Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Psychology

Vilde Sannerud Kalin

Does Native Language Experience with Vowel Length Transfer to Non-Native Audiovisual Vowel Perception?

Master's thesis in PSY3914 Psykologi – studieretning læring – hjerne, atferd, omgivelser
Supervisor: Dawn Behne
May 2022

NTNU
Norwegian University of
Science and Technology

Vilde Sannerud Kalin

Does Native Language Experience with Vowel Length Transfer to Non-Native Audiovisual Vowel Perception?

Master's thesis in PSY3914 Psykologi – studieretning læring – hjerne, atferd, omgivelser
Supervisor: Dawn Behne
May 2022

Norwegian University of Science and Technology
Faculty of Social and Educational Sciences
Department of Psychology

NTNU
Kunnskap for en bedre verden

Foreword

This Master's Thesis is an extension of a study conducted by Redmon and colleagues (2020). The experiment stimuli was developed by Leung et al. (2016), Tang et al. (2015) and Redmon et al. (2020), and the experiment was developed by Redmon et al. (2020). In the present study, a native Norwegian perceiver group was tested, and the data collection was done by me. Data results from an English perceiver group was made available by the authors of the Redmon et al. (2020) study, and was used as a comparison to the Norwegian perceiver group. The data analysis used for the present study was conducted by me. The aim of the study is based on the work of Redmon et al. (2020), but was extended and formulated by me.

Acknowledgements

First, and foremost, I must give a huge thank you to my supervisor, Dawn Behne. Never have I met an academic who cares as much for her students as she does. Thank you for being available at all hours, day and night, weekdays and weekends; for our weekly Wednesday discussions; for all support and kindness when I have struggled; for the detailed feedback on my first drafts; for creating a great environment for cooperation between us master students and you, despite the challenges of the pandemic. I appreciate that I have been lucky enough to have you as my supervisor both for my Bachelor- and Master thesis.

Thank you, my fellow master students, Alexander Sævild Ree and Ine Stensholm Elveland, for all help and support with my project, and especially for all the conversations and laughter during breaks.

Thank you, Yue Wang and Beverly Hannah at Simon Fraser University, for our discussions, answering my questions, and all the help you have provided with sharing your data results, guiding me for setting up the experiment, and the experiment procedure.

Thank you, Darren Rhodes, for taking the time to guide and help us with Mixed Models analysis.

Thank you, Tom Knudsen, for helping me with the installation of Bootcamp Assistant, and other technical issues.

Thank you to my family who always supports and believes in me, and to Ingrid Kalin Lauritsen for all support in times of need.

Table of Contents

Sammendrag

Denne studien er en utvidelse av forskning utført av Redmon og kolleger (2020), hvor personer med engelsk (Vest-Canadisk) og Mandarin som morsmål ble testet i deres forståelse av audiovisuelle (AV) tense/lax vokalkontraster i vanlig og klar tale. I denne studien ble personer med norsk som morsmål testet med de samme betingelsene, og deres resultater ble sammenlignet med resultatene til personene med engelsk som morsmål. Målet var å undersøke om bruk av hint på varighet i persepsjon av norske vokaler, ville overføres til annenspråks persepsjon av engelske AV tense (/i, ɑ, u/) og lax (/ɪ, ʌ, ʊ/) vokalpar produsert i vanlig og klar tale. En klar tale-fordel ble funnet for tense vokaler i «audio-only» (AO) og «video-only» (VO) for begge språkgrupper, men kun de engelske deltakerne viste en klar tale-fordel for tense vokaler i AV. Der de engelske deltakerne viste en klar tale-fordel for lax vokaler i AO, ble ikke dette funnet for de norske deltakerne. Begge språkgrupper viste ingen klar tale-fordel for lax vokaler i AV, mens en klar tale-ulempe ble funnet i VO. Resultatene viste et lignende mønster på tvers av språkgruppene, og presenterer implikasjoner på påvirkning fra et morsmål på annenspråks vokalpersepsjon, hvor erfaring med temporal timing fra en deltakers morsmål øker sensitivitet for varighetskontraster i et annetspråk.

Abstract

The present study is an extension of research conducted by Redmon and colleagues (2020), where native (Western Canadian English) and non-native (Mandarin) perceivers were tested, in their intelligibility of audiovisual (AV) English tense/lax vowel contrasts in plain and clear speech. In the present study, Norwegian perceivers were tested with the same conditions, and their results were compares to those of the native English. The aim was to investigate whether the use of duration cues in Norwegian vowel perception, would transfer to non-native perception of English AV tense (/i, ɑ, u/) and lax (/ɪ, ʌ, ʊ/) vowel pairs produced in plain- and clear speech styles. A clear speech advantage was found for tense vowels in audio-only (AO) and video-only (VO) with both perceiver groups, while only the English showed a clear speech advantage for tense vowels in AV. Where the English perceivers showed a clear speech advantage for lax vowels in AO, the Norwegian perceivers did not. For both perceiver groups, lax vowels in AV resulted in no clear speech advantage, while a clear speech disadvantage was found in VO. The results showed a similar pattern across perceiver groups, and present implications of native vowel system influence on non-native vowel perception, where experience with temporal timing from a perceivers' native language increases sensitivity to duration contrasts in a second language.

Does Native Language Experience with Vowel Length Transfer to Non-Native Audiovisual Vowel Perception?[1]

During communication in challenging listening situations, speech becomes articulatorily, and thereby acoustically, modified to increase the intelligibility for a perceiver. This clear speech style is produced with extreme acoustic features and hyperarticulation, which make speech more salient both auditorily and visually (Ferguson & Kewley-Port, 2007; Han et al., 2021; Lindblom, 1990; Searl & Evitts, 2013; Tang et al., 2015). Research have shown that both native and non-native perceivers find clear speech more intelligible than plain, conversational speech (Bradlow & Bent, 2002; Redmon et al., 2020; Smiljanić & Bradlow, 2011). Recent studies have addressed how English tense/lax vowel pairs are modified in clear speech, and has established both auditory (A) (Leung et al., 2016) and visual (V) (Tang et al., 2015) attributes that become more extreme in clear speech when compared to plain speech. A question arising in a recent study (Redmon et al., 2020), is whether native vowel systems influence the perception of visual and auditory cues in native and non-native vowel perception. In the study, native (Western Canadian English) and non-native (Mandarin) perceivers were tested in how intelligible they found audiovisual (AV) tense/lax vowel pairs produced in plain- and clear speech styles (Redmon et al., 2020).

Where Mandarin does not have vowel tensity, Norwegian distinguishes vowels by vowel length, which is similar to the English tense/lax vowel distinction. However, Norwegian vowel perception involves articulatory and acoustic features which corresponds to the distinction of attributes of plain versus clear speech in English. Therefore, Norwegian perceivers can further address how native vowel systems can influence non-native vowel perception. On this premise, the present study set forward to extend Redmon et al.'s study (2020), by testing native Norwegian perceivers´ intelligibility of English AV tense/lax vowel pairs in clear and plain speech, and comparing results to those from the native English.

Clear Versus Plain Speech

In challenging listening situations, such as noisy environments or when talking to someone with a hearing impairment, intelligibility will be more constrained for a listener (Lindblom, 1990). Such constraints on intelligibility during communication are the basis of the hyper- and hypoarticulation (H&H) theory, which suggests that a speaker will adjust the

[1] A preliminary paper was written by the author of this master thesis in the spring semester of 2021 in the subject PSY3114, about preparations for the current study: "Does Temporal Duration Experience from Native Language Vowel Distinction Transfer to Non-Native Audiovisual Vowel Perception: A Study Proposal"

visual articulation and audio signal of their speech based on a listener's requirements (Lindblom, 1990). According to Lindblom (1990), with few or no constraints for intelligibility for a listener, speech production can be effortless and more economic, resulting in hypoarticulated speech (i.e., conversational or plain speech). By contrast, in communicative contexts with more constraints of intelligibility for the listener, speech production requires more articulatory effort with the purpose to make speech more intelligible, resulting in hyperarticulated speech (i.e., clear speech) (Lindblom, 1990).

Two types of modifications separates clear speech from plain speech: signal-based and code-based modifications (Bradlow & Bent, 2002). Signal-based modifications serve to enhance the saliency of speech signals, such as speaking in a slower pace and increasing the use of longer pauses between words, or increasing the pitch range and the amplitude. Code-based modifications serve to increase the distance between phonological categories (Bradlow & Bent, 2002), which, for instance, can be done by increasing the difference between two vowels (Bradlow, 2002), making tense vowels longer than lax vowels (Leung et al., 2016) and, as much as possible, avoiding the vowel reduction which is common in hypoarticulated speech (Bradlow, 2002).

A clear speech style has been demonstrated to improve listener intelligibility when compared to a plain conversational speech style (Calandruccio et al., 2020; Han et al., 2021; Ferguson & Kewley-Port, 2007). Clear speech has also been found beneficial for the intelligibility of a listener when speech perception occurs in background noise (Calandruccio et al., 2020; Smiljanić & Gilbert, 2017), and in different perceiver groups, including hearing-impaired perceivers (Ferguson, 2012; Liu et al., 2004) and non-native perceivers (Bradlow & Bent, 2002; Redmon et al., 2020; Smiljanić & Bradlow, 2011). This basis gives reason to believe that the English tense/lax vowel distinction will become more intelligible for non-native perceivers when compares to plain speech.

Clear Speech and Vowel Tensity

In several language groups, including English, studies have demonstrated that vowels involve plain-to-clear speech modifications. These spectral modifications makes vowels longer in duration (Ferguson & Kewley-Port, 2007; Leung et al., 2016), and more dynamic (Ferguson & Kewley-Port, 2002, 2007), and increases the vowels intensity and fundamental frequency (f_0) which is associated with a greater mouth opening (Cooke & Lu, 2010; Kim & Davis, 2014). Vowels produced in clear speech also involves modifications of visible articulatory features which involves greater jaw- and lip movements (Kim & Davis, 2014; Tang et al., 2015).

In recent research, English tense and lax vowels produced in plain- and clear speech has been analyzed. English distinguishes between tense and lax vowels, where tense vowels differ from lax vowels in their longer duration and extreme spectral features during speech production (Hillenbrand et al., 1995). These same features are what makes vowels produced in clear speech distinct from vowels produced in plain speech (Leung et al., 2016). Therefore, Leung et al. (2016) analyzed English tense/lax vowel pairs produced in clear and plain speech styles, and found an interconnection of vowel tensity and clear speech where tense and lax vowel-distinctiveness was maintained through temporal modifications. Both tense and lax vowels were lengthened in clear relative to plain speech, but the tense vowels increased more in duration in clear speech than the lax vowels. Additionally, the spectral properties of the lax vowels increased more in clear speech than of the tense vowels, because of the peripheral spectral properties of tense vowel production, which makes bigger spectral changes unachievable. Thus, lax vowel production in clear speech involves spectral modifications to achieve an enhancement of intelligibility for the listener, to maintain a shorter duration than for the tense vowels to retain their distinctiveness (Leung et al., 2016).

Studies have also demonstrated a difference in visible articulatory features of vowels produced in clear versus plain speech (Kim & Davis, 2014; Tang et al., 2015). Tang et al. (2015) investigated lip- and jaw movements during the articulation of six English tense (/i, ɑ, u/) and lax (/ɪ, ʌ, ʊ/) vowels in plain and clear speech styles. They found that vowels produced in clear speech involved an increase in lip- and jaw movement when compared to the vowels produced in plain speech. Further, tense vowel production involved a greater lip- and jaw displacement than lax vowel production did (Tang et al., 2015). That being so, production of English tense and lax vowels in clear speech entail both acoustic and visual modifications.

As a follow-up study to Leung et al. (2016) and Tang et al. (2015), Redmon et al. investigated the intelligibility for perceivers of English AV vowel tensity in plain and clear speech styles (2020). Intelligibility of tense/lax vowel pairs in clear- and plain speech styles presented in audio only, audiovideo (AV) and video only (VO), was tested with two perceiver groups: native Western Canadian English speakers who have vowel tensity in their first language (L1), and native Mandarin speakers, who’s L1 does not have lax vowels. The results showed a clear speech benefit for tense vowels for both perceiver groups, when compared to tense vowels in plain speech, in all three modalities. Because of the increase in duration (Leung et al., 2016) and more extreme visual attributes (Tang et al., 2015) of tense vowels in clear speech, tense vowels became more intelligible for the perceivers than when produced in plain speech (Redmon et al., 2020). For lax vowels, the native English perceivers showed a clear speech benefit in AO, while in VO and AV they showed no clear speech benefit. With the non-native Mandarin perceivers, no advantage for clear speech was observed for lax vowels in neither of the modalities (Redmon et al., 2020). These results were explained by a decrease in tense/lax vowel contrasts when visual stimuli were available, because of the lax vowels' increase in duration and stronger articulation when produced in clear speech (Hillenbrand et al., 1995). As a result, lax vowels were sometimes mistaken for tense vowels in both perceiver groups, but even more so by the Mandarin perceivers because of inexperience with lax vowels from their L1 (Redmon et al., 2020).

The modality speech stimuli are presented in to non-native perceivers, influence their perception. As face-to-face communication involves information from both acoustic and visual features, research has found an audiovisual benefit in non-native speech perception over audio-only and video-only. The study also found that speech produced in VO was less intelligible for the perceivers than in AO and AV (Hazan et al., 2006). Furthermore, a study found that non-native perceivers attended more to visual information in audiovisual speech perception than native perceivers did (Hazan et al., 2010). On that account, a central issue is how visual and auditory cues from vowel distinction in a second language (L2) are used by non-native perceivers.

The visual and auditory cues of vowel duration are not available at the exact same time. In a study by Gracco and Löfqvist (1994), productions of the vowels /ae/, /U/, /I/, and /ai/ were carried out, and the time of jaw lowering- and closing, lip opening- and closing, as well as glottal opening- and closing was measured. They found that jaw lowering and lip opening preceded glottal opening, that is, visible articulatory movements preceded the sound production. Furthermore, glottal closing preceded jaw- and lip closing, meaning that visible information from the mouth and lips lasted longer than the sound production (Gracco & Löfqvist, 1994). The difference in timing of visual and auditory cues might therefore lead to visual cues of duration being less precise than auditory cues.

Temporal Experience with Vowel Length

Where the English vowel system uses spectral features and duration to distinguish tense and lax vowels (Hillenbrand et al., 1995), the Norwegian vowel system distinguishes long and short vowels by the same acoustic characteristics (Behne et al., 1996). However, the distinction of English tense and lax vowels and Norwegian short and long vowels differ.

English vowels are distinct in their phonological resonance, also referred to as vowel quality, and vowel length, however, the spectral properties of English tense vowels are the root to their lengthened duration when compared to lax vowels (Hillenbrand et al., 1995). During speech production of tense vowels, e.g., *keyed* [kiːd], the articulators use a longer time to carry out more extreme articulations than during production of lax vowels, e.g., *kid* [kɪd]. These articulatory properties of tense vowels generally makes their duration longer than lax vowels (Hillenbrand et al., 1995). During perception of English vowels, native English perceivers rely mostly on spectral ques, while duration is a secondary cue (Hillenbrand et al., 2000).

By contrast, Norwegian distinguishes vowels by vowel quality, as in English, but they are also distinct in their vowel quantity, which refers to variations of duration length of vowels with similar quality (Behne et al., 1996). For example, the Norwegian word *hat* [ha:t] “hate”, has a phonologically longer vowel than the vowel in *hatt* [hat:] “hat”. Norwegian consonants are also affected by the length of the preceding vowel, where a long vowel has a shorter subsequent consonant, and a short vowel has a longer subsequent consonant (Behne et al., 1996). When a vowel precedes a voiced consonant (i.e., postvocalic voicing), the vowel is longer in duration than when it precedes a voiceless consonant, and a voiced postvocalic consonant is longer in duration than a voiceless postvocalic consonant. Additionally, when a postvocalic consonant is voiced, the prevocalic consonant is shorter than when the postvocalic consonant is voiceless (Behne & Moxness, 1994). Vowel duration in Norwegian also affects the prevocalic consonant, where a short vowel has a preceding longer consonant, and a long vowel has a preceding shorter consonant (Behne & Moxness, 1994). Taken together, Norwegian vowels and consonants affect each other’s duration during speech production.

In Table 1, the four Canadian English tense/lax vowel pairs (Clopper et al., 2005), and the nine Norwegian short and long vowel pairs (Kristoffersen, 2000, p. 13), are presented. As the Norwegian vowel inventory consists of more short and long vowel pairs than the amount of Canadian English tense and lax vowel pairs, native Norwegian speakers acquire an experience with duration contrasts from several vowel sounds from their L1. As these durational vowel contrasts in Norwegian speech are what Norwegian perceivers attend to, together with vowel quality, during auditory vowel perception (Behne & Nylund, 2004), duration differences from short and long vowels, prevocalic consonants, and voiced and voiceless postvocalic consonants, might result in a timing experience from Norwegian.

Table 1

Canadian English tense and lax vowel pairs and Norwegian short and long vowel pairs

Canadian English tense/lax vowel pairs		Norwegian short and long vowel pairs	
Lax	Tense	Short	Long
ɪ	i	ɪ	iː
ʊ	u	ʏ	yː
ɛ	eɪ	ʉ	ʉː
ʌ	ɑ	ʊ	uː
		ɛ	eː
		œ	øː
		ɔ	oː
		æ	æː
		ɑ	ɑː

Temporal Timing Experience from a Native Language

Studies have found that experience with temporal timing derived from a perceiver's L1, can affect non-native speech perception (Kondaurova & Francis, 2009; Redmon et al., 2020). Furthermore, the speech cues used during vowel perception in a L2 can be influenced by the native vowel system of a perceiver (Flege et al., 2004; Flege et al., 1997; Kondaurova & Francis, 2009; Lengeris, 2009). Flege et al. investigated L2 vowel perception of similar Canadian English vowel pairs in participants with Italian as their L1, and compared them to participants with Canadian English as their L1 (2004). The Italian participants scored lower in vowel identification than the English participants, as Italian have fewer vowels than English. This made the Italian participants perceive similar vowels as the same vowel, which indicated L1 vowel system influence on L2 vowel perception (Flege et al., 2004).

Several research findings support that perceivers with vowel systems with duration differences in their L1, will use duration to distinguish vowels in a L2 which has vowel duration contrasts (e.g., Behne & Nylund, 2004; Ylinen et al., 2005;). However, studies have also found that duration is used to perceive vowels in a L2, although the perceivers do not have vowel duration in their L1 (e.g., Bohn & Flege, 1990; Cebrian, 2006; Kondaurova & Francis, 2009). The desensitization hypothesis (Bohn, 1995, pp. 294-295) suggests that when a perceiver do not have sufficient information through spectral differences to distinguish vowels in a L2, they will use duration instead, as they lack enough experience with the spectral differences in the given L2. This hypothesis points in the direction of duration being a speech cue that is available to all perceivers, regardless of their native vowel systems.

However, research on timing experience have demonstrated an improved sensitivity to timing differences. A study on audiovisual asynchronous speech perception demonstrated an increase in sensitivity for timing differences when perceivers had temporal experience from their native language (Behne & Wang, 2018). Further, it has been established that duration experience from music background makes perceivers more sensitive to duration differences in non-native speech (Behne et al., 2013; Chobert et al., 2014; Cooper & Wang, 2009; Cooper et al., 2016). These findings demonstrate that temporal timing experience makes the perceiver more sensitive to duration contrasts during speech perception.

In Redmon et al.'s study, the native Mandarin perceivers who do not have lax vowels in their native language, did not show a clear speech benefit for lax vowels in any of the modalities, especially not when visual information was available (2020). Further, an analysis on tensity specifically showed that the Mandarin group made most errors based on tensity misperception, where lax vowels were mistaken for tense vowels in clear speech. The authors suggested that the Mandarin perceivers distinguished lax vowels from tense vowels, based on the same perceptual cues that make vowels in plain and clear speech distinct, i.e. vowel duration and vowel spectrum (Redmon et al., 2020). In line with the Desensitization theory (Bohn, 1995, pp. 294-295), the Mandarin perceivers might have used vowel duration to identify the tense and lax vowels, although they do not have vowel tensity in their L1. If this was the case, in the context of AV vowel presentation in clear speech, duration was not a favorably cue for identifying lax vowels in clear speech. However, the Mandarin perceivers' lack of experience with vowel duration differences might also have contributed to the misperception of lax vowels as tense vowels.

Given the evidence that a perceiver's native language influences non-native speech perception, an assumption is that native Norwegians' use of duration cues is addition to spectral features to distinguish vowels in Norwegian (Behne & Nylund, 2004), might transfer to vowel perception of non-native languages with duration contrasts (i.e., English). One study found that native Norwegians used both duration and formant movement to recognize English vowels (Iverson & Evans, 2007), which implies the possibility of Norwegian perceivers' in the present study will do the same.

Although the native Mandarin perceivers in Redmon et al. (2020) seemed to use duration to distinguish tense and lax vowels, they did not have the experience with vowel duration from their L1, that Norwegian perceivers do. Norwegian might therefore contribute to an increased sensitivity to vowel length in non-native vowel perception.

The Present Study

A clear speech benefit has been observed in speech perception by non-native perceivers (Bradlow & Bent, 2002; Redmon et al., 2020; Smiljanić & Bradlow, 2011), and the native language of a perceiver has been found to influence non-native speech perception (Flege et al., 2004; Flege et al., 1997; Kondaurova & Francis, 2009; Lengeris, 2009). This brings into question whether perceivers will be influenced by their L1 when perceiving the acoustic and visual articulatory attributes of non-native clear speech. As Redmon et al. (2020) included these research topics into one study, they hypothesized that the intelligibility of the interconnection of AV tense/lax vowels in clear speech could depend on the perceivers' native language. This hypothesis was tested with native perceivers (Western Canadian English) who have vowel tensity in their L1, and non-native (Mandarin) perceivers who do not have experience with vowel tensity in their L1 (Redmon et al., 2020). However, non-native speech perception of the English tense/lax vowel distinction in clear speech, by perceivers who have experience with vowel length from their native language, remains to be investigated.

A testing ground for native vowel system influence on audiovisual non-native vowel perception, are English tense and lax vowels produced in clear speech, where English tense and lax vowels differ in spectral features and duration (Leung et al., 2016), and vowels produced in clear and plain speech are distinct by the same features (Hillenbrand et al., 1995). Norwegian also contrasts short and long vowels by these same acoustic attributes. However, native English perceivers primarily attend to spectral cues in vowel perception (Hillenbrand et al., 2000), while native Norwegian perceivers attend to vowel length in addition to spectral features (Behne & Nylund, 2004). Additionally, Norwegian adjacent vowels and consonants affect each other's duration (Behne & Moxness, 1994). As timing experience can result in an increased sensitivity to duration differences in non-native speech perception (Behne et al., 2013; Behne & Wang, 2018; Chobert et al., 2014; Cooper & Wang, 2009; Cooper et al., 2016), and L1 vowel systems have been found to affect L2 vowel perception (Flege et al., 1997; Kondaurova & Francis, 2009: Lengeris, 2009), experience with Norwegian short and long vowels, may influence native Norwegian perceivers' sensitivity to English tense and lax vowels in clear speech. Therefore, an investigation of Norwegian perceivers' intelligibility of AV tense/lax vowel contrasts in clear speech, can extend the findings in Redmon et al.'s study (2020).

On this premise, the present study set forward to extend Redmon et al.'s study (2020), by testing native Norwegian perceivers´ intelligibility of English AV tense/lax vowel pairs in clear and plain speech, and comparing results to those from the native English. The aim is to investigate whether the use of duration cues in Norwegian vowel perception, will transfer to non-native perception of English AV tense/lax vowel pairs produced in plain- and clear speech styles.

The Norwegian perceivers' results are expected to show a lower response accuracy in all conditions than of the English, as the speech stimuli are in the English perceivers' native language.

Tense vowels produced in clear speech involve more extreme spectral features and longer duration (Leung et al., 2016), as well as an increase in lip- and jaw movement (Tang et al., 2015), when compared to plain speech. As tense vowels involve modifications in clear speech that enhance the properties that are present in plain speech, the Norwegian perceivers are expected to show a clear speech advantage for tense vowels in all modalities (AO, AV, VO), as was observed for the English perceivers (Redmon et al., 2020). However, the clear speech advantage for tense vowels in VO is expected to be smaller than when auditory information is present in AO and AV. As the auditory cues of vowel duration will not be available, the visual duration cues might be less precise than auditory cues due to lip-and jaw movements before and after voicing (Gracco & Löfqvist, 1994).

As lax vowels produced in clear speech involve a longer duration and an increase in spectral features (Leung et al., 2016), and bigger jaw- and lip movements (Tang et al.,2015), they become more similar to their tense vowel counterparts. As a result, the Norwegian perceivers are expected to misinterpret some lax vowels as tense vowels in all modalities, which will decrease the response accuracy in clear speech. They are still expected to show a clear speech advantage for lax vowels when auditory information is present in AO, like the English perceivers showed (Redmon et al., 2020), as they will have access to both spectral and duration cues from the acoustic information. This assumption is based on the Norwegian perceivers' experience with vowel distinction from vowel quality and quantity from their L1 (Behne & Nylund, 2004). In AV, the English perceivers showed only a slightly higher response accuracy for lax vowels in clear speech than in plain speech, because of the multimodal information from audio and video (Redmon et al., 2020). As the Norwegian perceivers are expected to use duration cues, which are available from both acoustic and visual features, they might be more sensitive to duration contrasts in clear speech and show a clear speech advantage for lax vowels in AV. Additionally, an audiovisual benefit has been found for non-native speech perception over unimodal speech perception (e.g., Hazan et al., 2006). However, as a result of the expected tensity misinterpretation of lax vowels as tense vowels, as was found for the English perceivers (Redmon et al., 2020), the Norwegian perceivers are expected to show a smaller clear speech advantage for lax vowels than tense vowels in AO and AV. In VO however, the Norwegian perceivers are expected to show no clear speech advantage, because of the lack of auditory cues of duration, as was observed with the English perceivers. In this modality, a bigger degree of lax vowels in clear speech is expected to be misinterpreted as tense vowels, because of a smaller contrast in tensity differences (Hillenbrand et al., 1995), as well as the possibly less precise visual cues of duration (Gracco & Löfqvist, 1994). On this basis, the response pattern of the Norwegian and English perceivers is expected to be quite similar in all conditions.

Method

Design

The present study is an extension of Redmon et al.'s (2020) study conducted at Simon Fraser University (SFU) in Canada. In their study, native speakers of Western Canadian English were tested (Redmon et al., 2020) in a speech intelligibility experiment, developed by Tang et al. (2015), Leung et al. (2016), and Redmon et al. (2020). The experiment tested the English participants in their intelligibility of English AV tense/lax vowel-pairs in plain and clear speech styles (Redmon et al., 2020). In the present study, native speakers of Norwegian were tested in the same experiment, to investigate whether the use of duration to distinguish vowels in perception of Norwegian transfers to vowel perception of AV English tense and lax vowels, and to compare the collected data to the English perceivers' data from Redmon et al. (2020). The Method section will describe the data collection of the Norwegian perceiver data conducted at NTNU, and compared to the English perceiver data collection conducted at SFU (Redmon et al., 2020), as well as the development of the experiment stimuli.

Participants

The number of native Norwegian participants was motivated by the sample size of the

English participants tested by Redmon et al., (2020). 25 native speakers of Norwegian (19 female, 5 male, 1 other) in the age range of 19-27 yrs. (*M* = 21 yrs.) participated in the study. They were recruited from the Norwegian University of Science and Technology (NTNU) in Trondheim in Norway with the use of the safe digital questionnaire Nettskjema (University of Oslo, n.d.). Since children in the Norwegian school system are taught English from second grade through their first or second year in high school, and are exposed to the English language from childhood through digital media, they receive English training frequently. In the EF English Proficiency Index (EF EPI) of 2020, Norway was ranked as the fifth best country of the included countries in English proficiency, based on reading and listening skills (Education First, 2020, pp. 6-7). The population of the EF EPI 2020 were adults, where 79% were under the age of 35 (2020, p. 44), and the population of the present study were young adults in their twenties with Norwegian as their native language. Consequently, a possibility was that Norwegian participants would be very proficient in English, and therefore would use qualitative contrasts to distinguish English vowels in a bigger degree than they would during Norwegian vowel perception.

As the present study will investigate the impact of vowel duration as the leading que for intelligibility of English tense and lax vowels in clear speech, the Norwegian participants were recruited based of inclusion criteria that could decrease their English proficiency (see Appendix A). Therefore, the native Norwegian perceivers had started their English training no earlier than in first grade and grown up in a Norwegian speaking home. The participants had not lived in an English speaking country, and followed a student program with little English lectures or curriculum. All included participants had reported normal or adjusted vision and normal hearing.

The participants were asked if they had any neurological history which could affect vision, hearing or attention, but this was not an exclusion criteria. While handedness was tested with a variation of the Edinburgh Handedness Inventory (Oldfield, 1971), this information was collected, but not an inclusion criteria. The participants' comprehension of the stimulus words was tested to establish that they understood the word meanings, by connecting the word definition to the right word in a table. If any word definitions were connected to the wrong words, this was brought to their attention. The participants gave a written consent in order to participate (see Appendix B), and the study was registered by the Norwegian Centre for Research Data (NSD) (see Appendix C for approval of the study given by NSD).

The English participants were 21 (19 female, 2 male) speakers of Western Canadian

English in the age range of 19-27 yrs. (*M* = 22 yrs.), and were recruited at Simon Fraser University in Canada (Redmon et al., 2020). They reported normal or corrected vision, normal hearing and had no history of speech or language disorders (2020).

Materials

The production and editing of the stimuli presented in the experiment in the current study were done by Tang et al. (2015), Leung et al. (2016), and Redmon et al. (2020). The stimuli included three tense/lax vowel pairs implemented in six /kVd/ words (see Table 2) produced in plain and clear speech styles which were presented in AO, VO and AV. The tense/lax vowel pair /ɑ-ʌ/, has in earlier research been established as equivalents in the temporal (Gopal, 1990) and spectral (Clopper et al., 2005) domains in Western Canadian English. The vowel pairs /i-ɪ/ and /u-ʊ/ are also distinguishable by the same features, and the vowel pairs have in common that the production places of the lax vowels /ʌ/, /ɪ/ and /ʊ/ are more centralized in the oral cavity than the tense vowels /ɑ/, /i/, and /u/ (Leung et al., 2016). The consonant context /kVd/ was chosen because the duration and formant movement of the vowel is maintained during coarticulation with these consonants (Moon and Lindblom, 1994, as cited in Leung et al., 2016).

Table 2

Stimulus Vowel Features

Vowel	Word	Tensity	Feature
/i/	keyed	Tense	Lip spreading
/ɪ/	kid	Lax	Lip spreading
/ɑ/	cod	Tense	Jaw lowering
/ʌ/	cud	Lax	Jaw lowering
/u/	cooed	Tense	Lip rounding
/ʊ/	could	Lax	Lip rounding

Talkers

Six young adult speakers of Western Canadian English, 3 female and 3 male in the age range of 17-30 yrs., were recorded to make the audiovisual stimuli (Redmon et al., 2020). The talkers were recruited from Simon Fraser University in Canada, and reported no speech-or language impairment history. These particular talkers were chosen based on an articulatory analysis done by Tang et al., that established that these talkers' plain versus clear speech productions had the most distinct visible articulatory features (2015). In the talkers' Western Canadian English dialect, the production of the stimulus word "cod" involves the target vowel /ɑ/ (Redmon et al., 2020) as a result of the /ɑ/ and /ɔ/ merger (Clopper et al., 2005).

Stimuli Recording and Editing

Audio-video recordings were done in the Language and Brain Lab at Simon Fraser University in a sound attenuated room. Video recordings were taken of the talkers' faces during production of the target words at a 29 fps recording rate with a Canon Vixia HF30 camera. Simultaneously with the video recordings, audio recordings were done of the audio signal of the talkers' production of the target words. Audio recordings were done with a Shure KSM microphone with a 45-degree angle, and a 20 cm distance, relative to the talkers' mouths, and a Sonic Foundry Sound Forge 6.4 with a sampling rate sat to 48 kHz. The elicitation of the audiovisual plain and clear stimuli was done with a previously established simulated interactive computer program, that the talkers were told would recognize the words they produced (Leung et al., 2016; Maniwa et al., 2009; Tang et al., 2015). During recording, the talkers sat in front of a computer screen, on which one of the six English words was presented. To elicit plain speech, the talker was told to say the word naturally. The program would then appear to guess which word had been spoken, and any incorrect guesses were followed by an instruction to say the word more clearly, which resulted in the clear speech productions. The program prompted talkers to produce each word three times in a random order, which resulted in three speech productions of each word in both clear- and plain speech style.

The recordings were edited by Redmon et al., (2020) into three modalities: audio-only (AO), audio-video (AV) and video-only (VO). Audacity 2.1. was used to edit the recordings for each word into two-second AO clips, from the audio recordings from the microphone. Adobe Premier Pro CC 2014 was used to make the AV stimuli by the audio recordings from the microphone together with the video recordings, and to remove the audio recordings from the video recordings to make the VO stimuli. The video clips were edited to last for four seconds and included both mouth opening and closing. The audio in the AO and AV stimuli were normalized to 60 dB, and three types of cafeteria noise of 75 dB were laid on top of the audio clips (i.e., 15 dB signal-to-noise ratio), to ensure enough errors for the plain and clear speech productions to be compared.

Procedure

Access to the experiment was provided by the authors of Redmon et al., (2020). The native Norwegian participants were tested in the Speech Lab at the Department of Psychology at NTNU Dragvoll in Trondheim on two different days. The participants were given the option to consent to partake in an audio recording, providing the possibility to do an audio analyses later on, and if not consented to they could still participate in the experiment. During the audio recordings, the participants sat in front of a microphone in a sound attenuated room, and were told to say the six stimulus words "keyed, kid, cod, cud, cooed, could" repeatedly in a monotone voice until asked to continue to the next word.

At SFU, three participants were tested in a sound attenuated room at a time. At NTNU, up to two participants were tested at a time in a quiet laboratory room. Since the testing room at NTNU was not sound attenuated, noise from outside the testing room occurred occasionally during testing of seven of the participants, and was therefore logged with reference to the possibly affected trials. The participants sat in a chair in front of a computer screen with a 70 cm viewing distance. At SFU, the visual stimuli were presented on Microsoft monitors with a 15" screen, while at NTNU, 27" iMac monitors were used to present the visual stimuli. The size of the stimulus window and instruction boxes were modified at NTNU to match the sizes presented at SFU on smaller monitors, with a ~12" x ~7" (width by height) stimulus window size and 15 x 22 cm head size. A screen resolution of 1024 x 768 was used for the monitors in both labs. The brightness of the computer screens at SFU was specified as a non-distracting, comfortable brightness, while at NTNU, the brightness of the iMac screens was set to 80% as this was evaluated as the most comfortable. AKG K141 Studio headphones were used to present the audio signal at SFU, and AKG K271 stereo headphones were used at NTNU. The two types of headphones were compared by their sensitivity, impedance and audio frequency bandwidth to ensure the same sound quality for both participant groups. The computer volume was set to a listening level that was reported as comfortable at SFU (Redmon et al., 2020), and to a 45% volume at NTNU as this was evaluated as the most comfortable. A computer mouse was used by the participants to respond during testing in both labs.

The experiment was developed in Paradigm (Perception Research Systems, 2007) by Redmon et al. (2020), and Paradigm Player (Perception Research Systems, 2007) was used to run the experiment under participant testing. At NTNU, Microsoft 10 was used on the iMacs through Boot Camp Assistant, in order to use Paradigm Player for experiment testing. The experiment was divided into two parts, each part containing one block for each modality (AO, AV, VO), where testing occurred on two different days. Each block had 108 trials, which resulted in 648 trials for the whole experiment. There were six English stimulus words ("keyed, kid, cod, cud, cooed, could") presented in café noise (SNR: -15 dB), which consist of three tense/lax vowel pairs: /i-ɪ/, /ɑ-ʌ/ and /u-ʊ/ (see Table 2). The vowel pairs differ in articulatory features, where /i-ɪ/ have lip stretching, /ɑ-ʌ/ have jaw lowering and /u-ʊ/ have lip rounding.

Before testing, there was a familiarization session, where the six target words were presented auditorily without background noise. Then, each modality was presented with two trials in each to let the participants become familiar with the task. Each part of the experiment which were carried out on different days, presented stimuli in three blocks, one for each modality (AO, VO, AV). A block consisted of 108 trials, where each talkers' production of the stimuli words in plain and clear speech were presented three times in a randomized order. The blocks' order was counter-balanced across the participants, and the presentation order of the stimuli trials was randomized.

In each trial, the stimulus was presented, and the participant's task was to indicate which word was perceived by clicking one of the six word options ("keyed, kid, cod, cud, cooed, could") on the screen with a computer mouse, with up to four seconds to respond. Each part of the experiment lasted for 60-80 minutes, which included instructions, familiarization and the main testing session with breaks.

Results

The output files from Paradigm for each participant on each of the two testing days were consolidated as a datafile using Microsoft Excel 16.56. For each trial in the datafile, Experiment Version, Vowel Tensity (tense and lax), and Day (day 1 or day 2 of testing) were specified.

Potential outliers were identified by calculating the percentage of correct responses out of all 648 trials for each participant, but none were considered outliers. Occurrence of noise during experiment testing mentioned in the Method section, led to 66 responses being treated as missing values.

To test the Norwegian participant's percent of correct responses of English AV tense/lax vowel pairs in plain and clear speech, and compare their results to the English perceivers' responses, a generalized mixed model was carried out in JASP 0.16 (JASP Team, 2021). Percent of correct responses was used as the dependent variable, which consisted of either correct responses by responding with the same word as the presented target word, and incorrect responses by responding with another word than the presented target word as well as non-responses. Language group (Norwegian and English), Modality (AO, AV and VO), Speech Style (clear and plain) and Tensity (tense and lax) were the fixed effects variables, and the random effects grouping factors were Experiment Version (any of the three experiment versions), Day (first and second day of testing) and Stimulus Talker. The analysis did not include random slopes, and all statistical assumptions were satisfied.

First, the broad pattern of results will be presented, followed by a presentation of the similar response pattern of the two language groups, as well as their differences. Lastly, clear speech advantage with the Norwegian perceivers and English perceivers will be compared.

Broad Patterns of Results

The main goal of the study was to investigate whether Norwegian participants' use of duration to perceive vowels in Norwegian, transfers to the perception of English AV tense and lax vowels in clear- and plain speech, and to compare their results with those of the English participants tested by Redmon et al. (2020). The main effects can provide a broad overview of the data results (see Figure 1).

The main effect of Language Group on percent correct responses was significant X^2 (1, $N = 46$) = 105.924, $p < .001$, and as expected, the English perceiver group showed a higher percent of correct responses ($M = 82$, $SD = 38$) than the Norwegian perceiver group ($M = 78$, $SD = 41$).

The main effect of Speech Style on percent of correct responses across both language groups was significant X^2 (1, $N = 46$) = 54.330, $p < .001$, where clear speech led to a higher percent of correct responses ($M = 82$, $SD = 39$) than plain speech ($M = 79$, $SD = 41$), which was expected based on previous research that have demonstrated that clear speech improves perception.

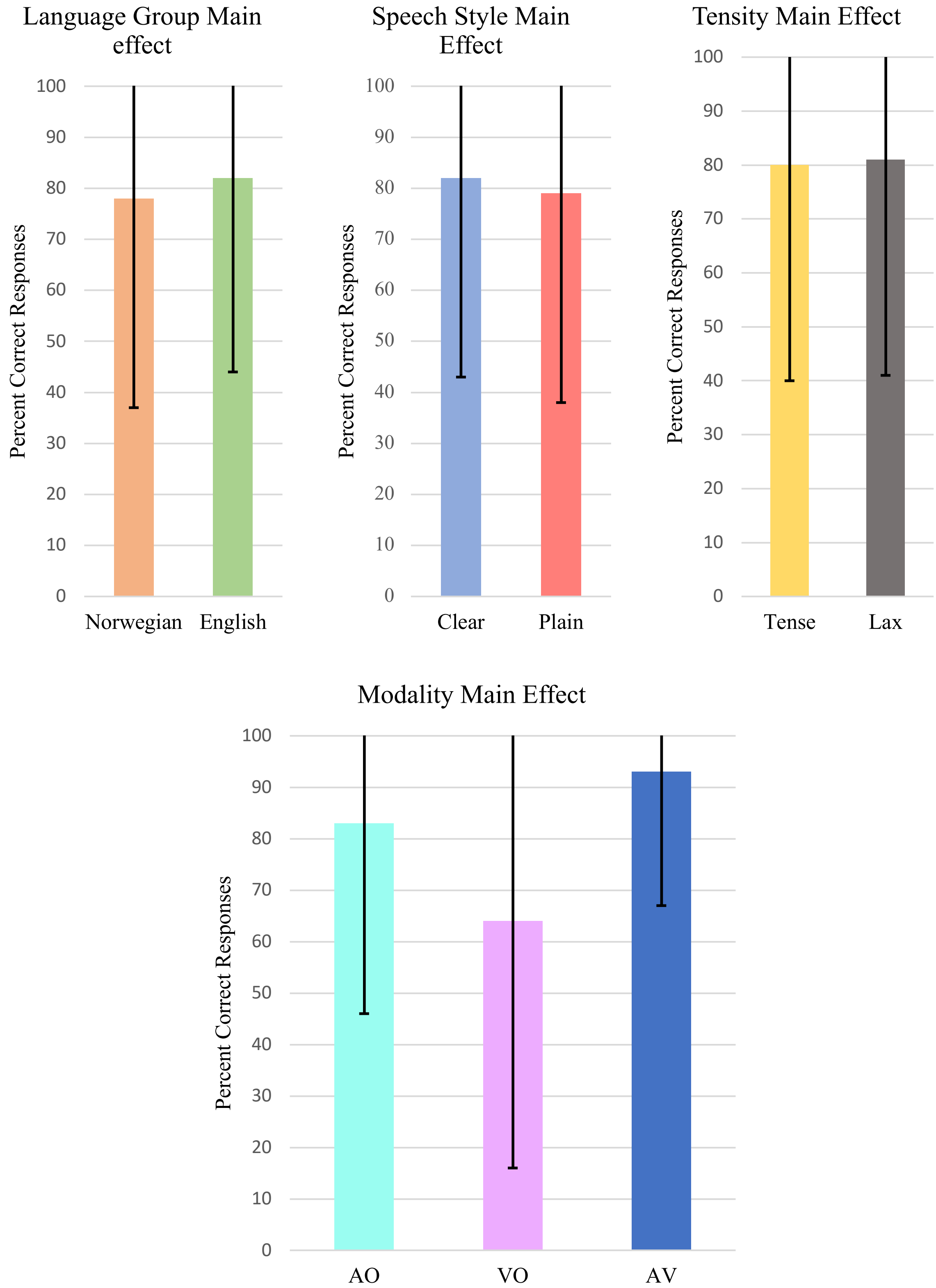
The main effect of Tensity on percent of correct responses was significant X^2 (1, $N = 46$) = 14.878, $p < .001$ across both perceiver groups, with a higher percent of correct responses for lax vowels ($M = 81$, $SD = 40$) than for tense vowels ($M = 80$, $SD = 40$). This result will be discussed later on.

The main effect of Modality on percent of correct responses for both perceiver groups was significant X^2 (2, $N = 46$) = 2675.106, $p < .001$. A Tukey post hoc test revealed a significant difference between all three modalities ($p < .001$), where the percent of correct responses was, not unexpectedly, the highest in AV ($M = 93$, $SD = 26$), a little lower in AO

(*M* = 83, *SD* = 37), and the lowest in VO (*M* = 64, *SD* = 48). These results show that when both audio and video were available during speech perception in the AV modality, the stimuli words were more intelligible for the perceivers. When only audio and no video was available in AO, the percent of correct responses was smaller than of AV, while when only video and no audio was available, the percent of correct responses was even smaller than of AV and AO.

Figure 1

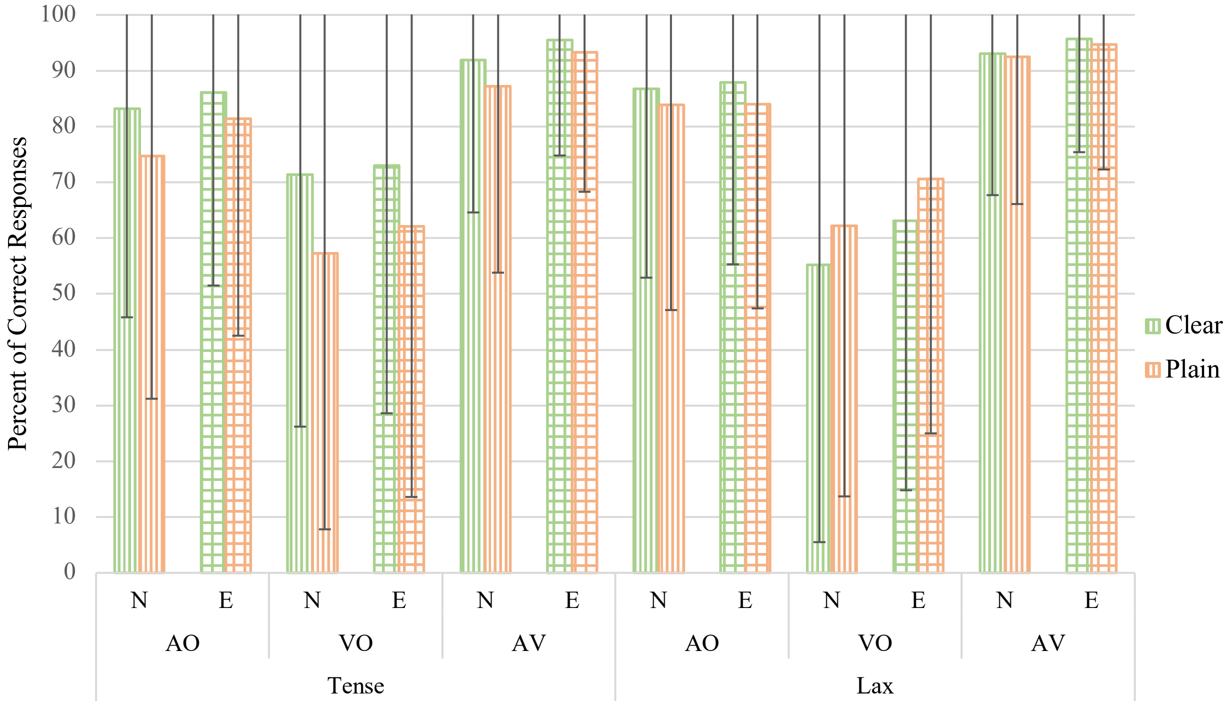
Main Effects of Language Group, Speech Style, Tensity and Modality



Note. The representation of the figures of each main effect shows mean percent of correct responses, with error bars for standard deviations of the means.

Figure 2

Norwegian and English Perceivers' Percent of Correct Responses



Note. Norwegian (N) and English (E) perceivers' mean percent of correct responses (y-axis) on stimuli words with tense and lax vowels (x-axis), presented in clear (green) or plain (orange) speech in the modalities AO, VO and AV (x-axis). Error bars visualizes the standard deviations of the means.

Table 3

Norwegian and English perceivers' mean percent of correct responses

	Norwegian Perceivers					
	Tense			Lax		
	AO	VO	AV	AO	VO	AV
M Clear	83	71	92	87	55	93
SD Clear	37	45	27	34	50	25
M Plain	75	57	87	84	62	93
SD Plain	44	50	33	37	49	26
	English Perceivers					
	Tense			Lax		
	AO	VO	AV	AO	VO	AV
M Clear	86	73	96	88	63	96
SD Clear	35	44	21	33	48	20
M Plain	81	62	93	84	71	95
SD Plain	39	49	25	37	46	22

Note. Means (M) and standard deviations (SD) of percent of correct responses.

Norwegian and English Perceiver Response Patterns

As visualized in Figure 2, the percent of correct responses of both language groups follows a similar pattern in all conditions. Therefore, the interactions of the fixed effects variables Speech Style, Modality and Tensity will be presented in this part across the two perceiver groups.

An important goal with this study was to investigate whether clear speech contributed to an increased intelligibility of audiovisual tense and lax vowels. The aforementioned main effect of Speech Style on percent of correct responses showed that clear speech led to more correct responses than plain speech. Further, there was a significant interaction between Modality and Speech Style X^2 (2, N = 46) = 13.390, p < .01. A Tukey post hoc test revealed a clear speech advantage in AO, with a significantly higher (p < .001) percent of correct responses of clear speech (M = 86, SD = 35) than of plain speech (M = 81, SD = 39). Further, there was a clear speech advantage in VO, with a significantly higher (p < .001) percent of correct responses of clear speech (M = 66, SD = 48) than of plain speech (M = 63, SD = 48). However, there was no significant difference (p = .065) between the speech styles in AV. The results show a clear speech advantage across both language groups in AO and VO, where either audio or video was lacking. In AV, clear speech did not contribute to a notably higher percent of correct responses as the access to both visual and auditory information in plain speech led to a smaller difference between the speech styles in intelligibility for the perceivers.

The aforementioned main effect of Tensity showed a higher percent of correct responses for lax vowels than for tense vowels. There was also a significant interaction between Modality and Tensity X^2 (2, N = 46) = 48.275, p < .001, with a significantly higher percent of correct responses of lax vowels (M = 86, SD = 35) than of tense vowels (M = 81, SD = 39) in AO. In VO, the pattern is the opposite, with a significantly higher percent of correct responses of tense vowels (M = 66, SD = 47) than of lax vowels (M = 62, SD = 48). In AV, there was no significant difference between percent of correct responses of tense vowels and lax vowels (p = .090).

These results are further elaborated through the significant interaction of Speech Style, Modality and Tensity X^2 (2, N = 46) = 30.419, p < 0.001) (see Figure 2 and Table 3). In AO, there was no significant difference (p = 0.309) between tense and lax vowels in clear speech, while in plain speech, lax vowels (M= 84, SD = 37) showed a significantly (p < .001) higher percent of correct responses than of tense vowels (M = 78, SD = 42). In VO, tense vowels (M = 72, SD = 45) showed a significantly higher (p < .001) percent of correct responses than lax vowels (*M*= 59, *SD* = 49) in clear speech, and lax vowels (*M* = 66, *SD* = 47) showed a significantly higher ($p < .001$) percent of correct responses than of tense vowels (*M* = 60, *SD* = 49) in plain speech. Lastly, in AV, no significant difference ($p = 1.000$) was found between tense and lax vowels in clear speech, and no significant difference ($p = .087$) between tense and lax vowels in plain speech.

When looking further on the significant interaction between Speech Style and Tensity, X^2 (1, $N = 46$) = 38.290, $p < .001$, the same pattern emerges. The interaction showed a higher percent of correct responses for tense vowels (*M* = 83, *SD* = 37) than lax vowels (*M* = 80, *SD* = 40) in clear speech, whereas in plain speech, there was a higher percent of correct responses for lax vowels (*M* = 81, *SD* = 39) than tense vowels (*M* = 76, *SD* = 43). As stated before, the Norwegian participants were expected to misinterpret lax vowels as tense vowels in clear speech when presented in VO, as was already observed with the English participants (Redmon et al., 2020). The results show a higher percent of correct responses for tense vowels than lax vowels in clear speech in VO, while no significant difference in percent of correct responses for tense and lax vowels in clear speech in AO and AV. When the tense and lax vowels were presented in plain speech in AO and VO, however, the percent of correct responses is higher for lax vowels than tense vowels.

Differences Between Norwegian and English Results

Language Group interacted significantly with Modality, and with Tensity and Modality, and can show what the differences between the two language groups' response patterns were.

The English perceivers were expected to show a higher percent of correct responses than the Norwegian perceivers, as stimuli were presented in English. The interaction of Language Group and Modality was significant X^2 (2, $N = 46$) = 14.419, $p < .001$. To assess the language groups' percent of correct responses in each modality, a Tukey post hoc test was conducted. It revealed a significantly higher percent of correct responses for the English perceivers (*M* = 84, *SD* = 36) than for the Norwegian perceivers (*M* = 82, *SD* = 38) in AO, a significantly higher percent of correct responses for the English perceivers (*M* = 67, *SD* = 47) than for the Norwegian perceivers (*M* = 62, *SD* = 49) in VO, and a significantly higher percent of correct responses for the English perceivers (*M* = 95, *SD* = 22) than for the Norwegian perceivers (*M* = 91, *SD* = 28) in AV. Based on percent of correct responses within each modality, the English participants performed better than the Norwegian participants in identifying the stimuli words, as expected. This is consistent with the previously mentioned main effect of Language Group on percent of correct responses, which showed that the English participants had a higher percent of correct responses than the Norwegian participants.

A significant interaction was found between Language Group, Modality and Tensity X^2 (2, N = 46) = 13.731, p < 0.001. A Tukey post hoc test showed that the Norwegian perceivers had a significantly (p < 0.001) higher percent of correct responses of lax vowels (M = 85, SD = 35) than of tense vowels (M = 79, SD = 41) in AO, while the English perceivers showed no significant difference (p = .693) between percent of correct responses of tense and lax vowels in this modality. In AV, The Norwegian perceivers did not show a significant difference (p = .078) between percent of correct responses of tense and lax vowels, while the English perceivers did show a significantly (p < 0.001) higher percent of correct responses of lax vowels (M = 95, SD = 21) than of tense vowels (M = 94, SD = 23). In VO, the Norwegian perceivers showed a significantly higher (p < 0.001) percent of correct responses of tense vowels (M = 64, SD = 48) than of lax vowels (M = 59, SD = 49), and the English perceivers showed a significantly higher (p < 0.001) percent of correct responses of tense vowels (M = 68, SD = 47) than of lax vowels (M = 67, SD = 47). In summary, the intelligibility of lax vowels for the Norwegian perceivers was higher than tense vowels in AO, while in VO, they found tense vowels more intelligible than lax vowels. The English perceivers found lax vowels more intelligible than tense vowels in AV, but tense vowels more intelligible than lax vowels in VO.

There was no significant interaction between Language Group and Speech Style X^2 (1, N = 46) = 0.218, p = .641, or between Language Group and Tensity X^2 (1, N = 46) = 1.527, p = .217.

Clear Speech Advantage with Norwegian Perceivers

The hypotheses of expected results for the Norwegian perceiver group could not be accounted for by the Generalized Linear Mixed Model, as no interactions of Language Group and Speech style were significant. The interaction was not significant between Language Group, Modality, Speech Style and Tensity X^2 (2, N = 46) = 0.487, p = 0.784, between Language Group, Speech Style and Tensity X^2 (1, N = 46) = 1.513, p = 0.219, or between Language Group, Modality and Speech Style X^2 (2, N = 46) = 0.391, p = 0.823. The lack of significant results in interactions with Language Group and Speech style involved, can be explained by high standard deviations of the means of percent of correct responses for clear and plain speech styles in both language groups (see Table 3).

To address the expected results from the Norwegian perceiver group, and compare them to the English perceivers' results from Redmon et al. (2020), a Generalized Linear Mixed Model was carried out with only the Norwegian perceiver data included. Percent of correct responses was used as dependent variable, Speech Style, Tensity and Modality were the fixed effects variables, and Experiment Version, Day and Stimulus Talker were the random effects grouping factors. The analysis did not include random slopes, and all statistical assumptions were satisfied.

The Norwegian participants' data was expected to show a clear speech advantage for the intelligibility of tense vowels in all modalities (AO, AV, VO). The interaction between Modality, Speech Style and Tensity was significant X^2 (2, N = 25) = 13.644, p = < 0.001, and a Tukey post hoc test was conducted. The Norwegian perceivers showed a significant (p < .001) clear speech advantage for percent of correct responses of tense vowels in AO (clear: M = 83, SD = 37; plain: M = 75, SD = 44), and in VO (clear: M = 71 , SD = 45; plain: M = 57, SD = 50), as was expected. This is consistent with the English perceivers' results. However, there was not a clear speech advantage found for tense vowels in AV, like there was for the English perceivers, with no significant difference (p = 0.091) between percent of correct responses. Another expectation was that there would be a smaller clear speech advantage for tense vowels in VO than in AO and AV, because of lack of auditory information. However, the opposite occurred, where the biggest clear speech advantage was found in VO.

The next expectation stated that the Norwegian participants would show a clear speech advantage for lax vowel intelligibility with present auditory information in AO and AV, but a slight smaller advantage than for tense vowel intelligibility. There was not a significant clear speech advantage for lax vowels in AO (p = 0.741) for the Norwegian perceivers, where the English perceivers did show a significant clear speech advantage (Redmon et al., 2020). Further, as was found for the English perceivers (Redmon et al., 2020), there was no clear speech advantage (p = 1.000) for tense vowels in AV with the Norwegian perceiver group. In VO, there was a significant difference between percent of correct responses of lax vowels in clear and plain speech, with a higher percent of correct responses in plain speech (M = 62, SD = 49) than in clear speech (M = 55, SD = 50), which can be seen as a clear speech disadvantage. This was expected based on the assumption that the lack of auditory information in VO would result in a misinterpretation of lax vowels as tense vowels, and the result is consistent with the English perceivers' results (Redmon et al., 2020).

In summary, the Norwegian and English perceivers' results show a similar pattern, with a generally higher intelligibility of tense and lax vowels produced in clear speech than in plain speech. Tense and lax vowels presented in AV were found the most intelligible for the perceiver groups, while the VO modality contributes the most poorly to intelligibility for the perceivers. When tense vowels were presented in clear speech in AO and VO, they were found more intelligible for the perceivers than the lax vowels, while in plain speech, lax vowels were found more intelligible than tense vowels in AO and VO.

The analysis with only the Norwegian perceiver data included, showed a clear speech advantage with Norwegian perceivers for tense vowels in AO and VO, while no clear speech advantage was found for lax vowels in any modalities, with a significant clear speech disadvantage for lax vowels in VO.

Discussion

Research has established a clear speech advantage over plain speech during speech perception (Calandruccio et al., 2020; Han et al., 2021; Ferguson & Kewley-Port, 2007; Searl & Evitts, 2013; Tang et al., 2015), which has been demonstrated with perceivers of both native and non-native speech stimuli (Bradlow & Bent, 2002; Redmon et al., 2020; Smiljanić & Bradlow, 2011). In recent studies, which acoustic and visual attributes make clear speech more intelligible for a perceiver than plain speech has been investigated (Leung et al., 2016; Tang et al., 2015). Following this, Redmon et al. (2020) investigated whether perception of non-native vowels in clear speech is influenced by the perceivers' native vowel system, in English perceivers who have vowel tensity in their L1, and Mandarin perceivers who do not have vowel tensity in their L1. Their results showed a clear speech benefit in both perceiver groups for tense vowels in AO, AV and VO. However, for lax vowels, the English perceivers showed no clear speech advantage in AV and VO where visual information was available, and the Mandarin perceivers showed no clear speech advantage in any of the modalities.

In English, spectral features and duration are what make tense and lax vowels distinct, and the same applies to Norwegian short and long vowels. However, English perceivers use spectral cues primarily to distinguish vowels (Hillenbrand et al., 2000), while Norwegian perceivers use duration in addition to spectral cues (Behne og Nylund, 2004). Furthermore, Norwegian adjacent vowels and consonants vary in duration as a result of vowel length and postvocalic consonant voicing (Behne & Moxness, 1994), which give native Norwegians an experience with duration differences from their L1. On this basis, an extension of Redmon et al. (2020) was offered through the Norwegian vowel system, and the present study investigated whether native Norwegian participants' experience with vowel length transfers to perception of English AV tense/lax vowel contrasts in plain- and clear speech.

Norwegian and English Perceiver Response Pattern

The English perceivers were expected to show a higher response accuracy than the Norwegian perceivers in all conditions, as the speech stimuli was in the English participants' native language. As expected, the English participants achieved a higher response accuracy than the Norwegian participants, for both tense and lax vowels in all modalities, in both plain- and clear speech. However, the Norwegian and English perceiver results showed a similar pattern as illustrated in Figure 2, and will therefore be investigated further in this part.

The results showed a clear speech benefit across language groups in all three modalities based on mean percent of correct responses (see Table 3), but only a significant difference between clear and plain speech in AO and VO, not in AV. Because of the available information in AV, plain speech became more intelligible for the perceivers, which resulted in a smaller advantage of clear speech. This is consistent with earlier research, where audiovisual speech stimuli improved speech perception when compared to unimodal speech perception (Hazan et al., 2006).

The main effect of the Modality variable can enlighten the significance of auditory and visual information during speech perception. The results showed that when both audio and video were available to the participants in AV, their response accuracy was the highest when compared to their response accuracy in AO and VO where either audio or video were lacking. The result conforms to earlier research which have demonstrated that multimodal information makes speech perception more intelligible for a perceiver (Hazan et al., 2006). As the lowest response accuracy was found in VO, the participants relied more on auditory information than on visual information when identifying the stimuli words. This is consistent with earlier research where the perceivers found speech stimuli presented in VO less intelligible than in AO and AV (Hazan et al., 2006).

The lower response accuracy of vowels in the VO modality, might be explained by visible articulatory features from lip- and jaw movements during vowel production. The onset-time of jaw- and lip movements has been observed to precede the onset-time of voicing of vowels, while the mouth closes after the voicing has ended (Gracco & Löfqvist, 1994). On this basis, visible duration cues during vowel perception in VO can arguably be less precise than auditory duration cues, which possibly results in a less accurate performance for tense/lax vowel distinction in VO.

The similar response pattern of the two language groups, can be explained by their native speech background. Both native English- and native Norwegian perceivers use vowel spectrum to distinguish vowels in their native languages, while native Norwegian perceivers use vowel length in a bigger degree during vowel perception than English perceivers (Behne & Nylund, 2004; Hillenbrand et al., 2000). Additionally, Norwegian adjacent vowels and consonants vary in duration in speech production (Behne & Moxness, 1994), which gives native Norwegian speakers an experience with temporal timing through their L1. Consequently, the Norwegian perceivers' experience with both vowel- and consonant duration, and vowel spectrum, when perceiving Norwegian speech, seems to have contributed to their sensitivity to the distinctions of English tense and lax vowels in clear and plain speech.

As the native Mandarin participants in Redmon et al.'s study (2020) did not have vowel tensity in their L1, a comparison with the Norwegian participants who have vowel duration in their L1, can substantiate the influence of the Norwegian perceivers' native vowel system on non-native vowel perception. Like the Norwegian participants, the Mandarin participants seemed to have used vowel duration to identify tense and lax vowels (Redmon et al., 2020), despite their lack of experience with vowel tensity from Mandarin, as other studies also have demonstrated (e.g., Bohn & Flege, 1990; Cebrian, 2006; Kondaurova & Francis, 2009), and in line with the Desensitization theory (Bohn, 1995, pp. 294-295). Regardless, the Norwegian participants' response pattern shows a higher performance in identifying the stimulus vowels than the Mandarin participants, and shows a higher response accuracy for lax vowels in clear speech than in plain speech in AO and AV, where the Mandarin participants did not show a clear speech benefit (Redmon et al., 2020). Taken together, the Norwegian participants were more sensitive to the durational contrasts of tense and lax vowels presented in clear speech, than the Mandarin participants who were lacking the same experience with vowel duration. This supports that experience with vowel length in a native language, transfers to non-native AV vowel perception.

Another explanation of the Norwegian participants' similar response pattern as the English participants', might be their English proficiency. Although the participants included in the present study were selected based on several criteria to diminish their English experience, the general English proficiency of native Norwegians is high (Education First, 2020, pp. 6-7), nevertheless. The participants' exposure to English speech in their daily life, for example through social media and streaming apps, could not be a reason for exclusion from the study, as recruitment of participants would be challenging for the desired age group.

As a result of the Norwegian participants' English language skills, they might have attended more to spectral cues during perception of the tense and lax vowels, as a native English perceiver would do (Hillenbrand et al., 2000). Such a result was found in a study, where Norwegian perceivers used both duration and formant movement to distinguish English vowels (Iverson & Evans, 2007). However, whether duration cues, spectral cues, or both were used to identify AV tense and lax vowels in clear speech by the Norwegian participants, cannot be accounted for by the native Norwegian data, but is a question that could be enlightened in future research. Native Norwegian participants with a higher English proficiency than the ones tested in the present study, could be tested in the same experiment, to compare the results from both groups and look into whether higher English proficiency results in higher response accuracies. However, as the English participants showed higher response accuracies than the Norwegian participants in all conditions in the present study, the Norwegian participants are less likely to have used spectral cues in the same extent as the English participants did.

Perception of English Tensity in Clear Speech

Recent research have found an interconnection of tense and lax vowels produced in plain and clear speech styles (Leung et al., 2016; Tang et al., 2015). This interconnection was observed in the results of the present study as well, and will be looked into further in this part.

The response accuracy of tense vowels in clear speech was higher than for the lax vowels, while in plain speech, the response accuracy was higher for lax vowels than for tense vowels, when both language groups were taken into consideration. The analysis showed a significantly higher response accuracy for tense vowels than lax vowels in clear speech in VO, while in AO and AV, the response accuracy for tense and lax vowels did not show a significant difference. Consequently, the higher response accuracy of tense vowels over lax vowels in clear speech derived from the VO modality. As argued by Redmon et al. (2020), the English participants seemed to misinterpret lax vowels as tense vowels when presented in clear speech in VO, which decreases the response accuracy of the lax vowels. While the English participants attend to spectral differences (Hillenbrand et al., 2000), and the Norwegian participants attend to duration differences together with spectral differences (Behne & Nylund, 2004), in vowel perception, both speech features are interfered with when lax vowels are produced in clear speech. Lax vowels increase both in spectral properties and duration in clear speech (Leung et al., 2016), which make the lax vowels' characteristics similar to those of tense vowels, and consequently, both language groups seemed to be deceived by the decrease in tense/lax vowel contrasts.

In plain speech, the higher response accuracy of lax vowels over tense vowels derived from the AO and VO modalities, while in AV, the difference between response accuracy of tense and lax vowels was not significant. As lax vowels have a shorter duration than tense vowels in general, lax vowels produced in plain speech can be visualized as the other extreme to tense vowels produced in clear speech. Lax vowels in plain speech remain short in duration when compared to tense vowels in plain speech, while tense vowels' already longer duration than lax vowels, becomes even longer in clear speech (Leung et al., 2016). Therefore, the higher response accuracy for lax vowels over tense vowels in plain speech in AO and VO, might be explained by a short duration which is easier to detect than the longer tense vowels.

Although the result patterns of the Norwegian and English perceivers were similar, the interaction of Language Group, Modality and Tensity showed some differences between the two perceiver groups. The results showed that the Norwegian perceivers found lax vowels more intelligible than tense vowels in AO, where the English perceivers showed no significant difference for vowel tensity. However, in AV, only the English perceivers showed a significant difference of response accuracy of tense and lax vowels, where lax vowels were found more intelligible than tense vowels. These differences in the perceiver groups in AO and AV, might be due to their native vowel systems, or due to errors done in tensity distinction. However, whether any of these are true cannot be accounted for by the collected data, and will not be further discussed. In VO, both perceiver groups found tense vowels more intelligible than lax vowels. As stated earlier, lax vowels were misperceived as tense vowels in a bigger degree in VO, which can explain the higher response accuracy for tense over lax vowels. Additionally, the tense vowels' longer duration and more extreme lip and jaw movements than of lax vowels (Tang et al., 2015), seems to have been visible features which were easier to detect than for the lax vowels.

Clear Speech Benefit in Norwegian Perceivers

The Norwegian participants were expected to show a clear speech advantage for tense vowels in all modalities. As expected, a clear speech advantage was found for the Norwegian perceivers in AO and VO, but unexpectedly, no clear speech advantage was found in AV. As the two unimodal modalities showed a clear speech advantage for tense vowels, which has been found to be less intelligible for perceivers than multimodality (e.g., AV) (Hazan et al., 2006), clear speech contributed to make the audio in AO, and video in VO, easier to perceive than when presented in plain speech, which is consistent with the English perceiver results (Redmon et al., 2020). Furthermore, auditory and visual duration cues from acoustic lengthening and extreme spectral features (Leung et al., 2016), as well as an increase in visual facial movements (Tang et al., 2015), of vowel production in clear speech, seems to have been beneficial in the unimodalities for the Norwegian perceivers. These are auditory and visual attributes of vowel duration, which the Norwegian perceivers seem to have attended to during perception of tense vowels, which became more salient and easier to perceive in clear speech. However, as the English perceiver group, which primarily attend to spectral features during vowel perception, showed a clear speech advantage for tense vowels in AO and VO as well, the more extreme spectral features from clear speech increased intelligibility. This might also have been the case for the Norwegian perceivers, as they use vowel quality to distinguish vowels in their L1 in addition to vowel length (Behne & Nylund, 2004).

The Norwegian perceivers were further expected to show a bigger clear speech advantage for tense vowels when auditory information was present in AO and AV, than in VO with only present visual information, as VO have been found to be the least intelligible of the three modalities (Hazan et al., 2006). However, the opposite occurred, where the biggest clear speech advantage was found in VO, which indicates that the visual cues of duration might have been used in a bigger degree than expected. The stimuli vowels /i-ɪ/ are produced with lip stretching, /ɑ-ʌ/ are produced with jaw lowering and /u-ʊ/ are produced with lip rounding. These are visible articulatory cues that are greater in clear speech production than in plain speech (Kim & Davis, 2014; Tang et al., 2015), which might have given the Norwegian participants enough information in the VO modality to distinguish tense vowels from lax vowels based on duration. Additionally, visual speech presentation offers duration cues from the time span of lip opening and closing, and jaw lowering and closing (Gracco & Löfqvist, 1994) that the Norwegian perceivers seem to have attended to when audio was lacking. This result is persistent with earlier research where perceivers used visual cues more than auditory cues during non-native speech perception (Hazan et al., 2010).

The perception of the lax vowels of the Norwegian participants was expected to show a clear speech advantage in AO and AV, because of the available auditory duration cues from the shorter length of lax vowels than of tense vowels (Leung et al., 2016), but not as big an advantage as for tense vowels. However, no clear speech advantage was present in AO and AV, as lax vowels in clear speech were not significantly more intelligible than in plain speech for the Norwegian perceivers. As noted earlier, lax vowels produced in plain speech make them keep the features that make them distinct from tense vowels, that is, short duration and smaller spectral features (Leung et al., 2016). Consequently, perception of lax vowels in plain speech was not significantly poorer than in clear speech. Furthermore, lax vowels were found to be misperceived as tense vowels in clear speech by the English and Mandarin perceivers in Redmon et al. (2020). This might have been the case for the Norwegian perceivers as well, as lax vowels in clear speech were not significantly more intelligible for the perceivers than lax vowels in plain speech.

In VO, the Norwegian perceivers were not expected to show a clear speech advantage, as they were expected to misperceive lax vowels as tense vowels as was observed for the English and Mandarin perceivers in Redmon et al. (2020). Additionally, visual duration cues are less precise than auditory cues of duration, as lip- and jaw movements precede the voice-onset time, and end after the vowel sound has stopped (Gracco & Löfqvist, 1994). This was expected to result in a lower response accuracy for lax vowels in VO if the Norwegian perceivers used duration to distinguish lax vowels from tense vowels. As expected, Norwegian perceivers did not show a clear speech advantage for lax vowels in VO. As Norwegian perceivers have been found to use both duration and formant movement to recognize English vowels in earlier research (Iverson & Evans, 2007), the lack of auditory cues from resonance and duration in the VO modality might have resulted in a smaller accuracy for lax vowels. The second cause is an increase in a misinterpretation of lax vowels as tense vowels when auditory information is lacking, because of a reduction in tense-lax vowel contrasts in clear speech (Hillenbrand et al., 1995).

The interactions where the Norwegian and English participant's response accuracy were included together with Speech Style, were not significant as a result of high SDs of the mean response accuracy of both language groups. The experiment setting and execution of the data collection done in different labs for each group should not be the reason for the high SDs, as these were observed with both language groups. The SDs of the Mandarin perceiver group from Redmon et al., (2020) were not available for the current study, but could be looked into to check whether the Mandarin group also showed high SDs for response accuracy means. Then, the possibility that language background was the reason for varying response accuracies across participants could be excluded. As high SD's were observed for both the Norwegian and English perceivers, the experiment task might explain the big variances between participants' response accuracies. The experiment task had six response alternatives in each trial, which probably required that the participants had to move their eyes to identify their perceived response word. Some participants might have been more comfortable with such a task, and found it easier to detect the perceived response alternative among all six response words, within the time frame of four seconds to respond.

Implications

In summary, the present study presents implications of a transference of native vowel system to non-native vowel perception. Norwegian perceivers' response accuracy of tense and lax vowels were higher than those of the Mandarin perceivers, who did not have experience with vowel tensity, tested by Redmon et al. 2020. This implies that the Norwegian perceivers' experience with duration contrasts from Norwegian adjacent consonants and vowels, made them more sensitive to duration contrasts in perception of English tense and lax vowels. Further, the Norwegian and English perceiver groups showed a similar response pattern, which indicates that the Norwegian perceivers' use of duration together with spectral features to distinguish vowels in their L1, might have been applied to English vowel distinction.

Within interactions with Modality, both language groups found tense and lax vowels in AV the most intelligible, and the VO modality as the least intelligible, which supports previous research findings.

Where the English perceivers showed a clear speech advantage for tense vowels in all modalities, the Norwegian perceivers did not show a clear speech advantage for tense vowels in AV, as a result of a higher intelligibility for multimodal perception which did not make tense vowels in clear speech easier to detect than in plain speech. For the lax vowels, the Norwegian perceivers showed no clear speech advantage in AO, where the English perceivers did. This was possibly a result of their attention to lax vowel duration which made lax vowels equally intelligible in both plain and clear speech styles, or as a result of misperceiving lax vowels as tense vowels in clear speech, which prevented a clear speech benefit in AO. As found for the English perceivers, a clear speech disadvantage for lax vowels in VO was found for the Norwegian perceivers, as lax vowels were misperceived as tense vowels because of lack of tense/lax vowel contrasts in clear speech. In AV, response accuracy of lax vowels produced in clear and plain speech did not differ significantly for any of the perceiver groups.

The results show that clear speech improves intelligibility of tense and lax vowels over all, but not in all combinations of tensity and modality. The interconnection of tense and lax vowels and plain and clear speech styles gives an example of an exception to the earlier demonstrated clear speech benefit. Further, the results imply that clear speech in general improves non-native intelligibility, however, the native language of the perceiver influences whether clear speech improves intelligibility or not.

Conclusion

The study implies that temporal timing experience from a native vowel system influences non-native vowel perception in clear speech, and that long term learning of temporal timing through a native language, influences non-native perception, and consequently, might influence acquisition of L2 vowel system learning. Clear speech contributes to a higher intelligibility of English vowel contrasts for non-native perceivers, but the interconnection of tense/lax vowel contrasts and plain-to-clear speech modifications constrain a clear speech benefit, especially when only visual information is available to perceivers.

References

Behne, D. M., Alm, M., Berg, A., Engell, T., Foyn, C., Johnsen, C., Srigaran, T., & Torsdottir, A. E. (2013). Effects of musical experience on perception of audiovisual synchrony for speech and music. *The Journal of the Acoustical Society of America, 133*(5), 3570. https://doi.org/10.1121/1.4806538

Behne, D. M., & Moxness, B. (1994). Effects of speaking rate, focal stress, and distinctive vowel length on syllable-internal timing in Norwegian. *The Journal of the Acoustical Society of America*, *96*(5), 3349–3349. https://doi.org/10.1121/1.410627

Behne, D., Moxness, B., & Nyland, A. (1996). Acoustic-phonetic evidence of vowel quantity and quality in Norwegian. *TMH-QPSR, 37*(2), 13-16.

Behne, D., & Nylund A. (2004). Transference or desensitization? – A study of vowel spectra and duration. *Proceedings of the 18th International Congress on Acoustics*, *1*, 603-606.

Behne, D., & Wang, Y. (2018). *Does native language temporal experience transfer to AV synchrony perception?* [Poster]. Department of Psychology, Norwegian University of Science and Technology & Department of Linguistics, Simon Fraser University.

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied psycholinguistics*, *11*(3), 303-328. 10.1017/S0142716400008912

Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 279-304). York Press.

Bradlow, A. R. (2002). Confluent talker-and listener-oriented forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 241-273). De Gruyter Mouton. https://doi.org/10.1515/9783110197105

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272-284. https://doi.org/10.1121/1.1487837

Calandruccio, L., Porter, H. L., Leibold, L. J., & Buss, E. (2020). The clear-speech benefit for school-age children: Speech-in-noise and speech-in-speech recognition. *Journal of Speech, Language, and Hearing Research, 63*(12), 4265–4276. https://doi.org/10.1044/2020_JSLHR-20-00353

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372–387. https://doi.org/10.1016/j.wocn.2005.08.003

Chobert, J., François, C., Velay, J.-L., & Besson, M. (2014). Twelve months of active musical training in 8-to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex (New York, N.Y. 1991)*, *24*(4), 956–967. https://doi.org/10.1093/cercor/bhs377

Clopper, C. G., Pisoni, D. B., & De Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3 I), 1661–1676. https://doi.org/10.1121/1.2000774

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America*, *128*(4), 2059-2069. https://doi.org/10.1121/1.3478775

Cooper, A., & Wang, Y. (2009). Effects of linguistic and musical experience on non-native perception of Thai vowel duration. *The Journal of the Acoustical Society of America*, *125*(4), 2773. https://doi.org/10.1121/1.4784740

Cooper, A., Wang, Y., & Ashley, R. (2016). Thai Rate-Varied Vowel Length Perception and the Impact of Musical Experience. *Language and Speech*, *60*(1), 65–84. https://doi.org/10.1177/0023830916642489

Education First. (2020). *EF EPI: EF English proficiency index*. https://www.ef.com/wwen/epi/

Ferguson, S. H. (2012). Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss. *Journal of Speech, Language, and Hearing Research*, *55*(3), 779–790. https://doi.org/10.1044/1092-4388(2011/10-0342)

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *112*(1), 259-271. https://doi.org/10.1121/1.1482078

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1241–1255. https://doi.org/10.1044/1092-4388(2007/087)

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., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in*

Second Language Acquisition, *26*(1), 1–34. https://doi.org/10.1017/S0272263104026117

Gopal, H. S. (1990). Effects of speaking rate on the behavior of tense and lax vowel durations. *Journal of Phonetics*, *18*(4), 497–518. https://doi.org/10.1016/S0095-4470(19)30411-5

Gracco, V. L., & Lofqvist, A. (1994). Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *The Journal of Neuroscience*, *14*(11), 6585–6597. https://doi.org/10.1523/jneurosci.14-11-06585.1994

Han, H. J., Munson, B., & Schlauch, R. S. (2021). Fundamental frequency range and other acoustic factors that might contribute to the clear-speech benefit. *The Journal of the Acoustical Society of America*, *149*(3), 1685–1698. https://doi.org/10.1121/10.0003564

Hazan, V., Kim, J., & Chen, Y. (2010). Audiovisual perception in adverse conditions: Language, speaker and listener effects. *Speech Communication*, *52*(11), 996–1009. https://doi.org/10.1016/j.specom.2010.05.003

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *Journal of the Acoustical Society of America* 119, 31740–1751. https://doi.org/10.1121/1.2166611

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–3022. https://doi.org/10.1121/1.1323463

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, *97*(5), 3099-3111. https://doi.org/10.1121/1.411872

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), 2842–2854. https://doi.org/10.1121/1.2783198

JASP Team. (2021). *JASP* (Version 0.16) [Computer software]. https://jasp-stats.org/

Kim, J., & Davis, C. (2014). Comparing the consistency and distinctiveness of speech produced in quiet and in noise. *Computer Speech & Language*, *28*(2), 598-606. https://doi.org/10.1016/j.csl.2013.02.002

Kondaurova, M. V., & Francis, A. L. (2009). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, *124*(6), 3959–3971. https://doi.org/10.1121/1.2999341

Kristoffersen, G. (2000). *The phonology of Norwegian.* Oxford Linguistics.

Lengeris, A. (2009). Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica*, *66*(3), 169-187. https://doi.org/10.1159/000235659

Leung, K. K. W., Jongman, A., Wang, Y., & Sereno, J. A. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *The Journal of the Acoustical Society of America*, *140*(1), 45–58. https://doi.org/10.1121/1.4954737

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. 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

Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F. G. (2004). Clear speech perception in acoustic and electric hearing. *The Journal of the Acoustical Society of America*, *116*(4), 2374–2383. https://doi.org/10.1121/1.1787528

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

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Perception Research Systems. (2007). *Paradigm Stimulus Presentation.* http://www.paradigmexperiments.com/

Redmon, C., Leung, K., Wang, Y., McMurray, B., Jongman, A., & Sereno, J. A. (2020). Cross-linguistic perception of clearly spoken English tense and lax vowels based on auditory, visual, and auditory-visual information. *Journal of Phonetics*, *81*, 100980. https://doi.org/10.1016/j.wocn.2020.100980

Searl, J., & Evitts, P. M. (2013). Tongue-palate contact pressure, oral air pressure, and acoustics of clear speech. *Journal of Speech, Language, and Hearing Research*, *56*(3), 826–839. https://doi.org/10.1044/1092-4388(2012/11-0337)

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America*, *130*(6), 4020–4031. https://doi.org/10.1121/1.3652882

Smiljanić, R., & Gilbert, R. C. (2017). Intelligibility of noise-adapted and clear speech in child, young adult, and older adult talkers. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3069–3080. https://doi.org/10.1044/2017_JSLHR-S-16-0165

Tang, L. Y. W., Hannah, B., Jongman, A., Sereno, J., Wang, Y., & Hamarneh, G.. (2015). Examining visible articulatory features in clear and plain speech. *Speech Communication*, *75*, 1–13. https://doi.org/10.1016/j.specom.2015.09.008

University of Oslo. (n.d.). *Nettskjema*. https://www.uio.no/tjenester/it/adm-app/nettskjema/

Ylinen, S., Shestakova, A., Alku, P., & Huotilainen, M. (2005). The Perception of Phonological Quantity based on Durational Cues by Native Speakers, Second-language Users and Nonspeakers of Finnish. *Language and Speech*, *48*(3), 313–338. https://doi.org/10.1177/00238309050480030401

Appendix A

Pre-test Questionnaire

Deltakerkode:_________

Dato: _________
Forsker: __________

Takk for at du deltar i denne studien!
All informasjon som samles inn vil oppbevares konfidensielt.

Alder: __________

Kjønn: Mann ☐ Kvinne ☐ Annet/ønsker ikke å svare ☐

Har du normalt/korrigert syn? (Korreksjon er f.eks. briller eller linser)

Ja ☐

Har du normal hørsel?

Ja ☐

Jeg bekrefter at jeg ikke har noen nevrologisk historikk som kan påvirke syn, hørsel eller oppmerksomhet (f.eks. hjernerystelse, ADHD, dysleksi eller lignende): ☐

Vennligst fyll inn nummeret til hvert av de følgende ordene ved siden av sin definisjon:
(1): cod; (2): cooed; (3): could; (4): cud; (5): keyed; (6): kid

Ord-nummer	**Definisjon**
	a simple past tense of can (fortidsformen av *can*)
	a child or young person (et barn eller en ung person)
	any of several soft-rayed food fishes of the family Gadidae, especially Gadus morhua, of cool, North Atlantic waters. (alle typer fettfinnet matfisk fra familien Gadidae, spesielt Gadus morhua fra kalde Nord Atlantiske vann.)
	the portion of food that a ruminant returns from the first stomach to the mouth to chew a second time. (maten som en drøvtygger bringer opp fra den første magen til munnen for å tygge en gang til.)
	to use a key (å bruke en nøkkel)
	to utter or imitate the soft, murmuring sound characteristic of doves. (å ytre eller å imitere den myke mumlende lyden som duer lager.)

Bakgrunnsinformasjon

1. Bakgrunn

	Ja	Nei
Er norsk morsmålet ditt?		
Startet du din engelskopplæring i 1. klasse på grunnskolen? (dvs. aktiv opplæring med en intensjon om å lære språket)		
Startet du din engelskopplæring før grunnskolen? (dvs. aktiv opplæring med en intensjon om å lære språket)		
Har du vokst opp i et hjem hvor engelsk ble brukt i dagligtalen?		
Har du bodd i et engelsktalende land?		
Har du vært på utveksling i et engelsktalende land?		
Har du gått studieretninger på universitetsnivå hvor det ble undervist på engelsk og deler av, eller hele, pensum var på engelsk?		
Har du tatt enkeltemner på universitetsnivå hvor det ble undervist på engelsk og deler av, eller hele, pensum var på engelsk?		
Har du gått på linjen International Baccalaureate (IB) på videregående skole?		

Takk for at du besvarte dette spørreskjemaet! Dine responser blir registrerte og oppbevares trygt og konfidensielt.

Appendix B

Information Form and Consent Form

Vil du delta i forskningsprosjektet *"Personer med norsk som morsmål sin forståelse av engelske audiovisuelle vokaler i klar tale"?*

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å undersøke personer med norsk som morsmål sin forståelse av engelske vokallyder. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Formålet med denne studien er å undersøke personer med norsk som morsmål sin forståelse av engelske vokaler produsert med forskjellige tale-stiler presentert som lyd og/eller video. Studien skal undersøke hvilken informasjon fra engelske vokaler som brukes av personer med norsk som morsmål til å forstå engelske ord.

Dine opplysninger som ikke er direkte personidentifiserende (alder, kjønn og datamaterialet som samles inn under deltakelse i denne studien) vil brukes i en masteroppgave og i et forskningsprosjekt som kan publiseres og brukes i internasjonale presentasjoner. Deltakere i dette prosjektet vil ikke kunne gjenkjennes i publikasjoner. Deltakerinformasjonen som samles inn vil kun bidra til å beskrive det samlede utvalget, ikke enkelt-deltakere.

Hvem er ansvarlig for forskningsprosjektet?

Det psykologiske institutt ved Norges teknisk-naturvitenskapelige universitet (NTNU) er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?

Du får spørsmål om å delta i studien fordi du er student ved NTNU, og fordi du oppfyller disse utvalgskriteriene for prosjektet:

- Du startet din engelskopplæring i førsteklasse på barneskolen, og ikke før dette.
- Du har ikke vokst opp i et engelsk-talende hjem.
- Du har ikke gått et studie tidligere som har mange forelesninger eller mye pensum på engelsk.
- Du har ikke vært på utveksling eller bodd i et engelsktalende land.
- Du har normalt eller justert syn og normal hørsel.

Alle studenter som oppfylte utvalgskriteriene, har fått spørsmål om å delta i prosjektet. Studiens utvalg vil bestå av 20-30 unge voksne (18-27 år).

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det følgende:

1. Informasjonen som ble samlet inn i det elektroniske spørreskjemaet Nettskjema som du besvarte da du viste din interesse for å delta i prosjektet, vil oppbevares som konfidensiell informasjon.

2. Du vil besvare et spørreskjema om din språkbakgrunn, kjønn, alder, hendthet, om du har normalt eller justert syn og hørsel, og en bekreftelse på at du ikke har nevrologisk historikk som kan påvirke syn, hørsel eller oppmerksomhet. Du vil og lese en liste med engelske ord og skal identifisere ordenes betydning. Dette vil ta omtrent 10 minutter. Dine svar vil registreres på papir, og oppbevares som konfidensiell informasjon.
3. Det vil gjøres et lydopptak av at du leser opp noen få engelske ord. Lydopptaket vil oppbevares på en minnepenn som vil oppbevares innelåst.
4. Du vil delta i et eksperiment hvor du skal lytte til en rekke engelske ord samtidig som du ser ansiktet til taleren. Du vil bli spurt om å respondere på hva du hører og/eller ser ved å bruke en data-mus til å trykke på en skjerm. Dette vil gjennomføres på to ulike dager, hvor hver økt vil ta omtrent 1 time. Det vil si at deltakelsen vil ta omtrent 2 timer til sammen. Dine responser vil registreres elektronisk, og lagres på en minnepenn som oppbevares innelåst.
5. Du vil svare på et kort spørreskjema som handler om eksperimentet du har deltatt i, etter du har gjennomført de to øktene. Dette vil ta omtrent 10 minutter. Dine svar vil registreres på papir, og oppbevares som konfidensiell informasjon.

Undersøkelsen vil finne sted ved Talelaben, Psykologisk institutt, Dragvoll ved NTNU, Trondheim.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- De som vil ha tilgang på dine opplysninger er masterstudent Vilde Sannerud Kalin og førsteamanuensis Dawn Behne som er veileder for dette masterprosjektet. Dawn Behne, førsteamanuensis ved Psykologisk institutt ved NTNU, og Yue Wang, førsteamanuensis ved Language and Brain Lab ved Simon Fraser University, Canada, vil få tilgang på data med ikke direkte personidentifiserende opplysninger som består av deltakeres alder, kjønn og datamaterialet som samles inn fra spørreskjemaene, eksperimentet og lydopptaket.
- Navnet ditt vil erstattes med en kode som lagres på en egen liste adskilt fra øvrige data. Opplysningene som ble samlet inn med Nettskjema da du viste interesse for dette prosjektet vil slettes etter datainnsamlingens slutt senest oktober i 2021. Datamaterialet vil lagres på en minnepenn og oppbevares innelåst. Dine svar på spørreskjemaene som besvares etter samtykke vil oppbevares innelåst.
- Masterstudent Vilde Sannerud Kalin vil være databehandleren som samler inn, bearbeider og lagrer data som blir samlet inn i dette prosjektet.
- Deltakere i dette prosjektet vil ikke kunne gjenkjennes i publikasjoner. Deltakerinformasjonen som samles inn vil kun bidra til å beskrive det samlede utvalget, ikke enkelt-deltakere.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?
Opplysningene dine vil anonymiseres ved slutten av datainnsamlingen senest ved utgangen av oktober 2021. Dine anonymiserte opplysninger (alder, kjønn og datamateriale fra spørreskjemaene og eksperimentet), vil oppbevares permanent som konfidensiell informasjon på papir og minnepenn innelåst.

Lydopptaket uten direkte personidentifiserende opplysninger vil lagres permanent på minnepenn innelåst på Talelaben ved Institutt for psykologi ved NTNU med det formål å brukes til videre forskning. Forskere ved Institutt for psykologi ved NTNU, Norge, og ved Language and Brain Lab ved Simon Fraser University, Canada, vil ha tilgang på lydopptaket uten direkte personidentifiserende opplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Psykologisk institutt ved NTNU har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke opplysninger vi behandler om deg, og å få utlevert en kopi av opplysningene
- å få rettet opplysninger om deg som er feil eller misvisende
- å få slettet personopplysninger om deg
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:

- Psykologisk institutt ved NTNU, ved
    - Masterstudent Vilde Sannerud Kalin på epost: vildeska@stud.ntnu.no
    - Førsteamanuensis Dawn Behne på epost: dawn.behne@ntnu.no
- Vårt personvernombud: NTNU ved Thomas Helgesen på epost: thomas.helgesen@ntnu.no

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:

- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Dawn Behne (Veileder)

Vilde Sannerud Kalin (Masterstudent)

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «Personer med norsk som morsmål sin forståelse av engelske audiovisuelle vokaler i klar tale», og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i eksperimentet på to ulike dager (omtrent 1 t. og 20 min. dag 1, og 1 t. dag 2).
- ☐ å delta ved å svare på to spørreskjema.
- ☐ å delta ved å lese opp noen få engelske ord som det gjøres lydopptak av.
- ☐ at mine opplysninger som ikke er direkte personidentifiserende, som består av alder, kjønn og datamateriale fra spørreskjemaene og eksperimentet, lagres permanent etter prosjektslutt, og vil brukes i et forskningsprosjekt.
- ☐ at lydopptaket uten direkte personidentifiserende opplysninger lagres permanent, og kan brukes til videre forskning av forskere tilknyttet Institutt for psykologi ved NTNU, Norge, og Language and Brain Lab ved Simon Fraser University, Canada.

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

----
(Signert av prosjektdeltaker, dato)

Appendix C

Approval from Norwegian Centre for Research Data (NSD)

Vurdering
19.08.2021

Referansenummer
264968

Prosjekttittel
Norwegian Perceivers' Intelligibility of English AV Vowels in Clear Speech

Behandlingsansvarlig institusjon
Norges teknisk-naturvitenskapelige universitet / Fakultet for samfunns- og utdanningsvitenskap (SU) / Institutt for psykologi

Prosjektperiode
01.08.2021 - 01.12.2023

Dato
19.08.2021

Type
Standard

Kommentar
Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 19.08.2021, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 01.12.2023. Direkte personidentifiserende opplysninger slettes innen prosjektslutt og innsamlede data (med lydopptak) lagres permanent ved NTNU til fremtidig forskning med lignende formål/forskningstema, basert på den registrertes samtykke. Personvernulempen ved permanent lagring av prosjektdata vurderes som svært liten.

LOVLIG GRUNNLAG
Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Lovlig grunnlag for behandlingen av personopplysninger i prosjektet, samt permanent lagring, vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER
NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:
· lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
· formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål · dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
· lagringsbegrensning (art. 5.1 e), ved at direkte personidentifiserende opplysninger ikke lagres lengre enn nødvendig for å oppfylle formålet. Permanent lagring av øvrige data har svært lav personvernulempe og er basert på den registrertes samtykke.

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), og dataportabilitet (art. 20). NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13. Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32). Yue Wang, førsteamanuensis ved Simon Fraser University i Canada, vil ha tilgang til innsamlede data uten direkte personidentifiserende opplysninger som består av deltakeres kjønn, alder og datamateriale, inkludert lydopptak. NSD legger til grunn at behandlingen oppfyller kravene til behandling av personopplysninger utenfor EU (personvernforordningen kapittel 5). For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde: https://www.nsd.no/personverntjenester/fylle-ut-meldeskjema-for-personopplysninger/melde-endringer-i-meldeskjema Du må vente på svar fra NSD før endringen gjennomføres.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene pågår i tråd med den behandlingen som er dokumentert.

Lykke til med prosjektet! Kontaktperson ved NSD Personverntjenester: Eva J. B. Payne

NTNU
Kunnskap for en bedre verden