab and Dellwo (2022) tested French-L1 speakers who did not have knowledge of a romance language with varying stress patterns (Spanish, Italian, or Portuguese). They placed these participants into three groups: those undergoing explicit training, non-explicit training, and no training. Both trainings were a total of four hours over eight sessions spread over two weeks and were delivered via computer. The explicit training included metalinguistic explanation of lexical stress and activities mimicking classroom methods. The non-explicit training was based on that of Schwab and Llisterri (2014), again learning to match Spanish pseudowords that differed only in stress placement with different arbitrary shapes. The increase in accuracy between a pre-test and post-test of stress identification in Spanish words was significantly greater in both trained groups than in the untrained groups, without a significant difference between the improvements in the two trained groups. Together, these studies show that stress deafness is malleable and can be lessened to some degree, though, as discussed before, stress deafness persists, even in advanced late learners (Dupoux et al., 2008) and in some simultaneous bilinguals (Dupoux et al., 2010) in a difficult task like sequence recall with phonetic variability.

1.5 Aims and Research Question

The present study aims to add to and advance this nascent literature on modification of stress deafness in French-L1 speakers, specifically in the context of English learning. We thus seek to answer the research question: Does a lesson aimed at increasing lexical stress awareness improve French-L1 speakers' L2 English lexical stress perception, as compared to a similar lesson that does not highlight lexical stress?

The present study seeks to narrow the large gap between training groups and control groups in studies on training for stress deafness. In the studies of Carpenter (2015), Schwab and Llisterri (2014), and Schwab and Dellwo (2022), the control group had no training or exposure to language whatsoever. By using a stress-aware training and a similar stress-ignoring training, the present study mimics what may happen in a language classroom, rather than simply comparing with no training; in both cases the same vocabulary is taught, but in one, stress is emphasized and explicitly addressed, thus promoting metalinguistic awareness of lexical stress, while in the other it is ignored. Participants in both groups are exposed to the same words pronounced authentically with their lexical stress. Rather than a training only focused on stress, this study seeks to implement something more comparable to what the CAPES jury recommended teachers do in their 2022 report: incorporate classroom rituals, like color coding or use of gesture, to systematically indicate stress placement in new words (Turin-Bartier, p. 32).

Another contribution of the present study is to use an English environment for the training and the perception task, rather than Spanish words or pseudowords. Spanish lexical stress differs from that of English in two crucial ways: it is marked orthographically when it does not follow the default pattern, and it does not involve vowel reduction. The orthographical marking via accent marks helps to make stress in Spanish more salient to learners and makes the explanation of lexical stress more difficult to avoid than in English, where it is not marked orthographically. Vowel reduction in English is an important part of stress contrasts and natural English speech, so using natural English stimuli for the perception task better approximates the real perceptual environment of English, whereas even studies comparing French speakers to English speakers have used pseudowords more similar to Spanish stress patterns (see Carpenter, 2015).

We hypothesized that the training would have a positive effect on accuracy for stress-based items, with the French speakers having undergone the stress-aware training outperforming those who underwent the stress-ignoring training. Expectations of such an effect were based on the positive effects of training shown by Carpenter (2015) and especially Schwab and Dellwo's (2022) findings of improvement in the similar explicit condition. Compared to Schwab and Dellwo's explicit training, which lasted four hours spread over eight sessions and two weeks, the training in the present study is short-term, but could still have an immediate effect, as did Carpenter's (2015) perceptual fading training that lasted 15 minutes.

2. Method

2.1 Participants

Participants were recruited through personal contacts including employees at universities and schools, relevant (such as English-learning) interest groups on the internet, and through a database of participants hosted by the CNRS (France's national scientific research center, see https://expestest.risc.cnrs.fr/public/index.php). Participation was voluntary, and the project was approved by the Ethics Committee of the Faculty of Humanities of the University of Amsterdam (EC number: FGW-5128). All ten participants were native speakers of French who were not native speakers of any other language, nor were they exposed to another language in the home before the age of six. All participants had grown up and were currently living in mainland France, in regions mostly in the northern half of France and where southern French is not spoken (Île-de-France, Hauts-de-France, Pays de la Loire, Normandie, Bretagne, Auvergne-Rhône-Alpes).[5] Participants reported not having any uncorrected visual or auditory impairments or any language-related disorders, such as dyslexia. No participants specialized in language studies at university.

Five participants (three female, two male) followed the stress-aware lesson and were 21 to 64 (m = 32.8, SD = 17.66) years of age. They scored an average of 70% (SD = 5.26%) on the LexTALE (Lexical Test for Advanced Learners of English, Lemhöfer & Broersma, 2012). Of these participants, four reported having studied German as their second foreign language and one reported no secondary foreign language. Five participants (one female, four male) took the stress-ignoring lesson and were 20 to 64 (m = 40, SD = 18.96) years of age. They scored an average of 66.00% (SD = 9.50%) on the LexTALE. Two of

[5] Again (see footnote 2), southern French functions somewhat differently from other mainland French dialects in regards to stress. Most research on stress deafness in French speakers has focused on non-southern French.

these participants reported studying German as a second foreign language, while three reported studying Spanish, two of whom reported studying a different third foreign language. Overall, scores on the LexTALE ranged from 50% to 76%, with an average of 68%. All but one participant performed at 60% or above, aligning with B2 level of the CEFR according to Lemhöfer and Broersma (2012, Appendix C, Table 9).

2.2 Materials

2.2.1 Lesson

Two lessons were developed as training for French-L1 participants with basic to intermediate English proficiency (approximately A2/B1 according to the CEFR). The lessons were recorded in PowerPoint to allow maximal similarity between the versions, with a voiceover and video of the instructor in the corner in order to be able to show some gesture, particularly counting syllables on fingers, and thus more closely resemble a classroom setting. Both lessons focused on a theme judged to be relevant and interesting to the target group of participants: renting an apartment in New York City. This also aligned with the experimenter conducting the lesson, a speaker of American English who had previously lived in the city, giving some cultural authenticity to the lesson in order to encourage participant motivation (see for example, Celik & Yıldız, 2019). The experimenter had training in teaching English to beginning and intermediate-level (A2 and B2 level) adults as well as two years of experience teaching full-time to francophones similar to the participants (A2 – C1). Both lessons focused on the same vocabulary and were made as similar as possible, with five and a half minutes of identical content, out of 17.5 minutes total, and the same basic PowerPoint slides, with some modification for each version. Lessons were conducted mostly in English with less than two minutes of French speech in each version, with clarity supported by language adapted to the learners' assumed level, rephrasing, and visuals and text on the slides.

Both lessons were interactive, designed to promote attention, comprehension, and learning. Each lesson contained nine comprehension questions, five of which were identical between the versions, and these were interspersed every one to three minutes throughout the lesson. These comprehension questions pertained to what had just been explained in the video and did not require participants to extend concepts beyond the presented examples; as such, they were relatively simple and served as a way to ensure participant attention and comprehension and have a record of these. After answering each question, a short clip of the relevant segment replayed. If the question had been answered correctly, the video continued. If it had been answered incorrectly, the participant was prompted to answer again and watch the replay until the correct answer had been given. Each lesson also contained two sorting tasks: a first identical task to sort words by number of syllables and a second sorting task after the end of the lesson based on the lesson type: sorting disyllabic words by stress pattern for the stress-aware lesson and sorting disyllabic and trisyllabic words by number of syllables for the stress-ignoring lesson. In each sorting task, a word appeared in the middle of a board, and participants had to drag and drop the word into the appropriate category. By clicking on the word, participants heard it pronounced by the instructor. Once all words were sorted, they were outlined in red (incorrect) or green (correct) in order to prompt participants to re-sort the word into the appropriate category until the task was completed successfully.

The stress-aware lesson focused on stress by explaining syllables, word stress, and vowel reduction, including the schwa, and then identifying stress in familiar words and words that were likely new to participants that were introduced throughout the lesson. Explanation of stress included a definition, examples, and the mention of phonetic correlates (longer, higher, louder), with the caveat that every stressed syllable does not necessarily present all of those characteristics. The stressed syllable was defined as being more important than the other syllables in the word and standing out, and it was also explained that stress could fall on any syllable of a word in English. Participants were also given counter-examples, in which stress was incorrectly placed, as in the stress training provided by Sippel and Martin (2023) for anglophone students learning German stress. Syllables were marked in example words with a circle above each syllable and a larger circle above the stressed syllable, similar to the marking in the training by Schwab and Dellwo (2022) for francophones learning Spanish stress. Participants were told that word stress in English is very different from the functioning of French, which makes it something especially important for francophones to learn, and were demonstrated the tendency (but not rule) for disyllabic nouns to be stressed on the first syllable and disyllabic verbs to be stressed on the second syllable.

In the stress-ignoring lesson, the number of syllables was identified and explained. As in the stress-aware lesson, syllables were marked with a circle above each syllable; however, all circles were of the same size since stress was ignored. It was explained that the number of syllables could differ in French-English cognates and that the number of syllables did not necessarily match the number of vowel groups in a written word. Otherwise, repetition of key vocabulary and overall length was made similar to that of the stress-aware lesson by adding more commentary on the similarities and differences between apartments and renting in New York City and French cities.

2.2.2 Perception Task

For the critical stimuli, three English disyllabic stress-based minimal pairs were selected. These pairs are nouns when stressed on the first syllable (trochees) and verbs when stressed on the second syllable (iambs). They were selected on the basis of not sounding close to a similar French word, having a relatively balanced frequency ratio between the noun and the verb form according to the SUBTLEX-US database (Brysbaert & Keuleers, 2012), and not containing any diphthongs. Six other English minimal pairs were selected as fillers: three that depend on the /i/-/ɪ/ vowel contrast, expected to be difficult for French speakers, and three that depend on the /s/-/z/ voicing contrast, expected to be easier for French speakers. The /i/-/ɪ/ contrast does not exist in French and is known to be difficult for French learners of English to produce and perceive (see, for example, Moore-Mauroux, 2010; Turin-Bartier, 2022, p. 32; Walter, 2001), while the /s/-/z/ contrast creates minimal pairs in French, such as /pwa.sɔ̃/ (*poisson*, 'fish') and /pwa.zɔ̃/ (*poison*, 'poison') or /kas/ (*casse*, 'break'[6]) and /kaz/ (*case*, 'box'), meaning that francophones are able to distinguish the two sounds, including in intervocalic and word-final positions (and such a contrast is thus not highlighted as a difficulty in guides to teaching English pronunciation to French speakers, such as Moore-Mauroux, 2010, and Walter, 2001). None of the words used in the lesson

---

[6] This is the pronunciation of the verb in first, second, and third person singular as well as in third person plural.

were reused in the ABX task, with the exception of "living" (a filler item), which was not analyzed for syllables, pronunciation, or stress pattern during the lesson.

Two young adult native speakers of west coast American English, a female from California and a male from Oregon, were recorded in a sound-proof booth. Their accents are similar to that of the speaker in the instructional video. The speakers repeated words from the minimal pairs described above and were instructed to emphasize word stress and include vowel reduction as in normal speech. Tokens of the critical stress contrast words spoken by both speakers were annotated in Praat (Boersma & Weenink, 2024) and assessed for vowel duration, pitch, and intensity. Due to the presence of creak in many vowels of the male voice, pitch was not able to be identified for all tokens. Creaky voice was not, however, taken to be problematic as it is common in the English of young Americans of both genders (e.g. Abdelli-Beruh et al., 2016) and is also very often present in the speech of the instructional video speaker.

These recordings presented considerable vowel reduction on a segmental level, most notably in the first syllable of the iambic words as opposed to their trochaic pairs. Thus, the first vowel in "refund" was reduced from /i/ to /ə/ (/ˈɹi.fʌnd/ to /ɹəˈfʌnd/), in "conduct" from /ɑ/ to /ə/ (/ˈkɑn.dʌkt/ to /kənˈdʌkt/), and in "extract" from /ɛ/ to /ɪ/ (/ˈɛk.stɹækt/ to /ɪkˈstɹækt/). The vowel reduction was more pronounced generally in the male speaker, and the amount of vowel reduction varied some between tokens of the female speaker.

For the stress-based minimal pair tokens, ratios of duration, pitch, and intensity in stressed syllables as compared to unstressed syllables were calculated using Praat and R with the tidyverse and dplyr packages (R Core Team, 2024; Wickham et al., 2019; Wickham et al., 2023). Duration was the most reliable indicator of stress in these tokens, with the vowel in the second syllable in both trochees and iambs being on average longer than the vowel in the first syllable, but the ratio of second to first vowel length being much greater in iambs (3.44 for the male speaker and 2.94 for the female speaker) than in trochees (1.13 for the male speaker and 1.32 for the female speaker). This pattern resembles that of similar stimuli in Höhle and colleagues (2009), and these stimuli are likely also influenced by final lengthening when recorded as isolated, disyllabic words. Because of the reliability of duration as an indicator, the most prototypical tokens were determined as those having the highest ratio of stressed to unstressed vowel for duration, as well as high values for ratios for pitch and intensity and sounding clearly stressed to an English-L1 speaker and a trained phonetician. To determine the most ambiguous tokens, the opposite criteria were used, first selecting for the lowest duration ratio, along with relatively low ratios for pitch and intensity.

The selected tokens were regularized to a constant peak of .99 via a Praat script in order to sound equally loud. The tokens used as stimuli A and B were from the male speaker and had a falling pitch contour, while the tokens used as stimulus X were from the female speaker and had a rising pitch contour. This provided a large amount of phonetic variation between A, B, and X stimuli, both in terms of the average fundamental frequency of each speaker and in terms of pitch contour. A high amount of phonetic variation in stimuli is needed to detect stress deafness in French speakers as this variation requires them to perceive stress at a phonological rather than phonetic level (see again, Dupoux et al., 1997 & 2001).

For each stress-based minimal pair, two sets of four ABX triads were constructed, for a total of eight triads per minimal pair. Each set contained all the possible orders of stimuli: trochee-iamb-trochee, trochee-iamb-iamb, iamb-trochee-trochee, iamb-trochee-iamb. One set used prototypically stressed tokens for X, expected to be easier items as they more closely aligned with the expected correlates of stress, and the other set used more ambiguously stressed tokens for X, expected to be more difficult items. This ensured a range of difficulty in the critical stimuli. Both sets used prototypically stressed tokens for A and B. One pair of tokens were used for the A and B stimuli for all triads in the first set, and a second pair of tokens were used for the A and B stimuli in the second set. See Table 1 as an example.

Table 1: Example Setup of Stimuli

	A token	A word	B token	B word	X token	X word	Correct
Set 1	Token 1M	ˈrefund	Token 2M	reˈfund	Token 3F	ˈrefund	A
	Token 1M	ˈrefund	Token 2M	reˈfund	Token 4F	reˈfund	B
	Token 2M	reˈfund	Token 1M	ˈrefund	Token 3F	ˈrefund	B
	Token 2M	reˈfund	Token 1M	ˈrefund	Token 4F	reˈfund	A
Set 2	Token 5M	ˈrefund	Token 6M	reˈfund	Token 7F	ˈrefund	A
	Token 5M	ˈrefund	Token 6M	reˈfund	Token 8F	reˈfund	B
	Token 6M	reˈfund	Token 5M	ˈrefund	Token 7F	ˈrefund	B
	Token 6M	reˈfund	Token 5M	ˈrefund	Token 8F	reˈfund	A

The task included 24 stress contrast trials, 16 voicing contrast trials, and 8 vowel contrast trials for a total of 48 trials. The voicing contrast and vowel contrast trials were constructed in a similar way to the critical trials, but the tokens were not analyzed in Praat for their characteristics and were instead chosen on the basis of sounding prototypical and clear. For each contrast, initially 24 triads were constructed, and then a subset was selected such that orders, minimal pairs, and A versus B correct responses were balanced, or as balanced as possible.

2.3 Procedure

Participants accessed the experiment online. They were instructed to complete the experiment in a quiet room and use a computer and headphones. The entire experiment was run in Experiment Designer (v2024.04, Vet, 2024). Because of participants' possibly low level of English proficiency, consent documents and background questions were in French, while instructions were given in English and French. In-lesson questions and activities were in English, but supported pedagogically by guidance and examples. After giving informed consent, participants proceeded to several biographical questions (see Appendix B), a LexTALE English task, and then watched either version of the lesson. After the conclusion of the lesson, including comprehension questions and two sorting activities, participants started the ABX perception task.

Each A, B, and X stimulus was separated by an interval of 1000 ms. Participants were informed that they would hear series of three words in English and needed to decide if the third word resembled the first or second word more and press the ‘f’ or ‘j’ key, respectively. The answer period ended upon press of the ‘f’ or ‘j’ key or after a deadline of 4000 ms if no press had been made. The next trial began after 1000 ms. This timing reproduces the timing of ABX tasks on stress deafness by Dupoux and colleagues (1997) and Carpenter (2015), except that the interval between A, B, and X in this experiment is twice as long (1000 ms as compared to 500 ms). The interval was extended to make the task less overwhelming for participants and also to encourage representation at a phonological level.

First, participants completed five practice trials, for which they were not given any feedback, and then proceeded to the experimental trials. The main task consisted of a total of 48 trials and the entire task lasted approximately six minutes. Trials were not organized into blocks but were arranged randomly, under the condition that no two consecutive trials concerned the same minimal pair. Participants were given one self-timed break halfway through the trials.

After the ABX perception task, participants answered a few final questions about their experience (see Appendix C) and could do so in English or French, as they chose. The experiment then concluded.

2.4 Exclusion

Any potential participant was automatically excluded if they received a score of 80% or above on the LexTALE, aligning with C1+ on the Common European Framework of Reference (CEFR) for Languages according to Lemhöfer and Broersma (2012, Appendix C, Table 9). Such people were excluded because the lesson would be much below their level. For the ABX task, all trials on which the participant did not respond in time (by 4000 ms after end of the last stimulus) were removed from the data. Four participants were excluded because they did not perform above 65% on at least one of the contrast types (voicing, vowel, or stress) in the ABX perception task. Performance near chance level (50%) could be due to not understanding the task and guessing, while much lower performance could be due to the confusion of keys. Relatively high (65%+) performance on at least one of the contrasts shows the participant has understood the task correctly and properly performed it. Additionally, one participant was excluded from analysis as the participant saw both versions of training due to stopping and restarting the experiment.

2.5 Analysis

To answer the main research question, an analysis of the effect of training was planned in R (R Core Team, 2024) through a generalized linear mixed-effects model (glmer, binomial family) from the lme4 package (Bates et al., 2015) on accuracy for the stress-contrast items. The presence of a main effect of training would answer the research question, with an expected positive effect of stress-aware as compared to stress-ignoring training. The most minimal model (and thus included in any valid model) is:

accuracy ~ training + (1 | participant) + (training | item)

Ideally, a model also controlling for X type (if the X is a more prototypical or ambiguous token) and for the specific minimal pair, as these are expected to influence accuracy, would be run:

accuracy ~ training * Xtype * pair + (Xtype * pair | participant) + (training | item)

Many other factors could also be controlled for through a model, such as effects of the order of stimuli or age of participants. A full list of such possible fixed effects for use can be found in Appendix A. Because of the small number of participants, fixed effects besides "training" were not included in the main model as they could not run without singular fit or failure to converge.

To clean data, the initial files from the experiment were aggregated and all unnecessary rows and columns were removed. Columns were renamed according to the names specified in Appendix A. Using R, all trials in which there was no response within the 4000 ms time limit were removed. Accuracy averages by contrast were then calculated for each participant, and exclusions were applied as discussed in the above section.

3. Results

3.1 Lesson Performance

All participants initially answered the nine in-lesson questions at above chance accuracy[7] (*m* = 86.00%) and answered at least eight of the nine questions correctly by the second attempt (*m* = 98.89%),

[7] 4 of the 9 questions, or 44%, is above chance level because the varying number of possible responses per question make chance level 38%.

after rewatching the relevant video clip. This demonstrated participants were paying attention and largely comprehending the lesson.

3.2 Participant Experience

In the post-ABX task questions (see Appendix C), three of four participants in the stress-aware training responded that the lesson was useful for the listening task.[8] When elaborating on why the lesson was useful, one of these particularly stated they had not known about vowel reduction, another generally cited the concepts of stress and vowel reduction, and the third attributed the usefulness generally to clear explanations. The participant in this condition who responded that the stress-aware lesson was not useful cited as a reason previous knowledge of the function of stress in English. When asked the purpose of the experiment, one participant specifically mentioned stress, guessing accurately that the experiment sought to determine the ease or difficulty participants had with English stress depending on their native language. The other participants referred more generally to studying ways of learning English pronunciation. Overall, these responses demonstrate that those in the stress-aware lesson generally understood that the perception task measured stress perception ability.

Of the five participants in the stress-ignoring lesson, only one responded that the lesson was useful for the listening task, and the remaining four responded that it was not. Three of those who responded that the lesson was not useful explained that the focus on syllables was not helpful and the fourth did not elaborate. No participant who had followed the stress-ignoring lesson mentioned any awareness that the goal of the experiment was related to word stress; one mentioned "l'accent" but elaborated that the perception exercise may have been used to measure capacity to distinguish British and American accents, so it is unlikely they meant "accent" as word stress. Guesses as to the purpose were generally more vague: "to know what is the best way to learn a new language or to know what sound can be detected by french [sic] people"[9] and measuring the effectiveness of English teaching. One participant

[8] Data was unavailable for one participant, whose ABX task ended early.
[9] Participants were free to answer these questions in English or French. This quotation is from the only person who answered in English. The other answers in this paragraph are English paraphrasing of the original French answers.

simply responded that they had no idea. Overall, it appears then that participants in the stress-ignoring lesson remained unaware of the lexical stress concept during and after the perception experiment. Comparing this to the responses of the group in the stress-aware training, it seems then that metalinguistic awareness of the lexical stress concept, at least for the time of the perception experiment, was successfully manipulated by the training.

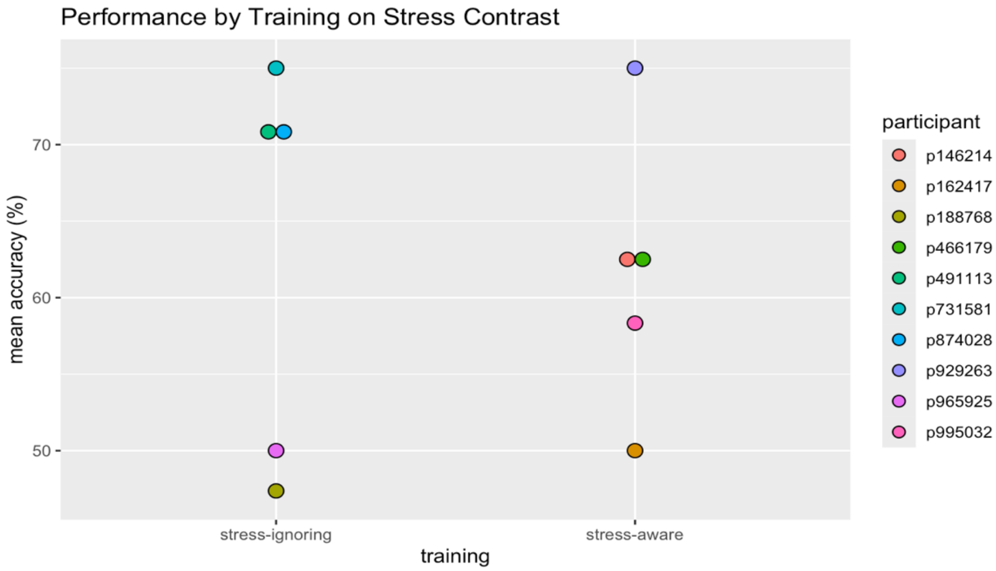
3.3 ABX Task

One participant in the stress-aware training reported encountering a technical difficulty partway through the ABX task, timing out on a series of items and then exiting the experiment. Therefore, this participant only answered 28 items of the 48, but responses before the series of timed-out items show that the participant was likely understanding the task. Therefore, the participants' results have been retained, though incomplete. Another participant, in the stress-ignoring training, timed out on 11 items in a row, presumably because of a distraction, but otherwise performed regularly and was also included. The remaining participants all timed out on one item or on no items, providing full or almost-full data sets.

3.3.1 Lesson Version Effect on Stress Contrast

Sample size for this study was very limited. This highly constrains any statistical analysis. Additionally, as gender, age and foreign language studied were not balanced between the training groups, any model that does not control for these factors may reflect their impact rather than simply the effect of training. Because larger models were not able to be run on such a small data set, the minimal model described earlier was chosen. Power of the minimal model to find an effect if it exists was calculated post hoc with the powerSim function from the simr package in R (Green & Macleod, 2016), with an α of 0.5 and for an effect size of 0.30, based on the observed effect size of training for French speakers in Carpenter's (2015) study also involving an ABX task. The resulting power was 12.67% (95% confidence interval running from 9.12% to 16.97%). Therefore, the model with the given data is unlikely to detect any such an effect, even if it does exist, so Type II error is likely.

Figure 1



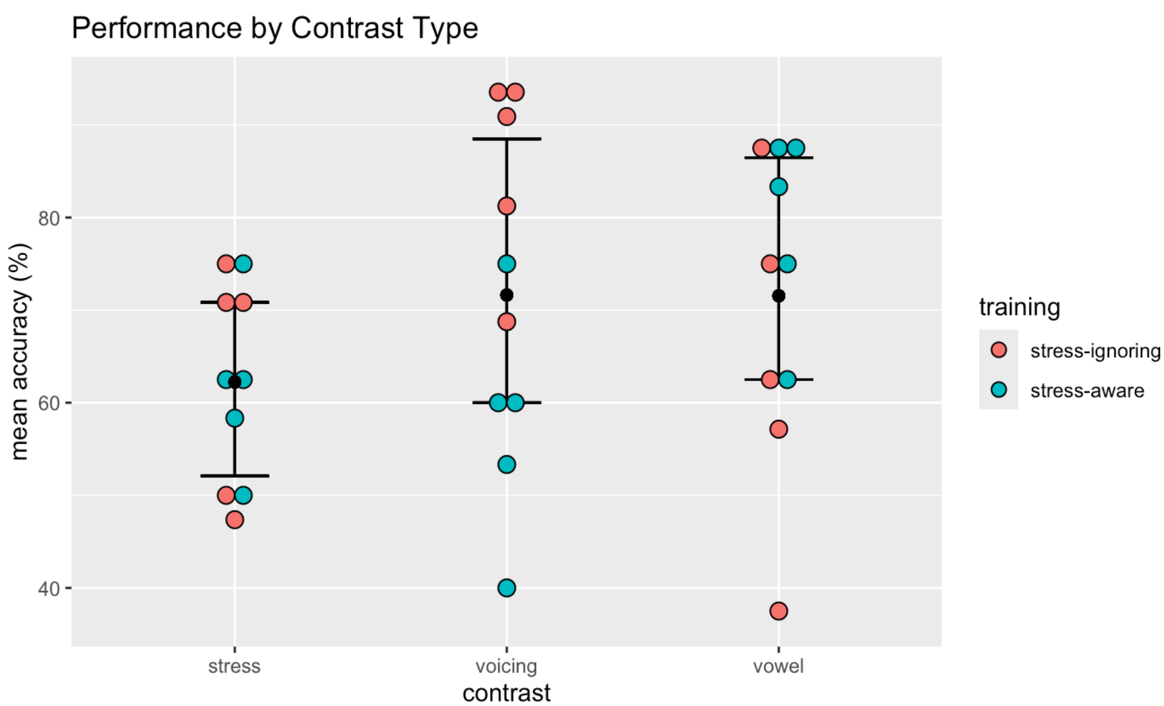
With that in mind, the planned minimal model was run for accuracy on the stress contrast items. Contrary to our hypothesis, odds of responding correctly to stress-contrast trials were 1.20 times greater for our participants who followed the stress-ignoring lesson than those who followed the stress-aware lesson, but this effect was not significant (z = 0.472, p from 1 = 0.64, 95% confidence interval running from 0.57 to 2.53). We thus are unable to conclude if a lesson affects stress perception in French-L1 speakers and cannot answer our research question. As can be seen in Figure 1, the performance on stress contrast items is variable in both training groups, without an indication of a strong trend. Most participants performed above chance, but none performed above 75% accuracy.

3.3.2 ABX Task Performance by Contrast

In general, the performance on the perception task was quite variable. Accuracy on each item ranged from 22% to 100%. French participants were expected to perform markedly better on the /s/-/z/ voicing contrast, similar to one in their native language, than on the stress contrast and /i/-/ɪ/ vowel contrast, which are not found in French. On average, participants did perform best on the /s/-/z/ voicing contrast (72.41% compared to 62.78% and 71.43%, respectively), with a high of 93.75% accuracy on the /s/-/z/ voicing contrast, as can be seen in Figure 2. However, for individual participants, the same pattern did not always appear; three participants performed worse on the voicing contrast than on the other two.

One participant, for example, performed with 75% accuracy on the stress and vowel contrasts but 53% accuracy on voicing. In any case, performance on the stress contrast was on average above chance, reinforcing the understanding that stress deafness is not absolute.

Figure 2



Note: Each colored dot represents one participant's average accuracy on the given contrast.

An unexpected observation from the data is that those who followed the stress-aware training performed worse than those who followed the stress-ignoring training on the voicing contrast items (40.85% error compared to 27.59% error, see Table 2 and Figure 2). As the experiment was not designed to address the difference between the stress contrast and voicing contrast or the voicing contrast alone, such results and interpretations must be taken with caution. The voicing and vowel contrasts had fewer items (16 and 8, respectively) and were not analyzed in detail for their phonetic properties. The larger range in accuracy on the voicing and vowel contrasts is also partly attributable to these smaller number of items. Any such analysis is thus exploratory and tentative.

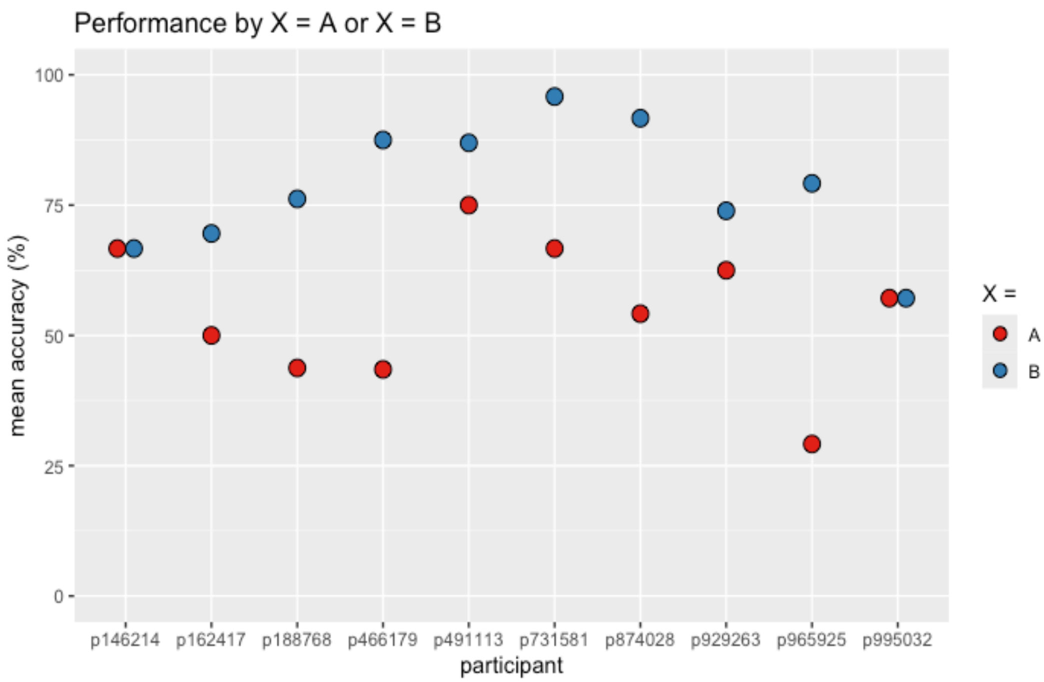
Table 2 Mean Error Rate and Standard Deviation

Training group	Stress contrast			Voicing contrast (/s/-/z/)		
	Pair	Error Rate (%)	*SD* (%)	Pair	Error Rate (%)	*SD* (%)
Stress-aware	ˈrefund/reˈfund	31.58	47.10	faces/phases	44.00	50.66
	ˈextract/exˈtract	47.06	50.66	dose/doze	39.13	49.90
	ˈconduct/conˈduct	36.11	48.71	device/devise	39.13	49.90
	mean	37.96	48.76	mean	40.85	49.50
Stress-ignoring	ˈrefund/reˈfund	33.33	47.76	faces/phases	14.81	36.20
	ˈextract/exˈtract	34.21	48.08	dose/doze	20.00	40.82
	ˈconduct/conˈduct	42.11	50.04	device/devise	9.10	29.42
	mean	36.52	48.36	mean	14.86	35.82
All	mean	37.22	48.45		27.59	44.85

3.3.3 Effect of X Matching A or Matching B

One effect observed in Dupoux et al. (1997) and Carpenter (2015) is that accuracy is higher when the X token matches A than when the X token matches B. The data in the present study follows this pattern, with on average 79.46% accuracy for B items and 55.20% accuracy for A items when calculated over all items for all contrast types (see Figure 3). This effect was verified through a post-hoc test. A glmer model for accuracy was run with the correct response type (if X matches A or X matches B) as the only predictor, random intercepts by item, and random intercepts and slopes by participant for correct response type. This model showed that for our participants, odds of responding correctly were 3.28 times greater when X matched B than when X matched A ($z = 4.283$, p from $1 = 1.9 \cdot 10^{-5}$, with a 95% confidence interval running from 1.90 to 5.65). We thus can confirm that, as attested previously, French-L1 speakers doing such an ABX task are more likely to perform accurately when X matches B than when X matches A.

Figure 3

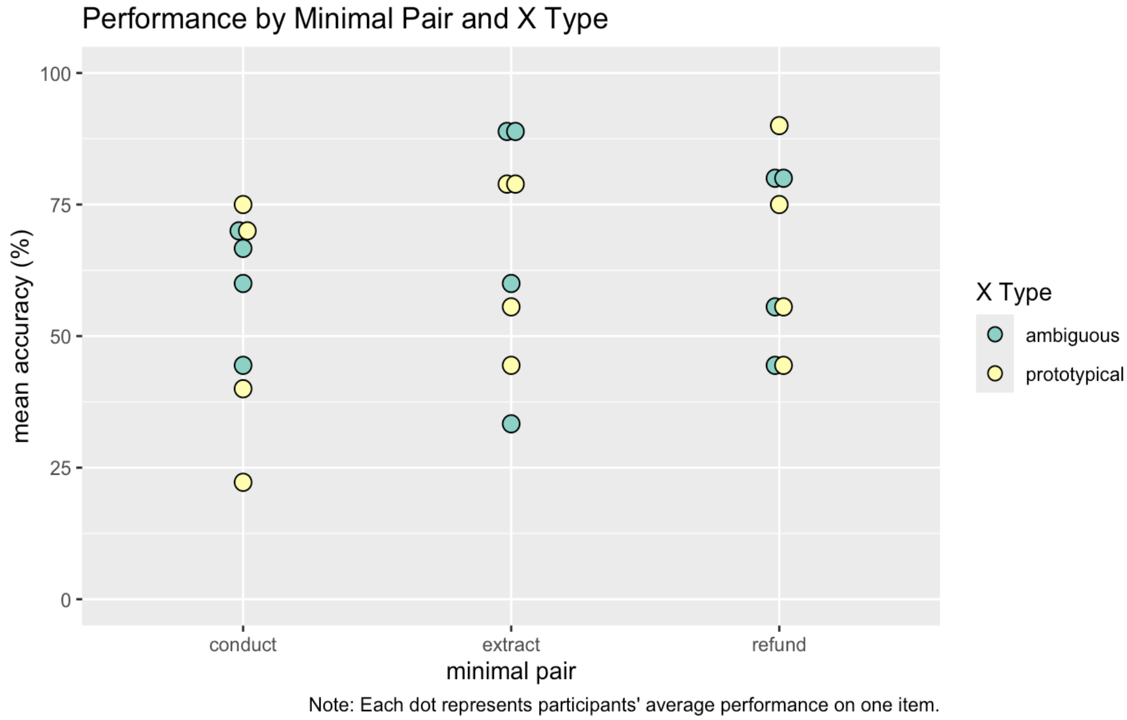


Additionally, it was observed that of the four participants excluded based on low accuracy, three had a marked preference for responding B rather than A (B constituted 85%, 88%, and 93% of responses for each of these participants). In general, participants seemed to gravitate towards B rather than A as a default response.

3.3.4 Analysis by X Type

For each stress contrast item, the X token was either prototypically stressed or relatively more ambiguous. It was expected that the more extreme, prototypical tokens would make items easier for participants as they corresponded to more extreme phonetic correlates: higher ratios of duration and usually pitch and intensity in the stressed syllable as compared to the unstressed syllable. Contrary to expectations, however, participants actually performed slightly better on average on the items with ambiguous X tokens than those with extreme X tokens (64.60% accuracy compared to 60.91% accuracy, see Figure 4). To check for an effect, a glmer model for accuracy was run with only X type (prototypical or ambiguous) as predictor and random slopes by X type per participant. Because of the small amount of data, the model resulted in a singular fit, but because the random slopes correspond to the effect that

Figure 4



answers the question, the model could not be simplified without losing validity. In any case, the p-value of the X type effect was 0.67 and thus not significant. We thus are unable to conclude from the present data if prototypicality of the X token has an effect on accuracy.

4. Discussion

The present study attempted to answer if a lesson focused on lexical stress would improve French-L1 listeners' stress perception of English words. Participants followed either a lesson that emphasized and explained lexical stress (stress-aware) or a similar lesson that ignored lexical stress (stress-ignoring) and then completed an ABX task in which they heard three real English disyllabic words and had to decide if a the third word (X) matched the first word (A) or second word (B). Words in each triad differed only in one of three ways: stress placement (first or second syllable), vowel contrast (/i/-/ɪ/), or voicing contrast (/s/-/z/). The study's main research question of the effect of lesson type on stress perception cannot be answered with the given data, and thus the question requires further research with more participants. Analysis of the effect of whether the X stimulus was prototypical or ambiguous was also inconclusive. As in previous studies using ABX tasks (Carpenter, 2015; Dupoux et al., 1997), our participants were significantly more likely (by 3.28 times) to respond correctly when the X token matched B than when the X token matched A. Though this one significant result was found for a relatively large effect, the extremely small sample size and thus low power of models for any inferential statistics means that, with the given data, these tests are unlikely to detect an effect even if one does exist, especially for smaller effects.

The limited current results are mostly in line with what is known about stress deafness and its detection on ABX tasks: French-L1 speakers perform above chance on stress-contrast items and X=B items are responded to more accurately than X=A items. Both of these patterns are consistent with the understanding that French-L1 speakers encode stress at a phonetic or shallow level. Information at this level is only stored for short periods of time, making the X=B items, in which the matching stimuli are presented closer together, easier than the X=A items, in which the matching stimuli are presented further apart and with an intervening (B) stimulus (Dupoux et al., 1997). Despite this general consistency, various ways in which the current results differ from previous similar studies may provide insight into what affects stress perception, especially on ABX tasks specifically.

4.1 Variability and Performance

The large amount of variability among participants' performance in this study is striking. Neither Dupoux et al. (1997) nor Carpenter (2015) addressed variability in accuracy in the ABX task performance for French-L1 speakers. However, it seems that their data may have been less variable, and the average performance was more accurate. In Dupoux and colleagues' (1997) experiments 1 and 3, mean error for a stress contrast in trisyllabic pseudowords was 19% and 10.8%, respectively. In Carpenter's (2015) experiment that reproduced Dupoux and colleagues' experiment 1 stimuli, mean error for stress contrast in the same trisyllabic pseudowords was 25% (30% for untrained participants and 19% for trained participants). In the present study, mean error on the stress contrast was 37%. Carpenter did not test any contrasts not based on stress. Dupoux and colleagues tested a consonant phoneme contrast in experiment 3, with French speakers performing with only 2.7% error. This stands in stark contrast to the 27.59% error in the voicing contrast in the present experiment. However, Dupoux and colleague's contrast included consonant pairs that differed both in place and manner of articulation like /f/ and /l/. These more different sounds should thus be easier to differentiate than /s/ and /z/, which differ only in voicing, and thus could explain some of this large difference.

There are several major differences between the present study and the studies of Dupoux et al. (1997) and Carpenter (2015) that could together account for these differences. Firstly, the present study had only 24 trials per participant for the stress contrast and 16 trials per participant for the voicing contrast, compared to 96 trials for each contrast type in Dupoux et al. and Carpenter's studies. The fewer number of trials partly explains higher variance in participants' individual averages on each contrast type. Secondly, both Dupoux et al. and Carpenter used ten practice trials with feedback on if the response was correct or incorrect, and items in which incorrect responses were given were repeated until the correct response was given. This constitutes a sort of training on the tasks and on perception of the contrasts and thus likely accounts partly for reduced error.

Thirdly, the present study uses real English words as stimuli, while Dupoux et al. (1997) and Carpenter (2015) used pseudowords that mimic Spanish in their stress patterns and lack of vowel reduction. This is a more complicated difference, as the impact of using real words in a language that the participants have studied is hard to quantify; there could be effects of familiarity and learning. Real Spanish words were used in the study by Schwab and Dellwo (2022), which found evidence of stress deafness, but their French-L1 participants had never studied Spanish, Italian, or Portuguese. Furthermore, the accuracy rates in their experiment are not directly comparable with those of the present study or with those of Carpenter (2015) and Dupoux et al. (1997) because Schwab and Dellwo did not use an ABX task and instead used odd-one-out and stressed syllable identification tasks in which chance level was 33% (rather than 50% as in an ABX task). As compared to Spanish words, the presence of drastic vowel reduction in English words should supply an additional cue to stress differences, with segmental differences in vowel quality, identifiable by the vowel formants. This additional cue should facilitate differentiation between words. However, it is not clear from this study's results if vowel reduction does facilitate the ABX task. For example, the first vowel in "ex'tract" was not always reduced by the female speaker who made the X stimuli. Token f63 of the female speaker includes less vowel reduction, but the items containing this token do not have lower accuracy rates.

Fourthly, the intonational pattern of the male (A and B token) voice was falling while that of the female voice (X token) was rising. This contrast in intonation, as it increases phonetic variability, likely introduced further difficulty compared to Dupoux et al. (1997) and Carpenter (2015), which had no such intonational variation. In the experiment by Schwab and Dellwo (2022), intonation was shown to affect stressed syllable identification, and the use of various intonation patterns led to lower performance on odd-one-out tasks as compared to odd-one-out tasks with only one type of intonation. As an increase in fundamental frequency is a phonetic correlate of stress, the variation in intonation can blur this cue and thus render lexical stress patterns more difficult to perceive. If French-L1 listeners do indeed rely more on phonetic rather than phonological means for identification of stress, it can then be expected that such variation in intonation would harm their discrimination ability based on stress placement.

Lastly, the ABX tasks of Dupoux et al. (1997) and Carpenter (2015) were conducted in person, while the present study was conducted online. Participants completing a task in person are more easily able to ask questions and receive guidance, and the added supervision by another person encourages them to stay focused and on-task. Experimenters supervising in person have more control over the experimental environment, such as background sounds and other distractions. Such differences in experiment setting have been shown to lead to differences in experiment outcomes in other auditory perception tasks (e.g. Pfeifer & Hamann, 2015). The online setting thus could also account for more variance. The potential undesirable effects of the present research being online were partly mitigated by the considerable length of the experiment (35 minutes), lack of monetary or material incentive, and regular questions and activities that required attention because these meant that participants included in the analysis were at least focused and determined enough to complete the entire experiment, rather than leaving partway through. However, the presence of background noises, distractions, variations in equipment, and other factors were not able to be controlled.

4.2 Unexpected Results of Voicing Contrast by Training

Upon post-hoc observation of the data, all participants who performed worse on the voicing contrast than on the stress and vowel contrasts had followed the stress-aware lesson, and, as can be seen in Figure 2, those in the stress-aware lesson generally performed worse on the voicing contrast than those in the stress-ignoring lesson. It is possible then that when attention was drawn to word stress, participants paid more attention to vowels, and thus performed better on word stress and vowel contrasts than on the voicing contrast, having diverted their attention from consonants. As all types of contrasts were presented in the same block, participants may have had difficulty switching between focusing on different characteristics. In previous experiments that included training (Carpenter, 2015; Schwab & Dellwo, 2022), participants only had to identify differences based on stress contrasts and not on any phonemic or other contrasts. Therefore, there is no previous literature that shows how having to identify differences based on multiple contrasts within the same task is affected by training. Dupoux and colleagues' (1997) experiment 3 did include stress contrasts, phoneme contrasts, and redundant stress-and-phoneme contrasts within the same ABX task. In this experiment, the French-L1 speakers performed best and very similarly on the phoneme contrasts and redundant stress-and-phoneme contrasts, showing a sort of default reliance on identification by (consonant) phoneme. This pattern is somewhat mirrored by the performance of those who followed the stress-unaware lesson and performed best in the voicing contrast.

Such a phenomenon of costs for switching focus between vowels and consonants or various types of contrasts could be analogous to language switching paradigms, which induce costs for bilinguals (e.g. Peeters, 2020). To explore this possible phenomenon, a post-hoc analysis was conducted to test for an interaction between training and contrast type. This was a glmer for accuracy with training and contrast as fixed predictors, and with random slopes and intercepts by participant for contrast and random slopes and intercepts by item for training. This model failed to converge, and as the random slopes are for the effect that answers the question, cannot be further simplified without losing validity. Therefore, generalizing from this small sample is not possible, and answering such a question would require further data.

4.3 Practical Application of Results

Though the present study does not add more to the argument that teaching about lexical stress can benefit French speaker's lexical stress perception, it does not provide any evidence against said teaching. In a classroom or pedagogical context, teachers can use those methods that were found to have an effect: Carpenter's (2015) short-term, perceptual fading technique training or longer-term explicit or game-like training as seen in the study by Schwab and Dellwo (2022). Such methods should be adapted for the specific pedagogical context, such as using real words in the target language. Because they are individualized methods delivered via computer, they can be offered to students as practice outside of the classroom; this, for example, would be particularly appropriate for the game-like implicit training method used by Schwab and Dellwo (2022). Additionally, the techniques and types of exercises used by Schwab and Dellwo in their explicit training could be adapted for teacher use in the classroom. In any case, a study in a more classroom-like context—repetitive, in-person, and longer-term—with a perception task would be useful and interesting to assess what would be best applied in schools generally.

4.4 Limitations

The online lesson, though meant to resemble what happens in the classroom, does of course have major differences. Being pre-recorded, there is no chance for specific corrective feedback and other teacher responses tailored to the needs of the student, which has been shown to improve perceptual accuracy in unfamiliar segmental distinctions (Lee & Lyster, 2016). Furthermore, without a larger, longer-term student-teacher context, there is no comparable opportunity for the building of rapport between teacher and student, which can help to boost motivation and learning (Bardorfer & Dolenc, 2022; Wilson et al., 2010).

Additionally, because the initial plan of recruiting through universities was not successful, participants were older and more spontaneously willing to volunteer themselves than what was anticipated. This may have led to the average LexTALE score (the proxy for proficiency) being higher than expected, and therefore, the intended English proficiency level of the audience for the lesson was likely lower than that of the actual participants. One participant, with a LexTALE score of 70 and in the stress-unaware lesson, noted that he found the lesson not to be helpful for the listening task because he had difficulty paying attention as it was "boring and very slow." This is the compounded danger of having the same, undifferentiated lesson for people of different proficiency levels and having a control group lesson that may seem (rightly) more superfluous.

5. Conclusion

In its current state, the present study is not able to add to the understanding of how pedagogical practices affect stress perception in French-L1 speakers. Nonetheless, the small amount of results still highlight aspects that should be considered in future similar studies, especially using ABX tasks, including the impact of attention to certain concepts related to pronunciation (such as lexical stress) on ABX task performance.

References

Abdelli-Beruh, N. B., Drugman, T., & Owl, R. R. (2016). Occurrence frequencies of acoustic patterns of vocal fry in American English speakers. *Journal of Voice*, *30*(6), 759.E11-759.E20. https://doi.org/10.1016/j.jvoice.2015.09.011

Andreassen, H. N., & Eychenne, J. (2013). The French foot revisited. *Language Sciences*, *39*, 126-140. https://doi.org/10.1016/j.langsci.2013.02.016

Avanzi, M., Schwab, S. & Dubosson, P. (2013). Prosodie et contact de langue : L'exemple du ‘français fédéral' [Prosody and language contact: The example of ‘federal French’]. *Revue française de linguistique appliquée, 18*(2), 77-90. https://doi.org/10.3917/rfla.182.0077.

Bardorfer, A., & Dolenc, P. (2022). Teacher-student rapport as predictor of learning motivation within higher education: The self-determination theory perspective. *Journal of Psychological & Educational Research*, *30*(2). https://www.marianjournals.com/files/JPER_articles/JPER_30_2_2022/Bardorfer_Dolenc_JPER_2022_30_2_115_133.pdf

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://10.18637/jss.v067.i01

Bijeljac-Babic, R., Serres, J., Höhle, B., & Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PloS One*, *7*(2), e30843. https://doi.org/10.1371/journal.pone.0030843

Boersma, Paul & Weenink, David (2024). Praat: Doing phonetics by computer [Computer program]. Version 6.4.19. https://www.fon.hum.uva.nl/praat/

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*, 991-997. https://doi.org/10.3758/s13428-012-0190-4

Capliez, M. (2011). Typologie des erreurs de production d’anglais des francophones : Segments vs. suprasegments [Typology of francophones’ English production errors: Segments vs.

suprasegments]. *Recherche et pratiques pédagogiques en langues*, *30*(3) 44-60. https://doi.org/10.4000/apliut.1645

Carpenter, A. C. (2015). Phonetic training significantly mitigates the stress 'deafness' of French speakers. *International Journal of Linguistics*, *7*(3), 94-108. https://doi.org/10.5296/ijl.v7i3.7661

Celik, B., & Yıldız, Y. (2019). The role of foreign language culture on teaching the language and learner motivation. *International Journal of Social Sciences & Educational Studies*, *5*(4), 150-161. https://doi.org/10.23918/ijsses.v5i4p150

Cours des Comptes (2024, October). La fonction ressources humaines au ministère de l'éducation nationale: Rapport public thématique [Human Resources Employment at the national Ministry of Education: Public Thematic Report]. https://www.ccomptes.fr/sites/default/files/2024-10/20241025-La-fonction-RH-au-ministere-de-leducation-nationale.pdf

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1568–1578. https://doi.org/10.1037/0096-1523.25.6.1568

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, *36*(3), 406-421. https://doi.org/10.1006/jmla.1996.2500

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness". *The Journal of the Acoustical Society of America*, *110*(3), 1606-1618. https://doi.org/10.1121/1.1380437

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress 'deafness' in simultaneous French–Spanish bilinguals. *Cognition*, *114*(2), 266–275. https://doi.org/10.1016/j.cognition.2009.10.001

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*(2), 682-706. https://doi.org/10.1016/j.cognition.2007.04.001

Frost, D. (2010). La surdité accentuelle : d'où vient-elle et comment la guérir ? [Stress deafness: Where does it come from and how can it be cured?] *Les cahiers de l'APLIUT*, *29*(2), 25-43. https://doi.org/10.4000/apliut.684

Frost, D., & Henderson, A. (2013). Résultats du sondage EPTiES (English Pronunciation Teaching in Europe Survey) : l'enseignement de la prononciation dans plusieurs pays européens vu par les enseignants [Results of the EPTiES survey (English Pronunciation Teaching in Europe Survey): Pronunciation teaching in various European countries as seen by teachers]. *Recherche et pratiques pédagogiques en langues, 32*(1), 92-113. https://doi.org/10.4000/apliut.3586

Giraudo, H., & Montermini, F. (2010). Primary stress assignment in Italian: Linguistic and experimental issues. *Lingue e Linguaggio*, *9*(2), 113–129. https://doi.org/10.1418/33323

Green, P., & MacLeod, C.J. (2016). simr: an R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493-498. https://doi.org/10.1111/2041-210X.12504

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*(3), 262-274. https://doi.org/10.1016/j.infbeh.2009.03.004

Kabak, B., Maniwa, K. & Kazanina, N. (2010). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology*, *1*(1), 207-224. https://doi.org/10.1515/labphon.2010.010

Ladefoged, P., & Disner, S.F. (2012). *Vowels and Consonants* (3rd ed.). Blackwell Publishing Ltd.

Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*(1), 35–64. https://www.doi.org/10.1017/S0272263115000194

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325-343. https://doi.org/10.3758/s13428-011-0146-0

MENJ, (2022). Les langues vivantes étrangères et régionales [Living Foreign and Regional Languages]. https://www.education.gouv.fr/les-langues-vivantes-etrangeres-et-regionales-11249

MENJ. (2022, December 12). Enseignement de l'anglais et des langues vivantes étrangères tout au long de la scolarité obligatoire: Mesures pour améliorer les apprentissages des élèves [Teaching of English and Living Foreign Languages throughout Compulsory Schooling: Measures to Improve Student Learning]. https://www.education.gouv.fr/bo/22/Hebdo47/MENE2234752C.htm

Ministère de l'enseignement supérieur, de la recherche et de l'innovation (2021). Licence professionnelle « Bachelor Universitaire de Technologie »: Techniques de commercialisation: Programme national [Professional Licence 'University Bachelor of Technology': Business and Marketing Techniques: National Curriculum]. https://cache.media.education.gouv.fr/file/SPE4-MESRI-17-6-2021/37/1/Annexe_25_TC_BUT_annee_1_1411371.pdf

Moore-Mauroux, S. (2010). Quelques outils méthodologiques pour aider l'étudiant francophone face à la prononciation de l'anglais [Several methodological tools to help the francophone student faced with English pronunciation]. *Les cahiers de l'APLIUT, 29*(3), 36-49. https://doi.org/10.4000/apliut.595

Onisep (n.d.). Les principales langues vivantes étrangères enseignées près de chez vous [The Main Foreign Living Languages Taught Near You]. https://www.onisep.fr/recherche?context=languages

Peeters, D. (2020). Bilingual switching between languages and listeners: Insights from immersive virtual reality. *Cognition*, *195*, 104107. https://doi.org/10.1016/j.cognition.2019.104107

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430. https://doi.org/10.1016/j.wocn.2010.04.001

Pfeifer, J., & Hamann, S. (2015). Revising the diagnosis of congenital amusia with the Montreal Battery of Evaluation of Amusia. *Frontiers in human neuroscience*, *9*, 161. https://doi.org/10.3389/fnhum.2015.00161

Polivanov, E. D. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4: 79-96. [English translation: The subjective nature of the perceptions of language sounds. In E.D. Polivanov (1974): *Selected works: articles on general linguistics*. The Hague: Mouton. 223-237]

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Schwab, S., & Dellwo, V. (2022). Explicit versus non-explicit prosodic training in the learning of Spanish L2 stress contrasts by French listeners. *Journal of Second Language Studies*, *5*(2), 266-306. https://doi.org/10.1075/jsls.21017.sch

Schwab, S., & Llisterri, J. (2014). Does training make French speakers more able to identify lexical stress? *COPAL, Concordia Working Papers in Applied Linguistics*, *5*, 624-636. https://web.archive.org/web/20180420214300id_/http://doe.concordia.ca/copal/documents/40_Schwab_Llisterri_Vol5.pdf

Sichel-Bazin, R. (2016). Le français du Midi : une variété de français à accent lexical [Southern French: A French variety with lexical stress]. *Langue française, 191*(3), 47-66. https://doi.org/10.3917/lf.191.0047.

Sippel, L. & Martin, I. (2023). Immediate and long-term improvement in lexical stress perception: the role of teacher and peer feedback. *International Review of Applied Linguistics in Language Teaching*, *61*(3), 1173-1195. https://doi.org/10.1515/iral-2021-0175

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*(4), 905–916. https://doi:10.1017/S1366728912000168

Turin-Bartier, M. 2022. Rapport du jury Concours : troisième concours du CAPES et du CAFEP-CAPES ; Section : langues vivantes étrangères : anglais ; session 2022. [Exam Jury Report: Third

Exam of the CAPES and CAFEP-CAPES; Section: Living Foreign Languages: English; 2022 Session]. https://www.devenirenseignant.gouv.fr/media/6467/download

Vaudin, M., 2024. Rapport du jury Concours : CAPES externe et CAFEP-CAPES ; Section : langues vivantes étrangères – Anglais ; Session 2024 [Exam Jury Report: External CAPES and CAFEP-CAPES; Section: Living Foreign Languages – English; 2024 Session]. https://www.devenirenseignant.gouv.fr/media/14701/download

Vet, D. (2024). *ED – Experiment Designer*. Universiteit van Amsterdam. https://www.fon.hum.uva.nl/dirk/ed.php

Walter, C. (2001). French speakers. In M. Swan & B. Smith (Eds.), *Learner English: A Teacher's Guide to Interference and Other Problems* (2nd ed., pp. 52–72). Cambridge University Press.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43). https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. https://CRAN.R-project.org/package=dplyr

Wilson, J.H., Ryan, R.G., & Pugh, J.L. (2010). Professor–student rapport scale predicts student outcomes. *Teaching of Psychology*, *37*(4), 246–251. https://doi.org/10.1080/00986283.2010.510976

Appendix A: Possible Variables and Explanations

The most maximal model would be:

accuracy ~ training * Xtype * pair * trialIndex * correct * response * LexTale * region * age + (Xtype * pair * correct * response * trialIndex | participant) + (training * trialIndex * LexTale * region * age | item)

The above maximal model is theoretically too complex; however, I believe no two independent predictors are inherently/necessarily correlated, as for example, response types (A or B) are evenly distributed between pairs and X types (extreme or ambiguous). However, the controlling via balancing and random order of trials is made less perfect by any trials excluded because no response was recorded (timed out).

Explanation of (possible) variables:

- accuracy: outcome (binary); *in raw results: testABX.Acc*
- participant: random predictor; *in raw results, subject*
- item: random predictor; *in raw results, testABX.Item*
- training: (RQ-answering) binary between-participant, within-item; 2 levels: stress (code +.5), nonStress (code -.5); *in raw results, respectively Version: 1, 2*
- Xtype: if X is a more extreme (prototypical) or ambiguous token, binary within-participant, between-item; 2 levels: ambiguous (-.5) or extreme (+.5), *in raw results, testABX.Xtype*
- pair: ternary within-participant, between-item; 3 levels: refund, extract, conduct; *in raw results, respectively, testABX.Pair: 1, 2, 3*
    - no immediately clear, well-founded reason to code any particular levels as +.5, 0, -.5
- trialIndex: within-participant, within-item, continuous; the order in which an item showed up during the task*, in raw results, testABX.index*
    - could include this in a model to control for order of variables, tiring as experiment continues, should be somewhat controlled for already through randomization
- correct: if in reality X = A or X = B, binary within-participant, between-item; *in raw results, testABX.Correct: A (-.5), B (+.5)*

- response: if the response given by participant is X = A or X = B, binary within-participant, between-item; *in raw results, testABX.Sel_ABX: A (-.5), B (+.5)*
- LexTale: score on LexTALE, proxy for proficiency, continuous; between-participant, within-item
- age: continuous; between-participant, within-item
- gender: binary, ternary, etc. depending on responses; between-participant, within-item
- regional origin: type depends on the range of origins of participants, could be binary, tertiary, etc.; between-participant, within-item
    - may be useful to include if there are participants from southern France

Appendix B: Pre-Questions

These questions were presented only in French; English translations by the author are provided below. Responses in bold excluded the person automatically from participation in the study.

1. Avez-vous une dyslexie ou un autre trouble lié au langage ? Do you have dyslexia or another language-related disorder ?	**oui** **yes**	non no			
2. Avez-vous une bonne vue et une bonne ouïe (ou une vue/ouïe suffisamment corrigée) ? Do you have good sight and hearing (or sufficiently corrected sight/hearing)?	oui yes	**non** **no**			
3. Le français est-il votre langue maternelle ? Is French your native language (mother tongue)?	oui yes	**non** **no**			
4. D'autres langues ont-elles été parlées chez vous pendant votre enfance (avant l'âge de 6 ans) ? Were other languages spoken where you grew up (before the age of 6)?	oui yes	non no			
5. Une de ces langues, a-t-elle été parlée plus de 25% du temps ? Was one of these languages spoken more than 25% of the time?	**oui** **yes**	non no			
6. Quelles langues avez-vous étudié à l'école ? (vous pouvez cocher plusieurs cases) Which languages did you study at school? (You can choose multiple answers)	anglais English	espagnol Spanish	allemand German	italien Italian	une autre langue another language
7. Dans quel département avez-vous grandi (passé la majorité vôtre enfance) ? In which department [official area of France] did you grow up (spend most of your childhood)?	01 - Ain	02 - Aisne	03 - Allier	04 - Alpes-de-Haute-Provence	05 - Hautes-Alpes …
8. Dans quel département habitez-vous actuellement ? In which department do you live now?	01 - Ain	02 - Aisne	03 - Allier	04 - Alpes-de-Haute-Provence	05 - Hautes-Alpes …
9. Étiez-vous ou êtes-vous spécialiste d'anglais ou d'une autre langue (par ex. LLCER, LEA) à l'université ? Were you or are you a specialist of English or another language at university (e.g. Applied	**oui** **yes**	non no			

Foreign Languages, or Foreign and Regional Languages, Literatures, and Civilizations)?					
10. Depuis combien d’années avez-vous votre baccalauréat ? For how many years have you had your baccalaureate [high school leaving exam]?	0	1	2	3	4 ou plus 4 or more
11. Quel âge avez-vous ? (veuillez utiliser les touches chiffres en haut du clavier) How old are you? (please use the number keys at the top of the keyboard)	[free numerical response]				
12. À quel genre vous identifiez-vous ? Which gender do you identify with?	femme woman	homme man	non-binaire non-binary	ne souhaite pas répondre do not wish to respond	autre other

Appendix C: Post-Questions

These questions were presented both in English and French, as shown below. Participants were free to respond in English or French.

1. Did you find the video lesson helpful for the listening activity? (Avez-vous trouvé que le cours en vidéo vous a permis de mieux appréhender l'exercice de compréhension orale ?)	yes (oui)	no (non)
2. Why did you find the video lesson helpful or not helpful? (Pourquoi avez-vous trouvé le cours en vidéo utile ou non?)	[free response]	
3. What do you think the experiment was about? (À votre avis, quel est le but de l'expérience ?)	[free response]	

Appendix D: Lesson Resources

Image credits used in stress-ignoring lesson in order of appearance and from top to bottom:

author's own photograph; floor plan courtesy of Kimberlee Jamieson; photos by Chastity Cortijo, Aaron de Sousa, Allef Vinicius, Stephen McFadden, Lisa Anna, Isaac, and Alex Tyson via Unsplash

Images in the stress-aware lesson are identical to those in the stress-ignoring lesson, except the images by Chastity Cortijo and Alex Tyson are not used.