SYNCHRONIC SOUND CHANGE AS A UNIQUE LOOK AT REPRESENTATION IN /aɪ/ RAISING IN FORT WAYNE, INDIANA

by

ALYSSA STRICKLER

B.A., Indiana University, 2017

M.A., University of Colorado, 2019

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Linguistics

2023

Committee Members:

Esther Brown

Jeremy Calder

Hannah Haynie

Bhuvana Narasimhan

Rebecca Scarborough

ABSTRACT

Strickler, Alyssa Rose (Ph.D., Linguistics)

Synchronic sound change as a unique look at representation

Thesis directed by Professor Rebecca Scarborough

There is an ongoing sound change in progress in the Fort Wayne, Indiana area. This change is well documented as several phonetically and phonologically conditioned patterns of American /aɪ/ raising. While the various locations, patterns and features of American raising are described in several studies, there is little work on the necessary change in the perceptual and/or representational systems of those speakers involved in the change. This dissertation combines a description of speech production and perception, both processes prevalent in production of /aɪ/ raising in Fort Wayne, and investigates perceptual patterns of speakers who both produce a raised /aɪ/ and also perceive it on a daily basis.

This work set out to document paths through sound change for both speech production and perception in /aɪ/ for participants in Fort Wayne, Indiana. What was revealed was variation in the approaches taken by individuals to produce raising, as well as variation without any apparent patterns within individuals in the perception data. These varied perception results, the target results in the clear speech task, and the lack of correlation between the production patterns and perception results within individuals are interpreted to mean that for Fort Wayne speakers, raised pre-voiceless /aɪ/ is not yet part of their representation.

CONTENTS

TABLES

FIGURES

CHAPTER I

INTRODUCTION

Background of American /aɪ/ raising

/aɪ/ raising is a well-documented ongoing sound change in Fort Wayne, Indiana, with several phonetically and phonologically conditioned patterns of American /aɪ/ raising. This version of /aɪ/ raising is similar to Canadian raising, but has innovated independently and without contact to Canada throughout the US in places including Indiana, Ohio, Philadelphia, and the northeast and upper Midwest (Berkson et al., 2017, Moreton and Thomas, 2007; Fruehwald, 2016; Vance, 1987). The US versions of /aɪ/ raising do not appear to be due to any type of dialect contact—this is clear due to the widespread nature of the phonological feature, as well as the seeming gaps in places that exhibit the pattern. Yet, there is a striking similarity between not only the patterning in the US, but also to that of Canadian English speakers, which has been the focus of linguistic work since 1942 at least. Details of the differences between Canadian raising and American versions of it are described here, but the main difference is that in American raising, only /aɪ/ is raised, not /aʊ/.

Canadian raising typically refers to a phonological process in which /aʊ/ is raised to /ʌʊ/ and /aɪ/ is raised to /ʌɪ/ preceding voiceless consonants in stressed syllables, for example /ɹʌɪt/ vs /ɹaɪd/. Typical Canadian raising is a classic example of opacity, with the process applying to flaps that are underlyingly voiceless but not those that are underlyingly voiced, and specifically occurs in stressed syllables (e.g., *writer* is raised but *rider* is not) (Chambers, 1973). The very first description of Canadian raising, in which there was a "Dialect A" and a "Dialect B" found in the speech of school-aged children in Toronto, depicted a degree of variation in production of /aɪ/. Speakers of Dialect A produced the word *typewriter* as /tʌɪpɹʌɪɾɚ/, while speakers of Dialect B produced the same word as /tʌɪpɹaɪɾɚ/, crucially without raising in the second syllable (Joos, 1942). This difference in pronunciation led to the thought that while raising may have at some point been phonetically conditioned—Dialect B speakers raised only before the phonetically voiceless /p/ and not voiced /ɾ/–it eventually became a phonologically conditioned process, exemplified by the Dialect A speakers who also raised before underlyingly voiceless /ɾ/. Notably, by the time that the phenomena was written about again, thirty years later, only Dialect A, the one with phonological raising, was found in Toronto (Chambers, 1973). Since that time, Canadian raising in both /aɪ/ and /aʊ/ preceding voiceless consonants has become a ubiquitous feature of Canadian English.

The patterns of /aɪ/ raising in the US tend to surface slightly differently, often with the same opacity regarding pre-flapped environments, but also in extended environments that don't include surface-level or underlying voicelessness such as in the word *spider* in several dialects (Fruehwald, 2016; Hualde, Luchkina, and Eager, 2017). These patterns have been recently dubbed "American raising," to mark their distinctiveness and lack of /aʊ/-raising that accompanies /aɪ/ raising in true Canadian raising (Davis and Berkson, 2021; Davis, Berkson, and Strickler, 2021). Raising processes in the US have been documented in Philadelphia, Chicago, and in Michigan, among other places (Fruehwald, 2016; Hualde et al., 2017; Dailey-O'Cain, 1997). Each of these raising processes differs slightly in the phonological environments involved in the process, but common of all dialects listed above, including canonical Canadian raising, is that /aɪ/ is raised to /ʌɪ/ preceding voiceless consonants in stressed syllables, including underlyingly voiceless consonants in stressed syllables. It has been stated that, in the case of /aɪ/ raising in Philadelphia in particular, that if there were a period of phonetic raising—raising preceding surface voiceless consonants in stressed syllables but not preceding underlyingly voiceless flaps (which surface as voiced) in stressed syllables—that it had happened too quickly to observe. This claim was made using evidence obtained from apparent time data in the Philadelphia Neighborhood Corpus, in which the transition from no /aɪ/ raising to /aɪ/ raising occurred beginning in speakers born in the 1920s (Fruehwald, 2016). Notably, there is no clear point in the data set when speakers in the corpus raise phonetically, or only preceding surface voiceless consonants: when there is raising, it always occurs not only preceding stressed, surface voiceless consonants, but also preceding the stressed, pre-underlyingly-voiceless-flap. This follows the long documented phonological patterns of /aɪ/ raising as it occurs in Canada, hence the nominal "Canadian raising."

Perception of /aɪ/ raising

While a majority of work documenting varieties of American raising have focused on the production of /aɪ/ raising, documentation of perception of American raising in the US has been done in Chicago area speakers (Hualde, Luchkina, and Eager, 2017). They investigated the perceptual salience of raising in college-aged speakers from Chicago, an area which also produces /aɪ/ raising. Participants could identify a target word that was raised in a context in which raising was the only cue as to the word (i.e., pairs were words like *writing* and *riding,* which would differ only in the vowel in a dialect with /aɪ/ raising, but would be near homophones in dialects without /aɪ/ raising) at above-chance accuracy, but nowhere near the accuracy levels of the answers of "easy" words, or those with multiple phonetic cues, in words like *write* and *ride*. In other words, while listeners in Chicago were able to use the raised vowel in otherwise identical-sounding words at above chance levels, it was not at the same accuracy levels as the monosyllabic words with multiple cues in addition to the raised vowel. These results indicate that although listeners are able to utilize the raised variants in perception to an extent, they also rely heavily on other cues like vowel length and final consonant voicing to identify a word, and are unable to fully rely on vowel raising all the time.

Farris-Trimble and Tessier (2019) also studied perception of Canadian raising, although the speaker who produced the stimuli and participants were all native speakers of Canadian English, in which /aɪ/ raising is well established. They were interested in whether representations were rule-based or exemplar-based, and used the Canadian raising process as an example of rule opacity, postulating that if their participants took longer to process the raised stimuli in words like *writing,* where there is opacity in rule application, it would indicate that the processing of these words does in fact include application of the two ordered rules, rather than just one. On the other hand, they predicted that if processing time did not increase in words with the application of the two rules, it would indicate that the words are stored as exemplars. The authors did find that processing time was longer for the words with opaque rule ordering. Although this study is not directly relevant to the innovation of /aɪ/ raising as a phenomenon because it considers a dialect where pre-voiceless /aɪ/ raising is fully phonologized, it is interesting to note that even in a dialect with well-developed /aɪ/ raising, processing time is still longer in words that require multiple phonological processing such as flapping and vowel raising. Due to these results, I predict that in a dialect like Fort Wayne, with variable pronunciations and representations, the differences in perception experiment results will be even more pronounced than they are in a Canadian dialect, with much less variation.

Fort Wayne /aɪ/ raising

In Fort Wayne, there are several patterns identified as specific to this region: a phonetic pattern in which raising occurs preceding only phonetically voiceless consonants in unstressed and shortened syllables (the least advanced pattern), a phonological pattern in which raising occurs preceding phonetically voiceless consonants in both unstressed and stressed positions, as well as preceding underlyingly voiceless flaps as in *writing* (the most advanced pattern), and several intermediary steps between these two patterns of raising. There are also speakers, generally older, who exhibit no raising at all at this point, while the most advanced patterns are commonly produced by the youngest speakers so far (Berkson et al., 2017; Davis et al., 2021). This mix of production patterns within a raising dialect is unique to the Fort Wayne area, and variation of this type has not been documented as a characteristic of a raising dialect except for the initial description of Canadian raising in Toronto, published in 1942.

The introduction of raising in northeast Indiana is interesting because it has arisen seemingly without dialect contact with another variety of raising. Even in communities as nearby as Indianapolis, there is not prevalent raising, and there are no known contact routes passing through Fort Wayne to a place in the upper Midwest where raising is likely to have developed due to dialect contact. The phonetic variation in /aɪ/ in the Fort Wayne area appears to be a sound change in progress, for two reasons: first, the trend generally seems to be that younger speakers produce more novel forms than older speakers (Berkson et al., 2017); second, in other dialects in which /aɪ/ has phonetically split into /aɪ/ and /ʌɪ/, as in Canadian English, the split has remained and become a phonologized sound change rather than regressing back to a single phonological sound (e.g. Chambers, 1973). The apparent presence of both raised and unraised /aɪ/ vowels in Fort Wayne area speakers makes the community an example of conditioned phonetic and/or phonological variation, and this presumed sound change in progress. Fort Wayne is a unique place to study the progression of sound change in both speech production and perception because there are speakers at various points of progress in the sound change at the same time. If there is a way to document the parallel progressions of sound change in production and perception, an ideal place to study those progressions is in a dialect in which there is variation among speakers and listeners because it will make differences in progress between individuals more apparent than it would be if everyone in the speech community behaved the same in both production and perception studies.

In my own previous research investigating perception of /aɪ/ in Fort Wayne, including an effort to correlate production and perception behavior within individuals, I found that all speakers were able to identify monosyllabic words with no problems, regardless of their own pattern of raising in pre-voiceless /aɪ/. However, participants who did raise in underlyingly pre-voiceless flaps themselves were better at identifying those words when they included appropriate pre-underlyingly-voiceless raising, compared to speakers who did not produce pre-flap raising. Like Farris-Trimble and Tessier's listeners, the Fort Wayne raisers, who were able to perform better in the identification task compared to non-raisers, still did not identify the pre-flap words as quickly or with as much accuracy as they identified the monosyllabic /aɪ/ words. I intend to further describe the correlation between /aɪ/ raising in production and performance in perception tasks, including investigating overall sensitivity in changes to /aɪ/, identification results, and the preferences of listeners for raised or unraised /aɪ/ in pre-voiceless position, to approximate representation of /aɪ/.

Perception in theoretical sound change

Sound change is often studied only via the way that it is produced, and rarely is the effect that a changing phonological system has on perception studied. In fact, even outside instances of sound change, the relationship between speech production and speech perception within an individual is similarly rarely studied empirically–how is the way that speakers produce certain sounds related to the way that they perceive those same sounds, especially when there is systematic variation in pronunciation? At the intersection of these two topics is the relationship between speech production and perception in current sound change, where there are multiple forms of one sound produced in the same speech community. Through the lens of /aɪ/ raising as an ongoing sound change in northeast Indiana, I will consider sound change as it progresses in speech production, how that same sound change progresses in speech perception, and how the two processes work together within an individual.

The purpose of laying out the following perspectives on sound change is to acknowledge that previous research has considered the role that the listener plays in sound change, but also to point out the niche role that I hope to fill with this dissertation. My goal is to build on previous theories, rather than to propose an entirely new process of sound change. I have access to a speech community where there is a sound change in progress, which presents an ideal opportunity to study the progression of sound change as it happens in real time, in both production and perception.

The process of sound change is easily and often considered from the perspective of a speaker, both because it's easiest to describe sound change in terms of acoustic data, and to describe that acoustic data as being affected by perception. However, speech perception could and likely does shift and change during the process of language change, just like speech production. After all, all language users are both speakers and listeners; it makes sense that for an individual, both of these roles will be impacted and have the ability to impact sound change.

The following theories that are distinguished from others within the genre of sound change literature by a focus on within-speaker sound change at the cognitive level rather than community spread once the innovation has occurred within a single speaker; these theories include papers from Ohala (1981, 2012); Blevins (2004); Lindblom, Guion, Hura, Moon, and Willerman (1995); Beddor (2009); and Bybee (2003). What is generally addressed in each of the theories is how the sound change originally occurs in terms of articulation, and how that change is due to perceptual effects (e.g., misperception or to gradual shifts in pronunciation that lead to a shift in representation). The work published later than 1981 engages directly with Ohala and his theory of sound change, which at the time was unique in considering the listener's role in the process of sound change. These works do address perception in terms of the ways that perception could influence production of a sound change, but not how perception might change over the process of the sound change.

Ohala's theory is perhaps the most recognizable perspective on perception in sound change situations. The theory proposes that sound change is due to the misperception of sounds, which then leads language users to adjust their representation of those sounds, resulting in sound change. There is no motivated reason for certain sound changes to propagate throughout the speech community over others; some misperceptions simply persist as sound changes while others do not. Ohala expanded and updated this original theory in 2012, clearly defining this position that sound change is non-teleological and non-optimizing. In other words, changes due to physical constraints of the vocal tract do not occur in order to enhance perceptibility or to decrease articulatory cost, but rather as a natural result of using language over time. In this response, Ohala also distinguishes between his account that describes the *inception* of sound change without further thought given to how that incipient change might spread beyond the initial listener, and theories by other authors of sound change that are more concerned with how sound change moves through the speech community. In Ohala's theory, sound change is perpetuated as many misperceptions of everyday speech, only some of which spread through the lexicon and the speech community to become sound changes. Ohala is not clear about what factors contribute to those changes that are in fact reproduced more than once, only that the initial sound change is non-teleological and phonetically or articulatorily motivated. In the 2012 update, it is clearly stated that synchronic variation is not diachronic change, specifying that the two are separate processes, a point that sets this theory apart from the theories of both Blevins (2004) and Lindblom et al. (1995), which will both be discussed in further detail in the following paragraphs.

Blevins (2004) directly compares the process of sound change to that of biological adaptation, where the innovations arising from the mechanisms previously described that are best suited to their environment are those that spread from speaker to speaker and become sound change. The entire process of sound change is not driven towards some cause, but rather, due to natural adaptations that are evolutionarily advantaged, approaches naturally derived universals such as three and five vowel systems which are exceptionally common worldwide. This theory is quite similar to Ohala with respect to the role of misperception, but differs in that she specifies that sound change is *purely* phonetically motivated. Blevins describes an additional path to sound change that involves listeners forming different ideal representations of the same sound: this might look like a speaker with the northern cities shift developing an ideal representation of /æ/ as fronted, as compared to an American English speaker without the northern cities shift. The main difference between Ohala and Blevins is that while they share a basic premise for the mechanism of sound change originating from natural phonetic constraints that lead to misperceptions or misapplications of phonological rules, Blevins goes beyond the initial mechanism of sound change and includes the way that sound change might spread. These paths to the advancement of sound change are the ways that certain innovations are pervasive and constitute sound change as it is commonly thought of—produced by multiple speakers, on its way to becoming a part of the language itself—whereas Ohala cares about the mechanism and initiation of change only.

Similar to the way that Blevins's theory of adaptive sound change was built on Ohala's original consideration of the listener as playing a role in sound change, Lindblom et al. (1995) posited a theory of sound change that is both inclusive of Ohala's original ideas and an extension of Lindblom's earlier theory of the hypo- and hyperarticulation communication (1990). In the H and H Theory, speakers produce a variety of forms of a word in various contexts, and for the most part, listeners know this and understand that the different forms are related. While the easiest forms to perceive are hyperarticulated, the easiest to produce are a reduced or hypoarticulated form, and speakers are continually balancing these two interests. Listeners in turn understand this, and are able to interpret the hypoarticulated forms in reference to their hyperarticulated representations. Sound change occurs when the listener consistently reproduces hypoarticulated speech and lessens their hyperarticulation, and the "mutation" is accepted by the speech community, without assigning the new form to its previously-more-hyperarticulated form. This process is also referred to as "reduction," and its integral role in Bybee's model of sound change will be discussed below. Lindblom et al.'s model leaves room for the same listener misperceptions that are central to Ohala's and Blevins's models and describes equally non-teleological incipient sound change. It differs, though, in making clear that the changes that are perpetuated are those that are a good balance of both articulatorily easier to produce and perceptually distinct enough to be easily understood. There are clear parallels between this model and Blevins's model of adaptations: those changes that are best suited to their environment, be it articulatorily, perceptually, or socially, are the changes that will persist, whereas innovations that are not well-adapted will fade. Lindblom et al. are also clear that in addition to the listener misperceiving words produced by the speaker, speakers also share some agency in the perpetuation of sound change, because not only are they are the original listeners-turned-speakers who then reproduce what they've heard and potentially misinterpreted, the speakers themselves "choose" whether or not to accept the new pronunciation and reproduce it. Of course, this "choice" isn't necessarily described by Lindblom to be a conscious decision, but rather influenced by many factors that are individual to each speaker such as social influence and articulatory benefit.

Lindblom's speakers have a degree of control over their productions. By this I mean that speakers are able to vary their own productions along a continuum of hypo- to hyperarticualtion depending on the scenario, and these are the variations that eventually give rise to sound change, rather than strictly phonetically constrained variation. Lindblom et al.'s model of adaptive sound change is most similar to Blevins's model, as both propose frequent innovations, some that are pervasive "adaptations" similar to biological evolution and continue to spread from speaker to speaker as a sound change, while others are not selected and do not spread. Blevins and Lindblom et al. write that yes, sound changes begin as variations that are frequently produced, but that there is an element of natural selection in which of those micro-changes that happen within one or two speakers are spread throughout the grammars of other speakers. Lindblom et al. differ from Blevins in that they address that there is some speaker choice in whether or not to reproduce the novel forms that they have heard, based on a variety of factors that are not only articulatory, but also socially and communicatively motivated. Blevins, on the other hand, proposes that sound change innovates due to differences in pronunciation due to variation produced by phonetic constraints and are sustained when they are adaptively feasible within the sound system in which they are innovated. This difference can in fact be categorized by a difference in what exactly contributes to the selection of certain variants over others: Lindblom et al. believe that there are many factors that contribute to the viability of innovation, including choice on the part of the speakers.

Beddor (2009) focuses on the role of coarticulation in sound change. More specifically, her claim is that coarticulatory features of certain sounds can become a primary indicator for those sounds over time. While this theory is articulatorily motivated, and therefore directly concerns production, a wider view of the theory includes perception as well. Cues are perceptual, and if speakers reduce cues in instances where there is more than one cue, and then listeners learn to listen for a different primary cue, this is an instance of perceptual sound change. Her main example of coarticulatory cues that are used in place of other acoustic cues is that for some speakers in Michigan, coarticulatory anticipatory nasality on a vowel preceding a nasal consonant followed by a voiceless consonant is becoming the primary indication of nasality in the word (i.e., *bent* is pronounced as [bɛ̃t] with little to no nasal consonant). There are other possibilities that could follow the same pattern, such as vowels with raised F0 following voiceless consonants, which later turn to tonal cues. The mechanism for sound change to occur is coarticulation produced by the speaker and interpreted by the listener. While this may sound similar to Ohala, Blevins, and Lindblom et al., it is without the key point of misinterpretation on the part of the listener. While Beddor allows for the possibility of listener misperception of the acoustic signal, the main point of the theory is that speakers and listeners have knowledge about the usefulness of coarticulation to the listener in interpreting the speaker's signal. The listener doesn't necessarily filter out the coarticulation in the signal, as in the theories of Ohala and Blevins--the listener actively uses the coarticulatory signal as additional phonetic information, rather than misinterpreting coarticulatory effects and reproducing that misinterpretation in future utterances. She holds the view, unique among the authors discussed here, that listeners make active use of coarticulatory details when parsing acoustic signals, rather than being fairly passive participants. While this is a more specialized theory of sound change, in that it focuses on coarticulation, it gives the listener an independent and somewhat agentive role in sound change. Lindblom et al. also give their listeners a type of agency, albeit not in the interpretations of the signal or the choice to reproduce it, but rather in the option to hypo- or hyperarticulate a word depending on the situation. Lindblom et al. also leave it up to the other listeners in the speech community whether to accept the innovations they hear from other speakers or not to; while Beddor doesn't specifically state this view, it's not incompatible with her theory that listeners are able to use coarticulatory information in the speech signal, regardless of the extent of their own propensity to use coarticulatory features in their own speech.

Bybee's (2003) theory, termed usage-based phonology, is another contrasting view to the theories of Ohala and Blevins because it does not rely on misinterpretation of speech to advance sound change, though attributing a phonetic input to a representation not intended by the speaker would affect that representation very slowly over time, and be unlikely to happen often enough to have any lasting effect. Bybee also does not focus on the listener as an overall source of sound change. While the theory doesn't directly implicate the listener in the initiation of sound change, though, experiences that a listener has do directly affect their representation; therefore language users playing the role of listener do contribute to overall sound shift. Because of her reliance on exemplar representations, Bybee doesn't describe a single initiation point or mechanism that causes sound change; she rather describes an inevitable trend in language as a whole towards reduction. Bybee's exemplar explanation could be applied to Beddor's example above of nasal consonant deletion: if a word with a nasal consonant preceding a voiceless consonant (like *went*) is frequently produced, speakers will consistently reduce the phonetic form of the word, potentially until the nasal consonant is no longer produced as part of the word. Each instance of a reduced form of the word, in this case a nasal vowel with less and less evidence of nasal consonant, affects the speaker's mental representation and storage of the word. The idea that forms reduce more and more with continued use due to the assumption on the part of the speaker that the listener will be able to accurately perceive a reduced form, relates to Beddor's tenet that listeners are able to use coarticulatory information in the acoustic signal where it exists. Beddor implies that speakers understand that listeners have an understanding of coarticulatory effects, and are therefore not discouraged from coarticulating if it is articulatorily easier. Bybee, like Beddor, gives credit to speakers' and listeners' abilities to make use of information outside of the bounds of phonological representation, namely coarticulatory and reductive effects.

In this section, I made some comparisons between the ways that each author considers the mechanism of sound change. Ohala and Blevins directly implicate misperceptions as an initiation point of sound change, with slightly differing details about how these misperceptions might occur and whether they are a possible path to sound change or the only path. Beddor and Lindblom both give a certain level of agency to their speakers and listeners, though don't actually purport that sound change is intentional on the part of the speakers; this point is compatible with Blevins's explicit statement that sound change is non-teleological (occurs without purpose or design). Beddor and Bybee both describe a relentless and constant progression towards sound change, that allows for listeners some agency in incorporating information from the acoustic signal to innovative sounds that they later produce. In all of these theories, synchronic variation is a piece of the mechanism that initiates sound change, whether that be due to predictable variation like coarticulation or a mispronunciation or misinterpretation of the variation by the listener. The differences discussed here pertain to how that synchronic variation becomes sound change, including the perspective of the listener. One piece of the sound change puzzle that all of these theories agree on is that diachronic variation originates from synchronic variation, and I will use that component of theory to ground my research in Fort Wayne, that the patterned, synchronic variation in /aɪ/ is the start of diachronic sound change. /aɪ/ raising in Fort Wayne mirrors other varieties of /aɪ/ raising, which represent change that has permeated the lexicon in other dialects, such as Canadian English.

In all of these theories of sound change that center perception as the mechanism of sound change, speech sounds can change for a variety reasons, from making the pronunciation of the words easier (like the simplification of complex consonant clusters, e.g., Blevins, 2004) to making the words easier for the listener to perceive (in the way that a hyperarticulated form is generally understood to be easier to perceive, e.g., Lindblom, 1990) to a total misperception of the speaker's intended message (e.g., Ohala, 1981). What these theories address, though, is more rooted in how perception of speech might affect production that leads to sound changes, rather than how the process of speech perception changes over the course of a sound change. This section has focused on several theories of sound change and their inclusion of both the talker and the listener in the initial point of a sound change. I've referred to this starting point as the "mechanism," which for most of these theories may or may not develop into actual sound change in the language. These theories of sound change are included because they hypothesize about what happens to speakers' pronunciation of sounds during a sound change because of what happens to their perception of that same sound–in other words, the theories of sound change that actually address perception describe perception as a process that influences speech production, rather than a process in its own right. This is the necessary distinction of sound changing *for* a listener: I would like to consider speech perception in a sound change as a process that is subject to change, just as production is.

The relationship between production and perception in diachronic change

The process of a change in perception of a speech sound could also be described as a change in the grammatical representation of that speech sound. A person's representation of a speech sound affects both their production and perception of that speech sound, but representation isn't always obvious from a speaker's production alone. For example, a speaker with a raised pre-voiceless /aɪ/ in production could be producing that raised /aɪ/ for a number of reasons, including a representation that includes a raised pre-voiceless /aɪ/, but that raising could also be due to coarticulatory, phonetic influences like duration shortening on the vowel, and might not be related to that speakers' best representation of that vowel.

Representations of speech sounds are described in many different ways, but I'm using it here in a loose sense to describe the "target" productions of speakers, as well as the phonetic features of vowels that aid in speech perception. This research isn't designed to indicate whether speakers in Fort Wayne have an exemplar representation of /aɪ/ or whether there is an underlying phoneme with phonological rules acting upon it. Rather, my hope is to consider that perceptual representation of /aɪ/ is subject to change just as the target of the produced vowel is subject to change, and to compare the progress of the two processes. The studies discussed below focus on the intersection between representation and sound change, and how a change in production and a change in perception are related within speakers. In this research, I will consider how production informs representation and also how perception informs representation, and how the two together form a complete picture. I'm interested in both production and perception in sound change, the way that these processes rely on representation, and whether they seem to change in parallel.

In a study of diachronic sound change, Sankoff (2018) presents data from the same and comparable participants in Quebec in both 1971 and 1984. While she didn't exactly investigate raising, this is a very valuable resource, as many studies of sound change make use of apparent time measures, and her findings can be used to determine whether the current process of using apparent time is an accurate representation of how sound change actually progresses over real time. She describes three changes in progress in Quebecois French, and three different ways that these changes seem to be happening. The phonological change that she describes, [r] → [ʀ], seems to closely parallel the situation in Fort Wayne as it has been described so far in terms of progression. The data collected in 1971 indicate that older speakers produced the less advanced [r] form, whereas younger speakers tended to produce the more advanced [ʀ] form. During the second round of data collection in 1984, the researchers were able to locate several of the same speakers, and match new speakers on many social features. In this particular change, the speakers maintained the stage of the change that they had exhibited in the earlier data collection. This is a strong comparison of synchronic data to diachronic data, and indicates that synchronic demonstrations of diachronic phonetic or phonological change are valid evidence of sound change. This is the same way that I will be using the synchronic variation in Fort Wayne /aɪ/ as indicative of a larger pattern of diachronic change.

Instances of diachronic change have been used as a window into representation before, and have also proved fruitful environments for comparison of production and perception within a single speaker. Researchers investigated a near-merger of tense and lax vowels in Utah English, which presented specifically before [ɫ] in /ʊ/ and /u/ vowels (DiPaolo and Faber, 1990). The authors described this as a sound change in progress, with younger participants participating in more progressive (merged vowels) forms of the merger and older participants presenting more conservative (unmerged) forms. A vowel categorization production task involved participants reading a short wordlist with the same vowel, the experimenters pointing out to the participants that all of the words had the same vowel, and then the participants read another wordlist with other distractor items including the original vowel. The participants were then asked to categorize their own speech, matching (or near-matching) the vowels that they had produced with those in example words. In the perception task, participants had a forced-choice task in which they decided which word they had heard between minimal pairs formed by contrasting the tense and lax vowels (i.e., *pull* vs *pool*). The most interesting results in this study are that the researchers interpreted the results of the production categorization experiment as the older generations (the younger generation's parents and grandparents) have targets that are merged, but productions that are not, hypothesizing by extension that perception precedes production in situations of sound change. This is not what I found in preliminary work asking a similar question for /aɪ/ raising in Fort Wayne; albeit through a different task, I found that my participants who exhibited /aɪ/ raising in production were in fact better at choosing the correct word in a forced-choice perception task that were minimal pairs only by the distinguishing factor of /aɪ/ raising (*writing*) vs. non /aɪ/ raising (*riding*) (Strickler, 2019). In the Chapter 4 of this dissertation, I will further investigate correlations between individuals' production and perception of /aɪ/, by way of tasks designed to reveal listeners' sensitivity to small changes in /aɪ/, use of raised /aɪ/ in word identification, and their preferences for raised or unraised /aɪ/ to compare to DiPaolo and Faber's vowel categorization task.

In another experiment investigating the correlation between participants' production of coarticulatory features and perceptual compensation of those same features, Yu (2019) describes a modest correlation between his speakers' productions and perceptual compensation of /s/ and /ʃ/. The less distinguished the individual's productions of the two phonemes were, or the least amount of difference between /s/ and /ʃ/ they produced, the less certain they were about which sound they had heard in a forced choice perception probing the same phonemes. A large part of Yu's analysis is pointing out that there is a lot of variability not accounted for using the perception-production correlation, and lists some possible other sources including gender variability and vowel differences. The takeaway from this study is that there is a correlation between how categorically the speakers produce /s/ and /ʃ/ and how confident they are in filtering out coarticulatory information when perceiving /s/ and /ʃ/. This predicts a direct link between the variation that the speakers produced and their perception of variation on the same consonant, and also predicts that for my Fort Wayne listeners, there will be some correlation between their production of /aɪ/ and their perception of /aɪ/.

I am using the ongoing change in the Fort Wayne community as an opportunity to study individual representation change as the sound change (/aɪ/-raising) progresses, during this brief transitional period in which speakers are producing different patterns of the change. This situation affords me a unique opportunity to explore the relationship between a speaker's production and perception of a sound change in progress. This will be a valuable contribution to both the understanding of the relationship between production and perception and an individual's experience of perception in sound change.

Production and perception of /aɪ/ raising in Fort Wayne

This dissertation depicts a sound change as it progresses through both the production and perception of speakers in Fort Wayne. It is not yet clear how closely these two processes parallel each other, although it has been generally assumed that perception of a sound change precedes production, and this has been shown empirically in a few studies (DiPaolo and Faber, 1990; Beddor 2009). In the instance of /aɪ/ raising in Fort Wayne, the results of the dissertation will be a combined description of production, perception of production, and the extent to which listeners utilize raising as a perceptual tool. The results as a whole are a complicated picture of two dynamic processes interacting. There are fairly clear stages in production that have been identified in Fort Wayne raising, but we know nothing about the possible stages of perceptual change, if there are any at all.

These questions are addressed in two parts, production and perception of /aɪ/ by Fort Wayne speakers. Each part includes several experiments. The two production tasks include data in the form of a wordlist, to identify which pattern of raising each participant exemplifies; a second task will ask participants to produce clear speech in addition to citation speech. In Chapter 2, I walk through several methods of measuring raising in /aɪ/, and use generalized additive mixed models to classify speakers using both raising and fronting patterns in pre-voiceless /aɪ/. The clear speech experiment described in Chapter 3 demonstrates hyperarticulation compared to more casual speech; this comparison determined that participants are not striving to achieve a raised /aɪ/ target in hyperarticulated speech, but are rather treating the two component parts of the diphthong as though they are /ɑ/ and /i/, and enhancing the diphthong by lowering /ɑ/ in the first half of the diphthong and raising and fronting /i/ in the second half of the diphthong.

In Chapter 4, the results to the three perception tasks are discussed. The vowel-specific formant sensitivity task, or 4IAX task, determined that listeners could accurately hear the types of differences in raised versus unraised /aɪ/ that are produced in Fort Wayne. The identification task revealed that although listeners can hear the small differences in pre-voiceless and pre-voiced /aɪ/ as demonstrated in the 4IAX task, these differences do not help them identify words. Finally, the preference task results indicated that listeners did not prefer a raised /aɪ/ over an unraised /aɪ/ in pre-voiceless position, providing evidence that these listeners do not have a representation that includes a raised /aɪ/. Overall in the perception tasks, listeners were sensitive to vowel and consonant pairings that are not produced in normal speech, such as a raised /aɪ/ in pre-voiced position, but did not demonstrate differences between the vowel and consonant pairings that sounded acceptable to listeners, such as raised and unraised /aɪ/ in pre-voiceless position.

I am using the ongoing change in the Fort Wayne community as an opportunity to study individual representation change as /aɪ/-raising progresses, during this brief transitional period in which speakers are producing different patterns of the change. This situation affords me a unique opportunity to explore the relationship between a speaker's production and perception of a sound change in progress, and to explore whether there are targets that differ with respect to the varying productions. These results are a valuable contribution to both the understanding of the relationship between production and perception and an individual's experience of perception in sound change.

Chapter 2
Production

Introduction

This chapter describes /aɪ/ raising produced by 21 speakers from Fort Wayne, Indiana. This ongoing sound change in progress has been termed "American raising," in comparison to its counterpart Canadian raising. Canadian raising describes raising in the /aʊ/ vowel in addition to raising in /aɪ/; American raising varieties typically only affect /aɪ/, which raises to [ʌɪ] when it is followed by a voiceless consonant.

In this chapter, three methods to analyze /aɪ/ raising in the current study are discussed. First, the most common method of comparing F1 at the midpoint of the nucleus of the diphthong and using a threshold of 60Hz to determine whether there is a difference between pre-voiced and pre-voiceless /aɪ/ is a typical measurement used to determine raising (e.g., Labov, Ash, & Boburg, 2006; Berkson, Davis, & Strickler, 2017). This method considers only a small portion of the vowel, and only F1. This relatively simple approach leaves out much of the detail at other points in the vowel and in F2. To include more of the vowel, the second analysis method calculates the trajectory length of the vowel, or the sum of the Euclidian distances across the vowel between F1 and F2 which has been used to compare diphthongs generally (Fox & Jacewicz, 2009). This method includes more of the vowel in the analysis, but the results are still not highly detailed and are not as useful in classifying individual speakers' production patterns. Finally, the third method of analysis is to use generalized additive mixed models (GAMMs) to predict the formants of /aɪ/ for each speaker individually. GAMMs have been used to analyze /aɪ/ raising previously (Hualde, Barlaz, & Luchkina, 2022; Mielke & Thomas, 2021). GAMMs have the benefit of allowing for predictive analysis over time, with the ability to determine the significant difference between the pre-voiced and pre-voiceless formants when they occur during the vowel, rather than a broader measure that considers the entire vowel as a whole.

In this chapter, I argue that /aɪ/ "raising" is really more like /aɪ/ raising + fronting, because F2 is often just as involved in the differences between pre-voiced and pre-voiceless /aɪ/ as F1 is. The majority of the speakers in this study produce a raised and fronted /aɪ/ in pre-voiceless context in monosyllabic words, and a smaller subset of those produce a raised /aɪ/ preceding underlyingly voiceless flaps (i.e., in *writing*), which is the most advanced presentation of the sound change as described in the previous chapter, section "Fort Wayne /aɪ/ raising." Within those speakers who produce significant raising and/or fronting in monosyllabic, pre-voiceless /aɪ/, there are patterns suggesting different possible origin points within the vowel for that split between [aɪ] and [ʌɪ], providing evidence for a variety of approaches that speakers use to achieve that raised and fronted pre-voiceless /aɪ/.

Methods

There are 21 total participants in this chapter, with data collected in one of two study protocols. In a first round of data collection, fourteen participants completed the production portion of the experiment only. These data were collected on the platform phonic AI. Participants were asked to use headphones or earbuds with an attached microphone, and to read along to a video that showed the words to be spoken aloud one at a time. Participants completed recording and then consented to be contacted again to complete the perception experiments at a later date. Study retention rate using this method was low, so to obtain additional perception data, a second round of data collection was completed in which a short production task was included at the same time as the perception tasks. Seven additional participants recorded spoken word lists on the online experiment platform Gorilla, just before they completed the three perception experiments, and again were encouraged to use headphones with an attached microphone while reading along to a video showing words appear one at a time on-screen. Including both of these rounds of data collection, there are 21 participants' speech analyzed in this chapter. Each speaker produced 109 words, including a variety of words with /aɪ/ in both stressed and unstressed position, both monosyllabic and disyllabic words with flapped consonants, and distractor words with monosyllabic /ɑ/, /i/, and /i/. These words are listed in their entirety in Appendix 4.

The sound files were first labeled using an online forced-aligner (Kisler, Reichel, & Schiel, 2017) and then vowel boundaries were hand-corrected for accuracy. A Praat script was used to find formant measurements at 10 time-normalized points for each vowel, and the vowel duration. Measurements in Hertz were converted to Bark using the formula $freq_{bark}=(26.81/(1+(1960/freq_{Hz})))-0.53$ (Traunmüller, 1990).

Duration differences

The vowels in this study were analyzed in time-normalized divisions, with ten evenly spaced time points over the course of the vowel. Of course, especially in monosyllabic words, there is a well-documented difference between vowel length preceding voiceless consonants compared to vowel length preceding voiced consonants (Peterson & Lehiste, 1960). In this section, those vowel length differences are described, and for the rest of the chapter, vowels are analyzed using ten time-normalized time points.

Similar to previous research in Fort Wayne, the duration differences between the pre-voiced and pre-voiceless /aɪ/ is greater in monosyllabic words than in disyllabic words. Overall, the average duration of monosyllabic, pre-voiced /aɪ/ is 246 ms (n~14 words per participant) and the average duration of monosyllabic, pre-voiceless /aɪ/ is 131 ms (n~16 words per participant). The average duration of disyllabic, pre-flap words is 134 ms for underlyingly voiced flaps (words like “riding” and “biding”; n~4 words per participant) and 131 ms for underlyingly voiceless flaps (words like “writing” and “biting”; n~4 words per participant). Some speakers did not have a full set of words to analyze due to difficulties recording, but the participants who recorded full word lists recorded the number of words specified above. For the monosyllabic words, the voicing of the following consonant significantly predicted vowel duration, with random intercepts for speakers ($t=-91.85$, $p<0.05$). For the disyllabic, pre-flap words, the underlying voicing of the following consonant did not significantly predict vowel duration, although there is a marginal effect indicating a pattern of slightly shorter vowels preceding underlyingly voiceless flaps; this model again includes random intercepts for speakers ($t=-2.276$, $p=0.02$).

Difference in Hertz between pre-voiced and pre-voiceless F1 at the nucleus of the diphthong

Background

Canadian raising (predating American raising, but with a nearly identical phonological outcome) has historically been measured in the nucleus. Researchers have previously defined Canadian raising as a difference in the height of the nucleus of the diphthong in pre-voiceless compared to pre-voiced vowels. This focus on nucleus height is apparent even from the way that raising is transcribed ([aɪ] as opposed to [ʌɪ]), where the first vowel noted in the diphthong changes from a low vowel to a mid vowel, and the second vowel in the diphthong does not change at all. This is also true in other notations of the diphthong, as in [aj] and [ʌj]. The American versions of raising have come to be defined in the same way, with most attention being paid to the nucleus of the diphthong.

This raising process was first described using formant measurements by Labov, Ash, and Boburg in the Atlas of North American English (2006). Here, raising was described as a 60 Hertz difference in F1 between the pre-voiced /aɪ/ and the pre-voiceless /ʌɪ/, where the nucleus of pre-voiceless /aɪ/ was 60Hz **lower** than that of pre-voiced /aɪ/, resulting in a higher vowel (a lower F1 corresponds to a higher vowel). This was a fairly descriptive study with little statistical justification for choosing a difference of 60Hz as the threshold that defines raising. This method has been used by other researchers to classify /aɪ/ as raised or unraised, including in all previous work in Fort Wayne (Berkson et al., 2017; Davis et al., 2021). This method of measuring raising only considers the nucleus of the diphthong.

Analysis

The simplest way to measure raising is by using a simple difference between the F1 measurement at the nucleus of the diphthong in pre-voiced /aɪ/ and the same measurement in the nucleus of the diphthong in pre-voiceless /aɪ/. As in Berkson, Davis, & Strickler (2017), the pre-voiceless-consonant /aɪ/ vowel (n~16) and the pre-voiced-consonant /aɪ/ vowels (n~14) were averaged at 10 time-normalized points across the vowel, and at timepoint 3, if the difference is greater than 60 Hz, the speaker was considered a "raiser." The clearest context to consider these vowels is in monosyllabic words, where raising tends to appear first; in this section, only the monosyllabic words will be considered for analysis, but more morphologically complex words will be included in later analyses in this chapter.

Using this method of classification, only one out of all twenty-one speakers in this study could be considered a raiser. This speaker's average formant tracks across F1 (the two lines along the bottom of the graph) and F2 (the two lines at the top of the graph) are shown in Figure 1. The only difference measurement analyzed here is within the black box.

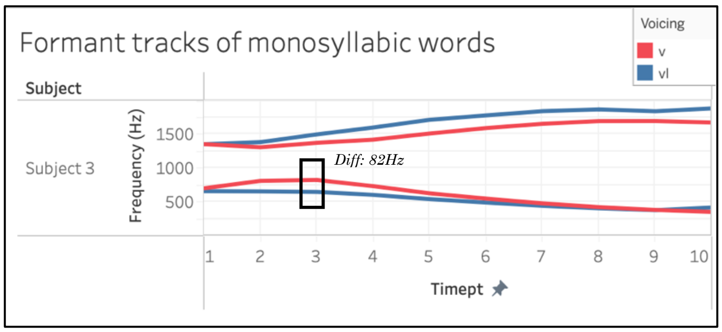


Figure 1 Formant tracks of monosyllabic words, speaker with >60Hz

Although there is only one speaker with an average difference of greater than 60Hz between pre-voiced /aɪ/ and pre-voiceless /aɪ/ at the third time point, four other speakers have a difference that is greater than 50 Hz at that same point; these speakers' formant tracks are pictured below in Figure 2.

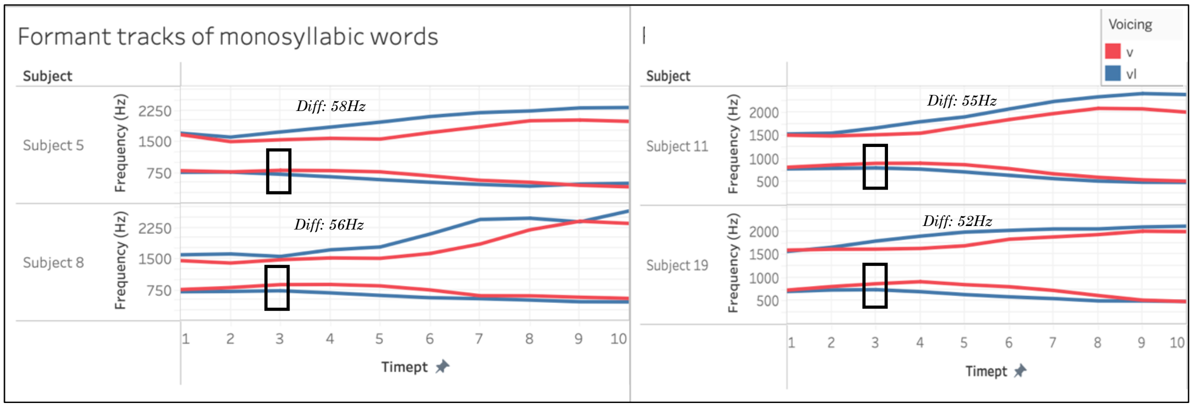


Figure 2 Figure 2 Formant tracks of the four speakers producing an average difference greater than 50Hz in F1 at timepoint 3

There are also five speakers with differences greater than 20Hz comparing pre-voiced to pre-voiceless /aɪ/ in F1 at timepoint 3 in monosyllabic words, depicted below in Figure 3.

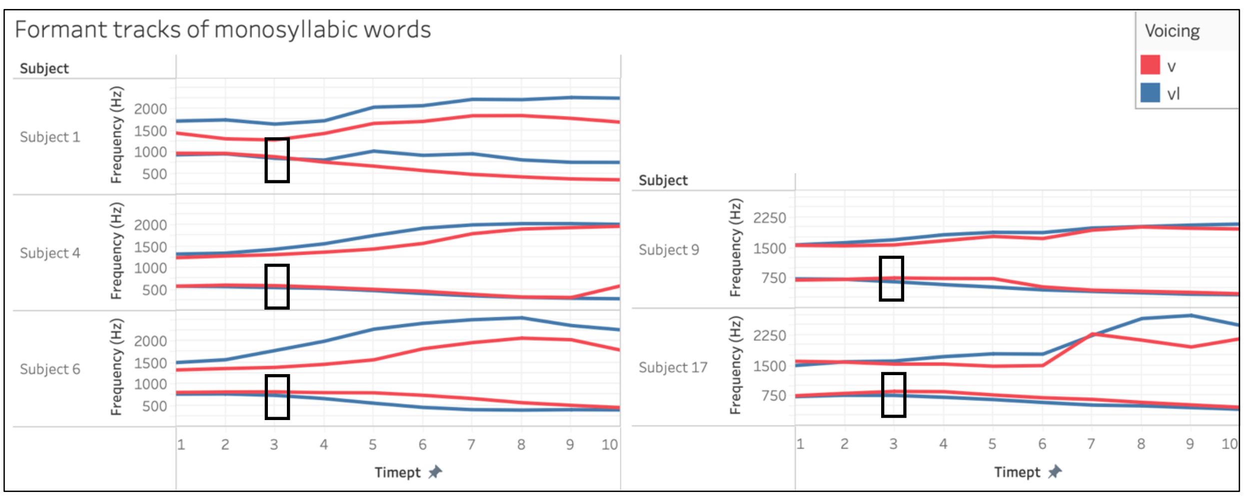


Figure 3 Formant tracks of the five speakers who have an average difference of greater than 20Hz in F1 at timepoint 3

Finally, there are eleven speakers who produce a difference of less than 20 Hertz at timepoint 3, shown below in Figure 4.

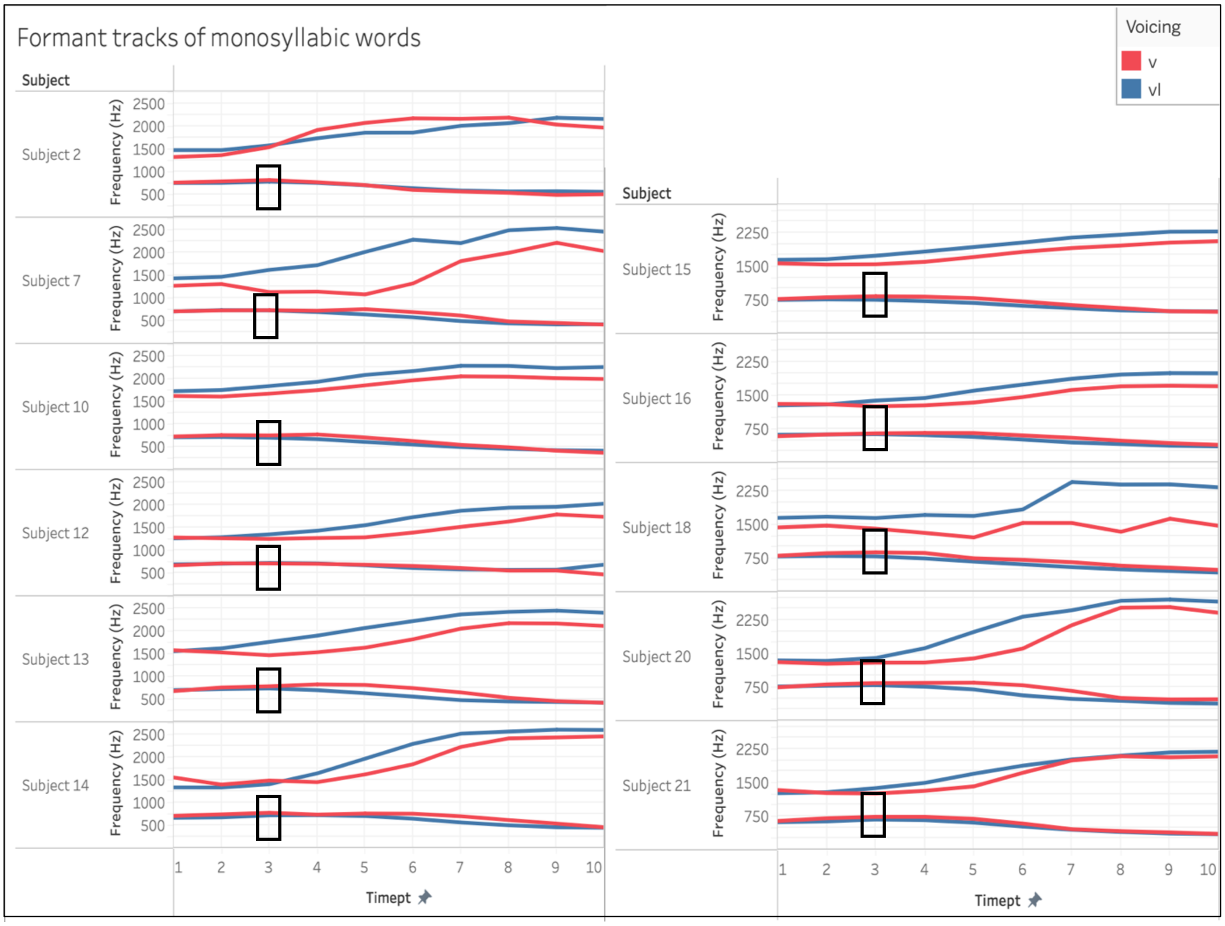


Figure 4 Formant tracks of the eleven speakers who produce <20Hz difference in F1 at timepoint 3

In comparing the rest of the formant measurements of F1, and the entire tracks of F2 in these vowels, there is a lot of variation between the pre-voiced and pre-voiceless /aɪ/ vowels that is not captured in the single difference measure in the nucleus of the diphthong. There are three points where this limited analysis of F1 differences at timepoint 3 fall short.

First, there are F1 differences at timepoints other than timepoint 3, or at the nucleus of the diphthong. Timepoint 3 was used for this analysis because it's the same point in the vowel that has been used in previous descriptions of raising in

Fort Wayne (Berkson et al. 2017). This choice of time-normalized point is also directly comparable to other studies that choose to analyze the maximum F1 value rather than choose a specific time point in each vowel (e.g., Moreton & Thomas 2007). Timepoint three captures the maximum F1 value, while also corresponding with the nucleus of the diphthong. However, there are individuals who produce visible differences between pre-voiced and pre-voiceless vowels at timepoint 4 or 5, which is also in the nucleus of the diphthong but not at the maximum F1 value, who are not included in this choice of timepoint analysis.

Further, there is a range of pre-voiced – pre-voiceless differences in the production of F1 from 20 Hz to 60 Hz, not a seemingly categorical split between differences greater than 60Hz and less than 60Hz, so using that cutoff does not describe the data at hand.

Finally, there are also visible difference in F2 between pre-voiceless and pre-voiced /aɪ/. In fact, in the above plots where the formant tracks are plotted in Hertz, these differences appear far larger than the differences in F1.
Some of these large differences are reduced, however, when plotting the formant measurements in Bark, as seen below. Plotting in Bark allows for an easier comparison of F1 to F2 because Bark is roughly a reflection of perceptual distances rather than physical distances between two measurements, as in Hertz. Humans are much more sensitive to differences at lower frequencies (i.e., where F1 usually occurs) than they are sensitive to differences at higher frequencies (i.e., where F2 usually occurs). In other words, in F1, differences plotted in Hertz will appear

smaller than the same differences plotted in Bark, while in F2, differences in Hertz will appear larger than the same differences plotted in Bark. For example in the formant tracks below, the same speakers' vowels are plotted in Hertz on the left and in Bark on the right. The differences in F2 in Hertz appear smaller in Bark, and the differences in F1 in Hertz appear larger in Bark; for Subject 13, the size of the differences in F1 and F2 look very different in Hertz, and much more comparable in Bark, and are also more comparable in speech perception when plotted in Bark. Therefore, Bark will be the measurement used to plot formants for the rest of this chapter. In Figure 5, there is a comparison between one individual's formants plotted in Hertz on the left and in Bark on the right. Note that the differences that appear large in F2 in the Hertz plot look smaller in the Bark plot, and the differences in F1 around timepoints 4-7 in the Hertz plot look more comparable to the F2 differences in the Bark plot.

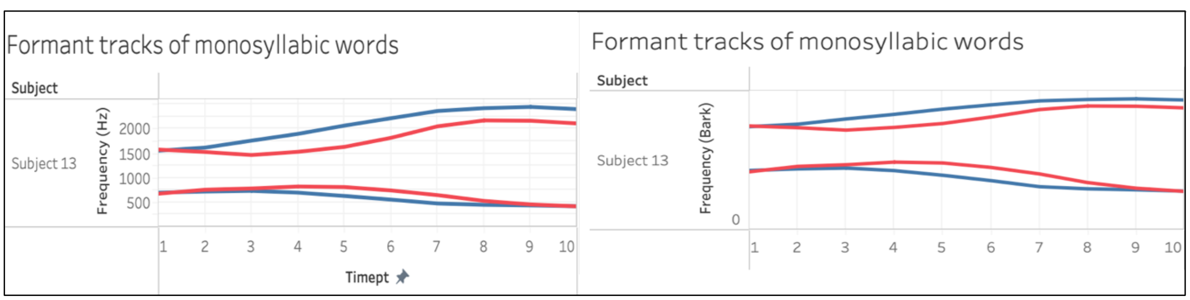


Figure 5 Subject 13 monosyllabic words, plotted in Hertz and Bark

Vowel trajectory length

Background: more inclusive studies of /aɪ/

Previous research has also included F2 in analyses of /aɪ/ in dialects where /aɪ/ raising also occurs. These studies note evidence of fronting in pre-voiceless context, as well as the raising that is apparent in the name. This fronting is sometimes described as a measure of offglide peripheralization (e.g., Moreton, 2004). Hualde et al. (2017) note that there is both nucleus raising and offglide fronting in pre-voiceless /aɪ/ in Chicago area speakers. Hualde, Barlaz, and Luchkina (2022) go as far as to suggest that the difference between the pre-voiced and pre-voiceless diphthongs should be transcribed as [aɪ] and [ʌi], respectively, because the fronting in the offglide is as prevalent as the raising in the nucleus. Offglide peripheralization is described for eastern Ohio speakers in Moreton & Thomas (2007) and Mielke & Thomas (2021). Moreton & Thomas describe /aɪ/ raising in Ohio, and suggest that the raising process may start as an offglide peripheralization process, which is common in vowels preceding voiceless consonants, as is nucleus shortening in diphthongs preceding voiceless consonants. Mielke & Thomas also describe speakers in Ohio beginning to peripheralize the second half of the diphthong before they raise the nucleus of the diphthong. In Mielke & Thomas (2021), this is described as an "unzipping" effect, in which speakers with the least separated pre-voiced and pre-voiceless /aɪ/ vowels will have a raised and fronted /i/ in the diphthong, and the distinction between the two phonological environments becomes more pronounced earlier in more advanced "raisers," moving backwards throughout the vowel. Both the Moreton & Thomas and Mielke & Thomas studies consider at the very least both parts of the diphthong, the nucleus and the offglide in their analyses, rather than only the nucleus.

There is evidence in previous research conducted on raising dialects outside of Fort Wayne that /aɪ/ raising often affects the entire vowel (both the nucleus and the glide), and at the least co-occurs with a degree of fronting as well as raising. Due to this evidence, in the next analysis, we will move from analyzing single points to the entire tracks of both F1 and F2.

Vowel trajectory length analysis

Fox and Jacewicz (2009) describe two ways of measuring vowel formant trajectories in dynamic vowels between different regional dialects (speakers in Wisconsin, Ohio, and North Carolina). For each vowel, they measure five equidistant time points throughout the vowel and use these to calculate vector length (VL) and trajectory length (TL); they also include spectral rate of change with these measures, to account for the amount of formant change over time. Fox and Jacewicz calculated the trajectory length for four vowel sections, one between each of their five measurements; they calculated Euclidian distances between timepoints using F1 and F2 across the length of the vowel, and then summed those four distance measures to calculate a total summed distance across the vowel. They called this measurement Trajectory Length (TL), and it's the measurement that will be used in this section.

$$VSL_n = \sqrt{(F1_n - F1_{n+1})^2 + (F2_n - F2_{n+1})^2}.$$

The overall formant TL was then defined as a sum of trajectories of four vowel sections:

$$TL = \sum_{n=1}^{4} VSL_n.$$

Figure 6 Trajectory length calculation, from Fox and Jacewicz (2009)

Trajectory length measures make the most sense for the purposes of analysis of /aɪ/, because they include the entire trajectory of the vowel, rather than the vector length calculation, which measures the distance between two points. For the Fort Wayne speakers, measurements were taken from ten timepoints across the diphthong; therefore, our measure of formant trajectory change will include nine sections of vowel, starting from the beginning of the vowel and continuing to the end of the vowel.

Using the same formant measurements (in Hertz) as in the previous section, and only considering the monosyllabic pre-voiced and pre-voiceless /aɪ/ vowels, the distance measurements for each individual word by speaker were calculated. A regression analysis of the trajectory length numbers by following consonant voicing with random intercepts for speakers did not find any significant differences between pre-voiced and pre-voiceless /aɪ/ when the TL measurement included the entire vowel, or the sum of the vector lengths of all nine sections (estimate=$1.99e^5$, t value=1.437, p=0.15).

There were, however, significant differences when considering the nucleus and coda of the vowel in separate models. The sum of the vector length measurements calculated from the first five timepoints were used to calculate the trajectory length of the nucleus, and the sum of the vector length measurements calculated from the last five timepoints were used to calculate the trajectory length of the coda. The nucleus and the coda have been documented as doing different things in the case of raising (e.g., Thomas & Mielke, 2021; Moreton & Thomas, 2007). Looking at the formant tracks of the vowels, it does appear that there is more separation between the pre-voiced and pre-voiceless formant tracks in the first half of the vowel compared to the second half of the vowel. When the trajectory length of the nucleus of each word was regressed on the voicing of the following consonant, with random intercepts for speakers, there was a significant prediction of a longer trajectory length for pre-voiceless /aɪ/ (estimate=1.991e^{05}, t=3.082, p=0.002). The same model predicting the trajectory length of the coda using the voicing of the following consonant as a main predictor with random intercepts for speakers has no significant results (t = -0.010, p=0.99163). A larger trajectory length prediction indicates that for the nucleui of pre-voiceless /aɪ/, there is more distance between F1 and F2 at each point; in practice, this looks like a lower F1 value (a higher vowel) and also a higher F2 value (a fronter vowel).

These results indicate that there are statistically significant differences between the formant tracks of the monosyllabic pre-voiced and pre-voiceless /aɪ/ vowels, when including both F1 and F2 in the measurements that lead to the prediction. These differences are significant in the nucleus, but not in the coda, which points to the raising described in previous research that originates in the nucleus in the Fort Wayne sound change, rather than the speaker in Ohio who seem to originate the sound change in the offglide.

This model also includes the formant tracks of all speakers in one model, which is a good step towards describing that there are community-wide effects of the sound change in Fort Wayne, but doesn't address the individual differences that have been previously documented in Fort Wayne and that are of particular interest in this study. Linear models are not typically run with data split by individual speakers, because there's not enough variation in measures when all of the values come from a single speaker. In the next section, speakers will be considered individually using generalized additive mixed models, which is a statistical technique that has been previously used to analyze individual speakers (e.g., Thomas & Mielke 2021).

Generalized Additive Mixed Models

Background

Generalized Additive Mixed Models (GAMMs) are used to analyze data that occur on a timeseries, with non-linear slopes and can include both continuous and categorical variables (Sóskuthy 2017). GAMMs are also useful for capturing non-linearity, as in a typical dynamic formant track, and can capture variation as it occurs along the formant track, rather than focusing on a single point. Previous research has used (Generalized Additive Mixed Models) GAMMs to analyze vowel formant trajectories (e.g., Renwick & Stanley, 2020) and specifically in analyzing

/aɪ/ diphthongs for American raising (Thomas & Mielke, 2021; Hualde, Barlaz & Luchkina, 2022).

Hualde, Barlaz, and Luchkina (2022) describe /aɪ/ raising in Chicago and the surrounding area. Recently, they compared GAMMs to the 60hz in F1 measurement technique in analyzing /aɪ/ raising. They used GAMMs to compare /aɪ/ in pre-/t/ context to /aɪ/ in pre-/d/ context, and another comparison for pre-t-flap /aɪ/ and pre-d-flap /aɪ/. They considered only alveolar stops as the following consonant context to reduce some of the variation due to coarticulation of other places of articulation, and used Hertz formant measurements for the GAMMs, and time-normalized data. They specifically note in their discussion section that a takeaway from their paper should be that F2 also matters in determining raising because it increases (indicating a fronter vowel) as F1 decreases (indicating a higher vowel). Hualde, Barlaz & Luchkina also note that while diachronic sound change observation was not their goal, their data support the theory that monosyllabic /aɪ/ raising precedes pre-flap /aɪ/ raising, because their participants demonstrate more separation in monosyllabic /aɪ/ in the pre-voiced vs pre-voiceless vowels, compared to the differences in pre-flap /aɪ/.

By comparison, Thomas and Mielke (2021) as described in the previous section use GAMMs to document /aɪ/ raising in Ohio speakers. They use normalized Bark measurements in the models, and analyze the GAMM plots of individual speakers. This is the method that I will take in the following analysis section.

Analysis of individuals using GAMMs

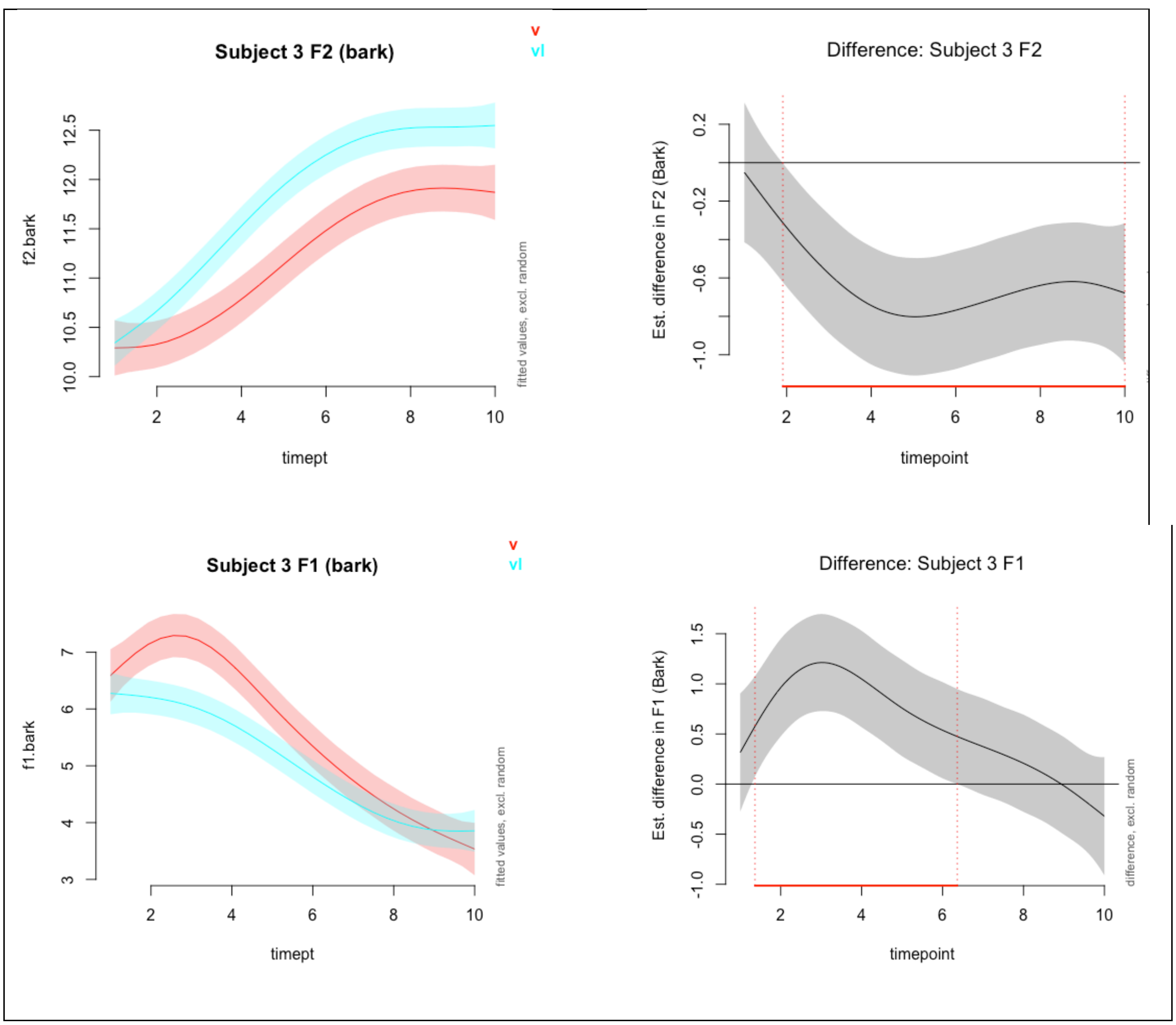
Monosyllabic words

In the previous section, the statistical testing of trajectory length measurements revealed that in the group as a whole, there was a significant difference between pre-voiced /aɪ/ and pre-voiceless /aɪ/. In this section, for each speaker and for each type of word, Bark formant trajectories were predicted with a main predictor of following consonant voicing, smoothing predictors for each timepoint by the voicing (to account for predictable variation in each timepoint that can be accounted for by voicing), with random smooths for each word (to account for any difference that can be attributed to within-word variation). These models were run for both monosyllabic and disyllabic, pre-flap /aɪ/ with the same model terms, and each individual speaker was considered in their own model so that the speaker can be classified in terms of their raising pattern.

The visual results of these GAMMs in one set are included Appendix 1, along with the difference plots that highlight which sections of the vowel have significant differences between pre-voiceless /aɪ/ and pre-voiced /aɪ/. There are separate models, and therefore separate model plots and difference plots, for F1 and F2. Significance using GAMMs is determined by a difference curve plot, which shows the predicted difference between the two predicted outcomes for F1, in this case pre-voiced /aɪ/ F1 and pre-voiceless /aɪ/ F1, with error bars around those predictions. In these difference curve plots, if the predicted difference is significantly different from zero, including the possible error bars, the formant tracks are determined to be significant at that point along the formant track. On the difference plots below, places where there is a significant difference between pre-voiceless /aɪ/ and pre-voiced /aɪ/ are highlighted in red along the x axis, which depicts time points.

First, models were run using only monosyllabic /aɪ/ words, which is typically the first phonological environment where raising occurs in Fort Wayne speakers (Berkson et al., 2017). Then, models were run using disyllabic, pre-flap /aɪ/ in both underlyingly voiced and underlyingly voiceless context, to determine whether any participants in this study were producing a difference in that phonological environment, which typically shows raising at a later stage, and always includes monosyllabic raising as well. Put another way, no speaker produces disyllabic pre-flap raising without also producing monosyllabic raising, and this is one reason the disyllabic raisers are considered the most advanced; the other reason is that raising preceding underlyingly voiceless flaps that are voiced in their surface-level pronunciation is a phonological example of opacity, and is considered more advanced than a purely phonetic pattern. The majority of speakers in this study produced a raised and fronted /aɪ/ in monosyllabic words preceding voiceless consonants, but did not produce a difference between the two phonological contexts in the pre-flap environment. These classifications are based on the GAMM results and difference plots of the /aɪ/ vowels from monosyllabic words first, and the same model results and difference plots from disyllabic pre-flap /aɪ/ words second.

Fifteen of the twenty-one speakers raised (lower F1) and fronted (higher F2) in /aɪ/ preceding voiceless consonants compared to /aɪ/ preceding voiced consonants in monosyllabic words. This raising and fronting was significant across at least two timepoints in each formant, although not necessarily the same two timepoints in every case, and in some cases was a significantly different full separation of the formant between pre-voiced and pre-voiceless context. A few examples of what this looks like in the GAMM output and difference plot are included below in Figure 7.



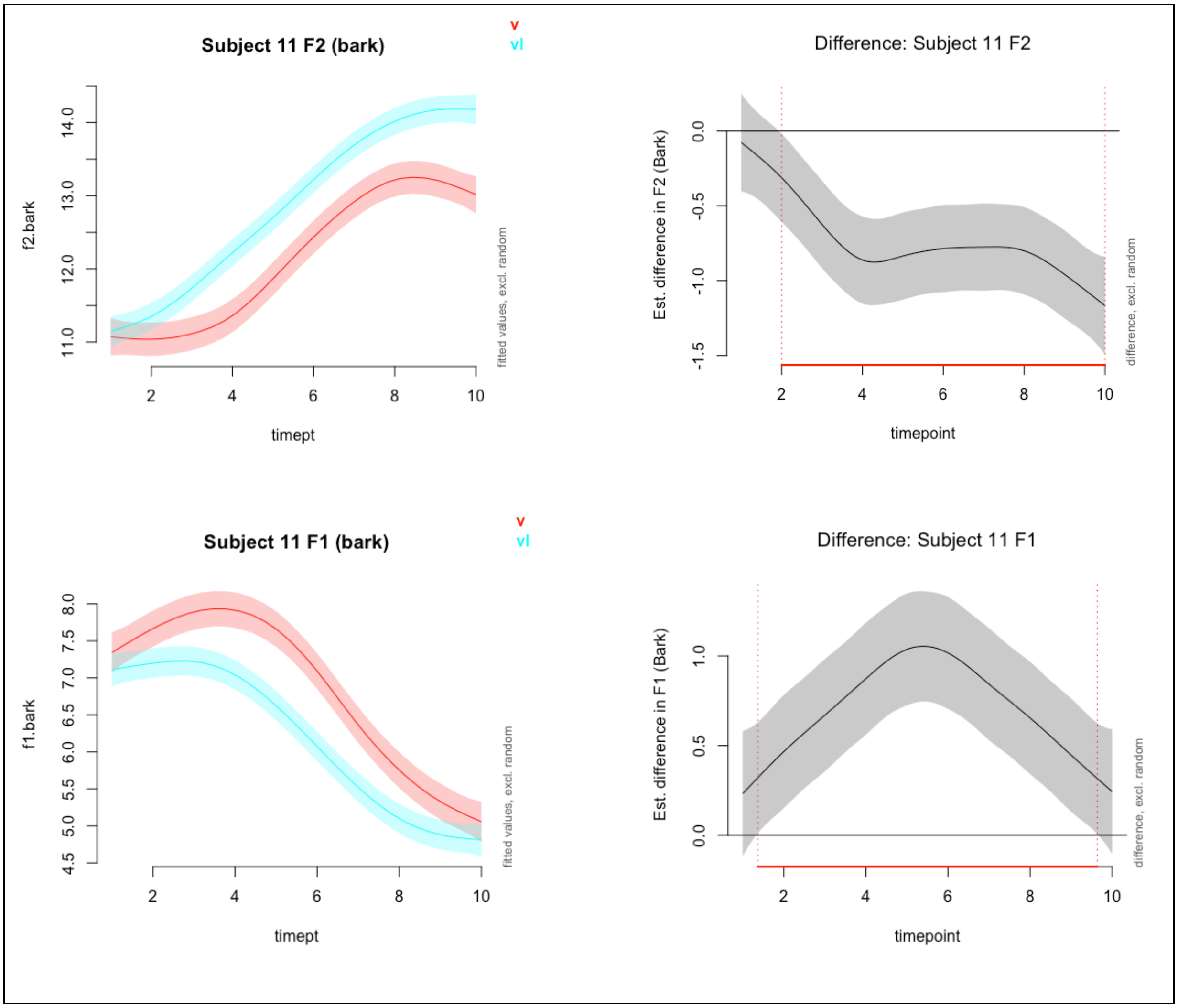


Figure 7 Raiser + fronters, Subjects 3 and 11 GAMM and difference plots

For Subject 3, F2 is significantly higher, or the vowel is fronted, along almost the entire duration of the vowel. F1 is significantly lower, indicating a raised vowel, in the first half of the diphthong but not in the second half. For Subject 11, both F2 and F1 are significantly different when comparing the pre-voiced and pre-voiceless formants, across most if not the entire formant track. F2 is significantly higher, or the vowel is fronter, and F1 is significantly lower, indicating a raised vowel, in the pre-voiceless context for this speaker. Both of these subjects are included in the group of fifteen raisers + fronters, characterized by both significant raising and fronting in the same individual, but they are examples of the variability in types of production patterns that were found in this category. It's difficult to find further patterns in these data, however, because speakers within the group don't seem to do the same thing. A few speakers have separation all along both formant tracks. Others, like Subject 3, have nearly full separation along either F1 or F2. There is also variation in where the separation or partial separation occurs: in the nucleus or the coda or sometimes right in the middle of the vowel, encompassing a bit of the nucleus and coda but not clearly one or the other. These differences in realization of raising and fronting among participants will be further discussed in the next section.

Of the six remaining speakers, three only produce significant differences in F2 as shown in Figure 8, two only produce significant differences in F1 as shown in Figure 9, and one produces no significant differences between pre-voiced and pre-voiceless /aɪ/ in either F1 or F2, as shown in Figure 10. So, apart from the raisers + fronters discussed above, there are fronters, raisers, and one participant who produces no significant difference.

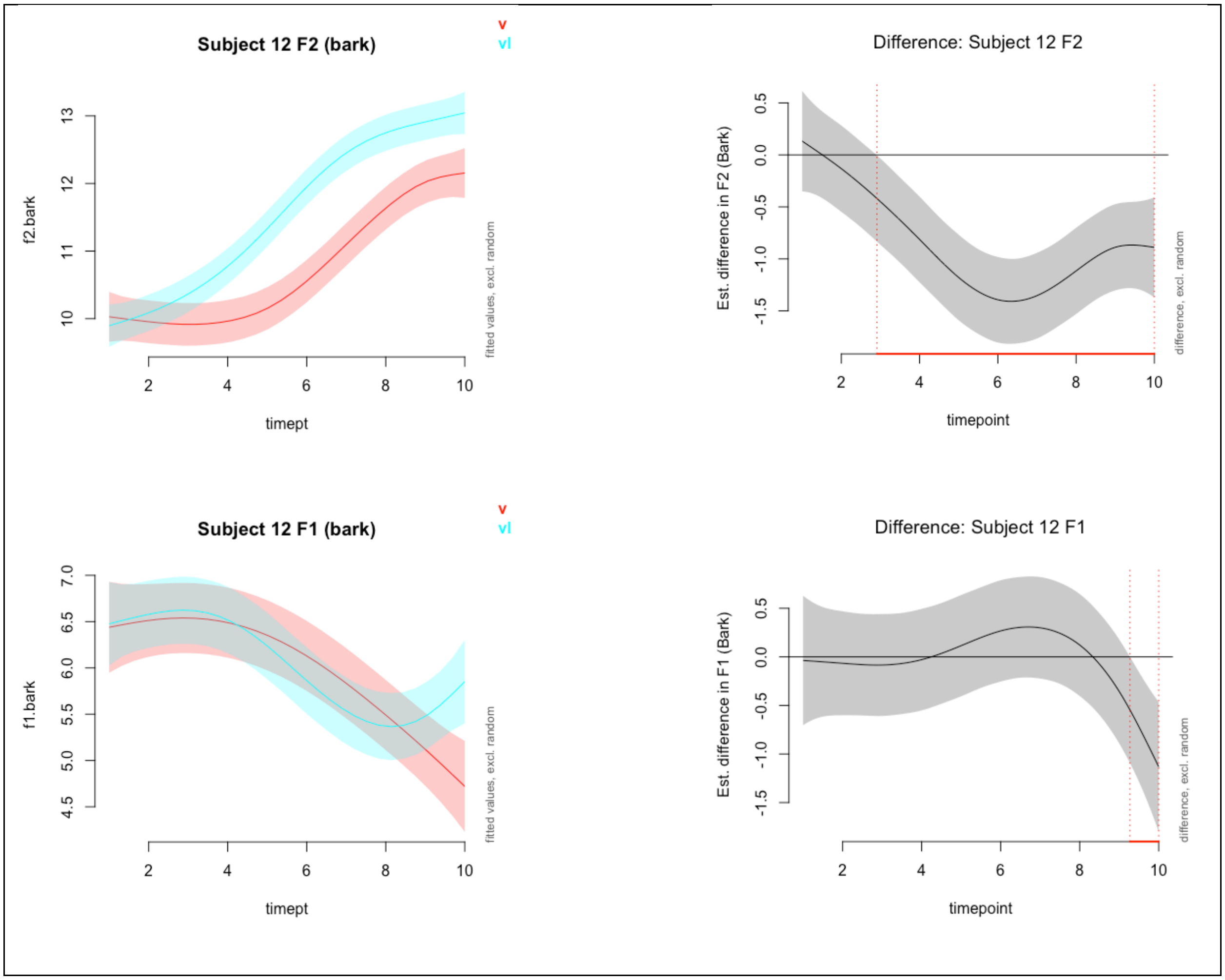


Figure 8 Fronter; Subject 12 GAMM and difference plot

Above, Figure 8 shows an example of a subject who produces a raised F2, or a fronter vowel with no raising (actually a very small amount of raising, but just right at the end) in F1.

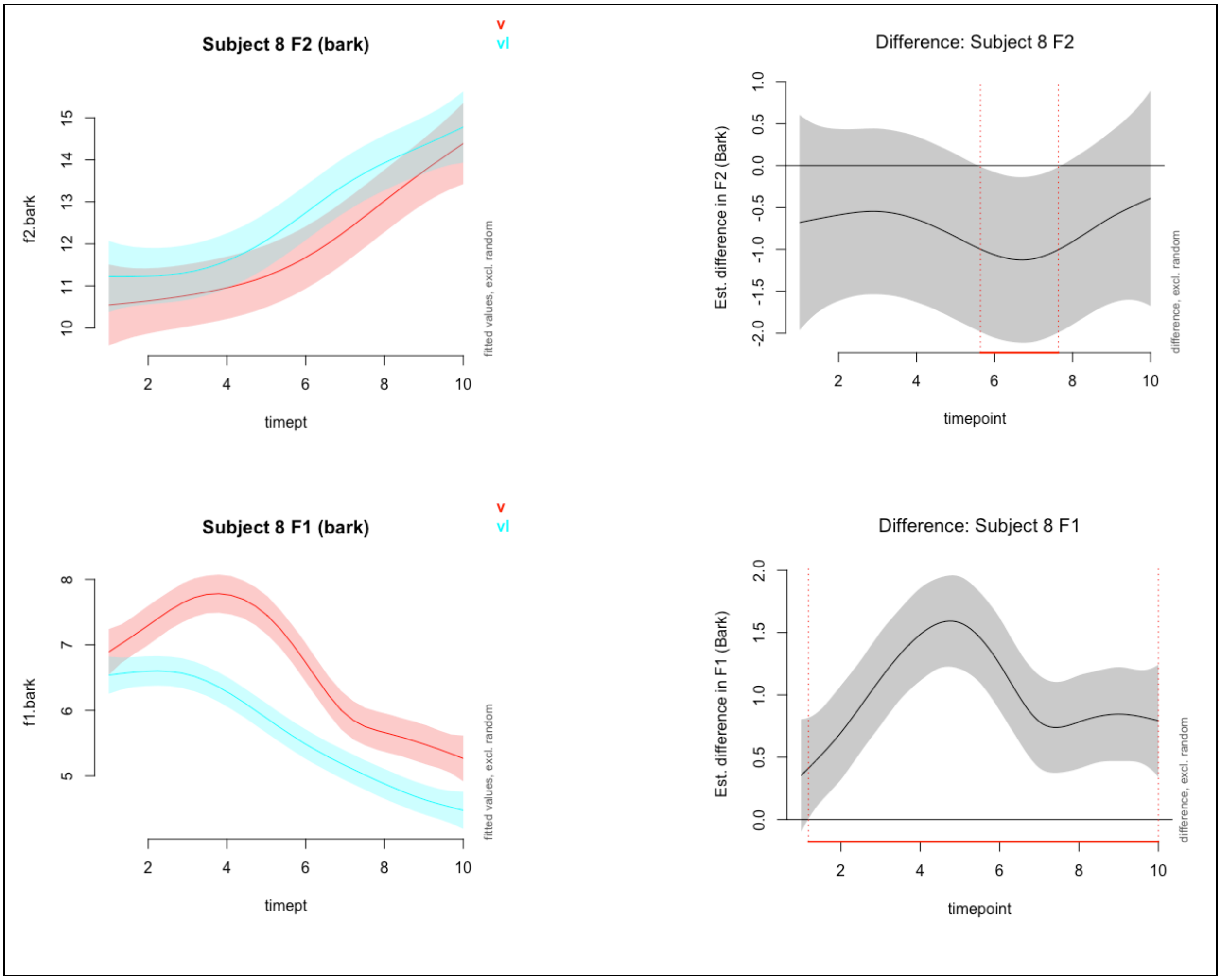


Figure 9 Raiser; Subject 8 GAMM and difference plot

Figure 9 shows one of the two subjects who produces a raised vowel (lower F1), but not a fronted vowel, in pre-voiceless /aɪ/. There is a tiny bit of fronting from timepoints 6-7 in F2, but the difference doesn't span two time points and the error bar in the difference plot is very close to 0, which would indicate a non-significant difference. The error bars in the GAMM plot on the left are also overlapping, and the model does not produce a significant prediction term in the output. So, while this subject may be on their way to producing a fronted /aɪ/ in addition to a raised /aɪ/, these data are not significantly fronted.

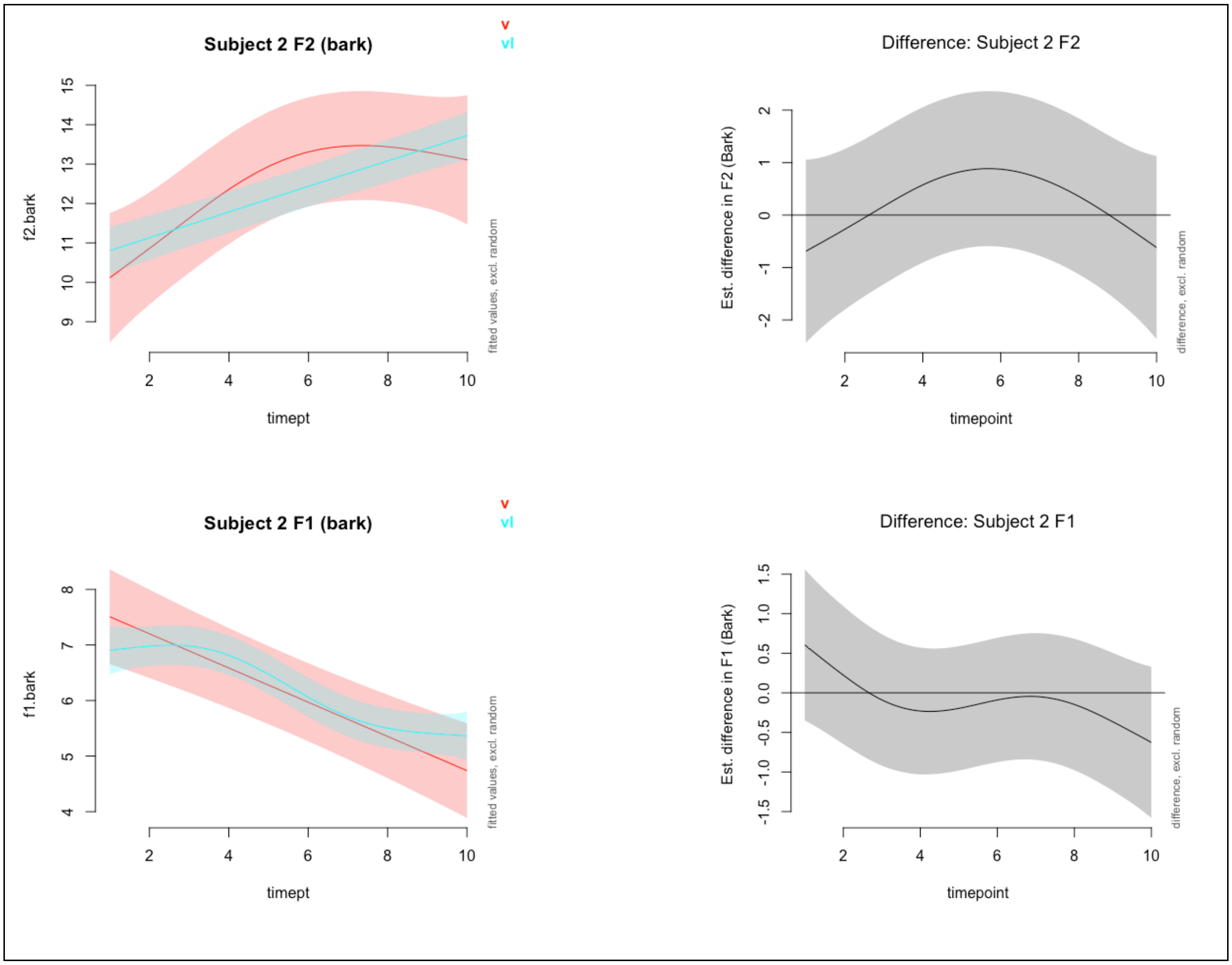


Figure 10 Non-raiser and non-fronter

Finally, shown in Figure 10, this speaker does not produce significant separation between pre-voiceless and pre-voiced /aɪ/ in either F1 or F2.

This classification system results in the following division of the twenty-one speakers in this chapter, in terms of comparing pre-voiceless /aɪ/ to pre-voiced /aɪ/ in monosyllabic words. The division of subjects is shown below in Table 1.

Raisers + fronters	15 participants	Subjects 3, 5, 6, 7, 9, 10, 11, 13, 15, 16, 17, 18, 19, 20, and 21
Fronters	3 participants	Subjects 1, 4, and 12
Raisers	2 participants	Subjects 8 and 14
No change	1 participant	Subject 2

Table 1 List of participants by change in pre-voiceless /aɪ/

Disyllabic words

Categorization of American raising typically includes analysis of /aɪ/ in disyllabic, pre-flap position as well as in monosyllabic words. In Berkson, Davis, and Strickler (2017), participants who raised in underlyingly-voiceless pre-flap position (i.e., before *writing*) were categorized as the most advanced participants in the sound change.

Out of the twenty-one total subjects in this chapter, six produced fewer than all eight disyllabic words including in the stimuli list due to either recording error or word mispronunciation and are excluded from this analysis due to lack of data. Of the fifteen remaining speakers, eleven produce no difference between pre-/t/-flap and pre-/d/-flap /aɪ/ in the four words in each category.

Four of the 15 raisers + fronters from the monosyllabic section above also produce nucleus raising (vowel raising, which is F1 lowering), with no significant difference in F2 in the flap context. They are Subjects 8, 10, 13, and 19.

Origin of the separation of pre-voiceless and pre-voiced formant tracks

Eight speakers exhibit some form of offglide raising or fronting in pre-voiceless /aɪ/, determined by a difference in the second half of the diphthong, or timepoints 5-10. For these speakers, the significant differences between pre-voiced /aɪ/ and pre-voiceless /aɪ/ are in either F1 (raising, as in Subjects 8, 14, and 17) or F2 (fronting, as in Subjects 1 and 12), or with significant differences in both formants (raising + fronting, as in Subjects 11, 16, and 18). These differences appear to be most prevalent in the offglide portion of the diphthong, similar to Thomas and Mielke's (2022) Ohio speakers, who they described as "unzipping" pre-voiceless /aɪ/ from pre-voiced /aɪ/, with differences at the end of the vowel in the earliest stages of the sound change to differences at the beginning of the vowel in the later stages. An example of a speaker who "unzips" the formants back-to-front is shown in Figure 11, in Subject 16.

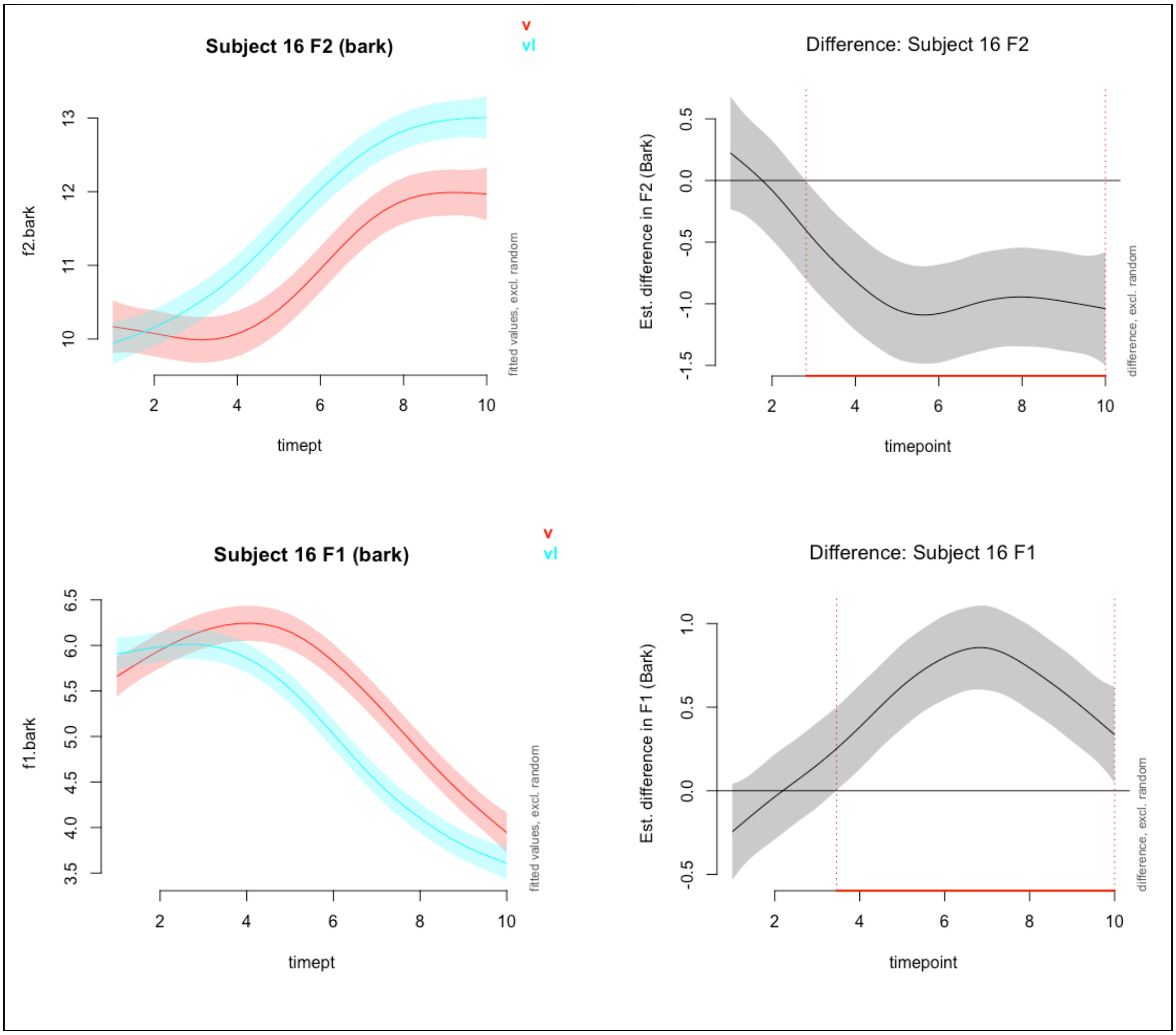


Figure 11 Example of speaker who separates formants starting from the end of the vowel

As a contrast to the above speakers with a split seeming to originate in the offglide, six speakers exhibit significant differences in pre-voiceless /aɪ/ compared to pre-voiced /aɪ/ originating in the nucleus. These don't appear to show the same "unzipping" from the end of the diphthong process, but rather a split that occurs in the nucleus of the diphthong and sometimes spreads to the glide. An example of a speaker who separates the pre-voiced and pre-voiceless formants from the middle is shown below in Figure 12, Subject 19.

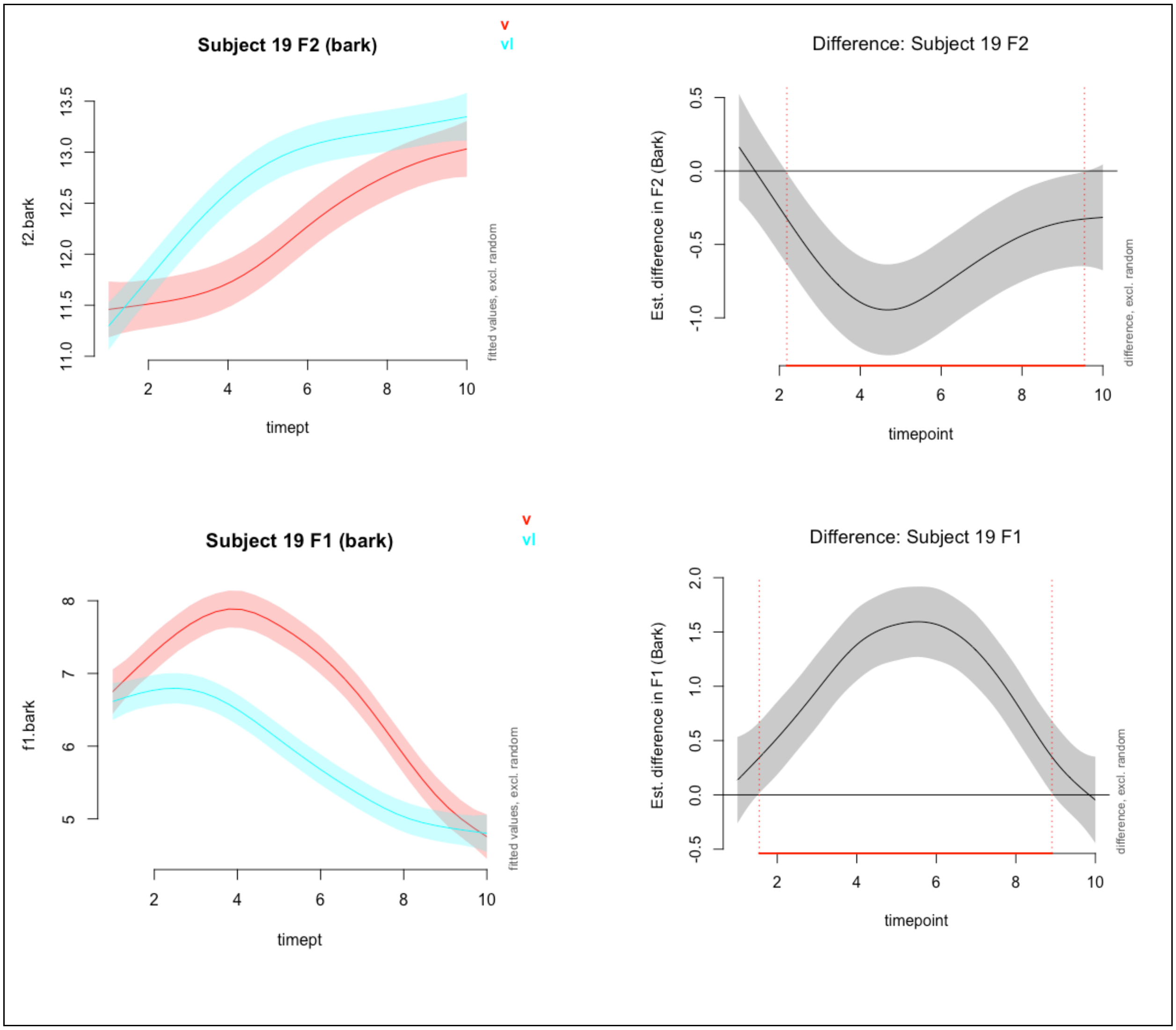


Figure 12 Example of a speaker with nucleus separation, extending towards the glide

Five participants demonstrate a combination of both origin points, and appear to separate F1 at the nucleus and F2 at the offglide. All five of the participants in this category followed this same pattern, which is an interesting combination of nucleus and offglide sound change—fronting /i/ in the diphthong while raising /ɑ/, and sometimes raising as well as fronting /i/. An example of this combined pattern is shown in Figure 13, Subject 3.

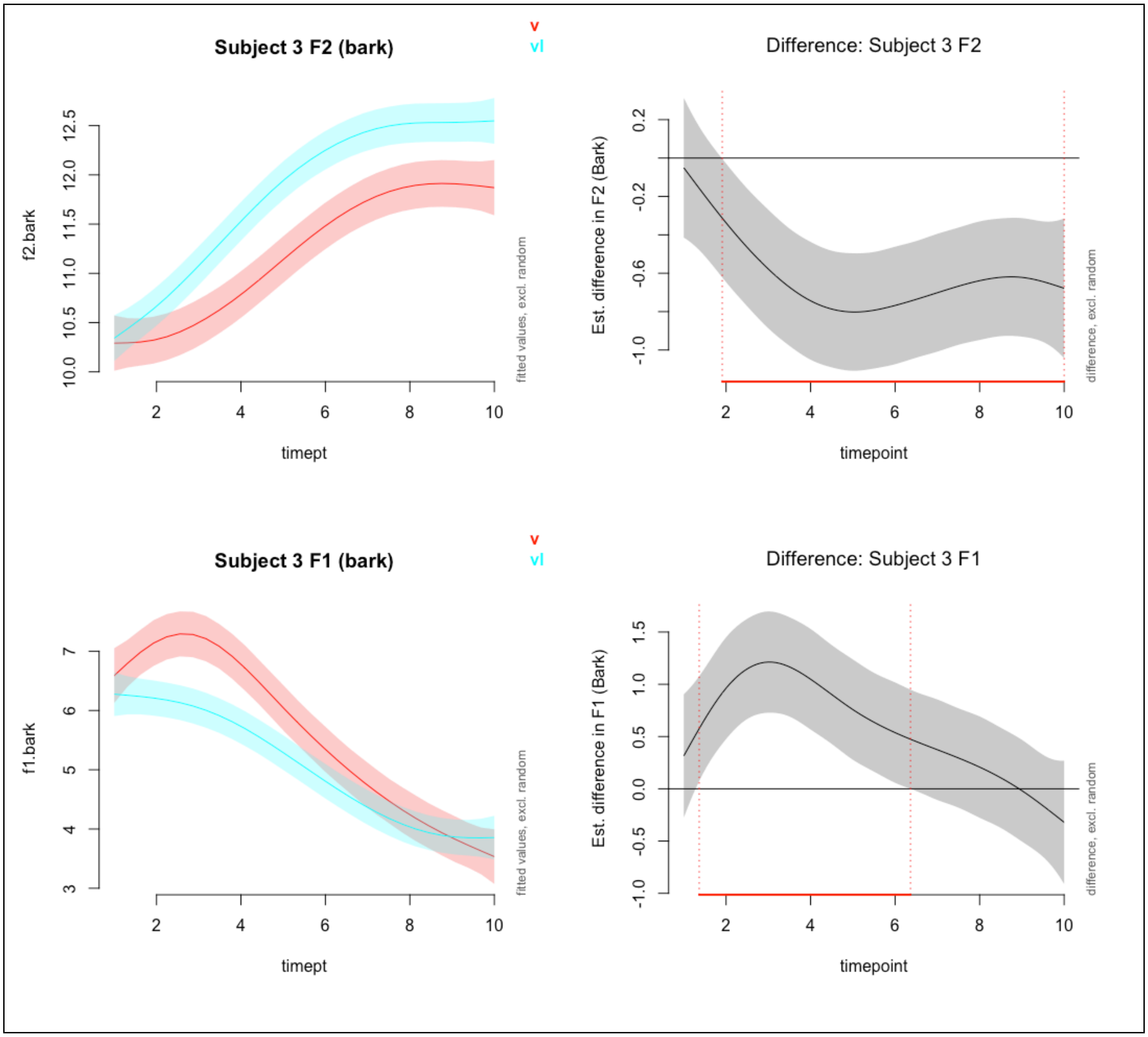


Figure 13 Example of a speaker with combination of nucleus raising and offglide fronting

One speaker produces a full separation of both formant tracks, so it is unclear whether the origin for the split is in the nucleus or the coda. This speaker is one of the most advanced in the sound change, and was discussed in the previous section as a pre-flap raiser. This speaker is pictured in Figure 14.

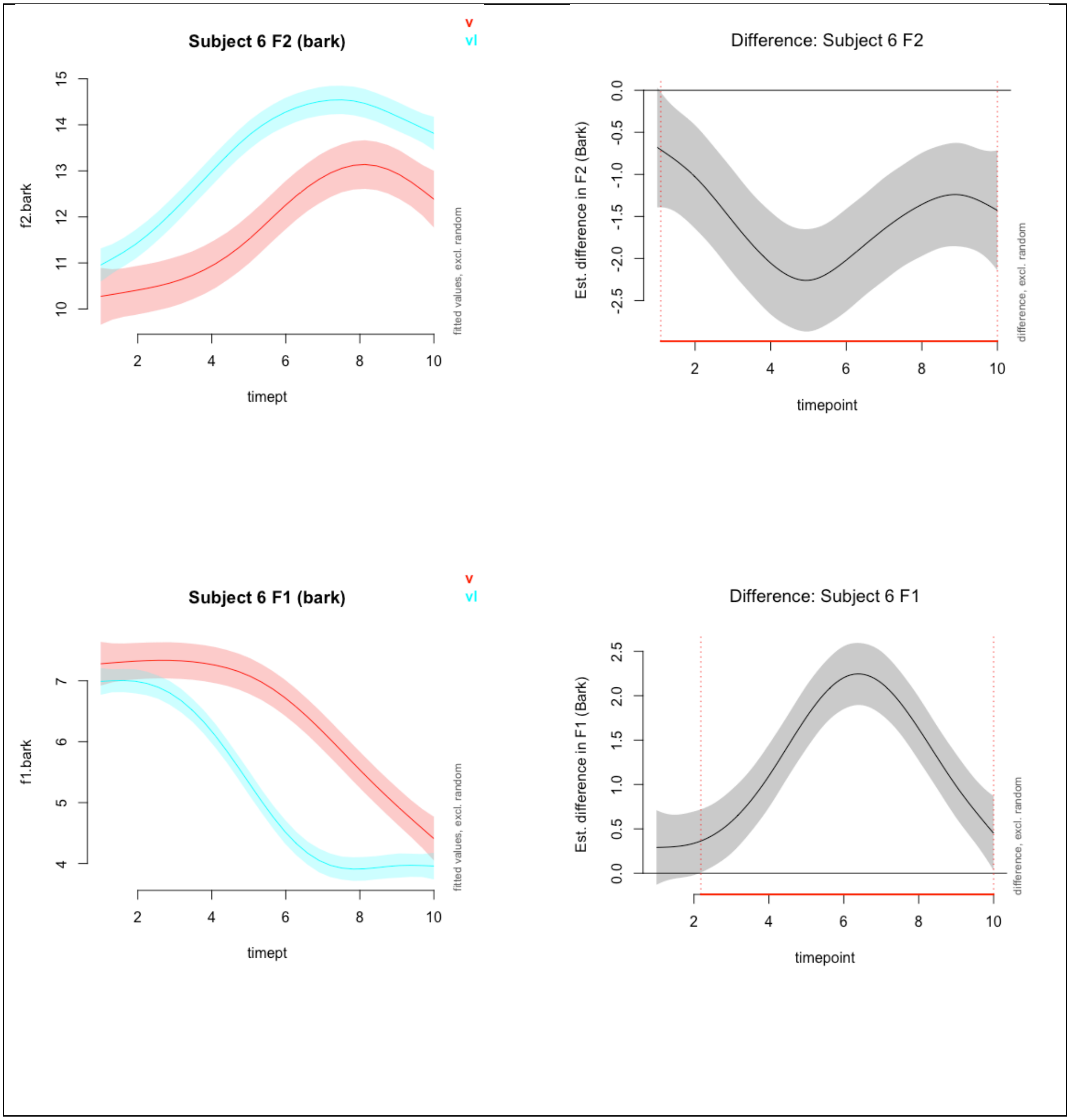


Figure 14 Example of a speaker with full separation in F1 and F2

The one remaining speaker, Subject 2 as mentioned in the previous section, does not produce a difference between pre-voiceless and pre-voiced /aɪ/ in monosyllabic words. However, among the other 20 speakers, there seems to be compelling evidence of both raising and/or fronting originating in the offglide and moving backwards through the vowel (e.g., Thomas & Mielke 2021) and raising originating in the nucleus (e.g., Berkson, Davis, & Strickler 2017; Labov, Ash, & Boburg 2006).

Discussion

In this chapter, I considered various ways to measure /aɪ/ raising, or American raising, in speakers from Fort Wayne, Indiana. Previous research has mostly relied on a single difference measure at the midpoint of the nucleus of the diphthong, and used a threshold of 60Hz difference between pre-voiceless and pre-voiced F1 at this point to determine raising (e.g., Berkson, Davis, and Strickler 2017; Labov, Ash and Boburg 2006). I demonstrated that in the speakers in this study, F2 also increases (indicating vowel fronting) for most speakers, along with F1 lowering (producing a raised vowel) in pre-voiceless position. Hualde, Barlaz, & Luchkina (2022) and Moreton (2004) also described vowel fronting realized in F2 as a feature in the process that is typically characterized only by the raising realized in F1. I wanted a method to analyze the data quantitatively, to include F2 as well as F1 in the analysis, and to consider more of vowel trajectory when considering the sound change, so I turned to the trajectory length analysis.

I used both F1 and F2 to calculate the trajectory length of the nucleus and coda of each word, and in a model that included all of the speakers in the study, the trajectory length of the nucleus was significantly predicted using the voicing of the following consonant. This method demonstrated that there are group-level difference between the nucleus of pre-voiceless [ʌɪ] and the nucleus of pre-voiced [aɪ]. While this method achieved the goal of including the entire vowel trajectory and F2 in the analysis of /aɪ/, it didn't capture the specific time in the vowel that differences in pre-voiced and pre-voiceless /aɪ/ occurred, and so I had to consider more sophisticated statistical methods.

Finally, I used Generalized Additive Mixed Models to analyze each speaker in-depth, and to be able to categorize the production patterns of individuals. An initial categorization using monosyllabic words determined that 15 of the 22 participants both raised and fronted /aɪ/ in pre-voiceless position, 3 participants fronted /aɪ/, 2 participants raised /aɪ/, and 1 did not produce any differences between the two voicing contexts. A second analysis of bisyllabic pre-flap /aɪ/ determined that 4 out of 15 participants raised /aɪ/ in the nucleus of the vowel, with no differences in F2 or in the offglide, and the other 11 participants produced no difference between the two underlying flap voicing contexts.

The GAMM analysis also provided support for two theories about the origins of the American raising sound change. One theory is that the sound change begins in the offglide, possibly in F2 as a fronting process (e.g., Moreton 2004; Thomas & Mielke 2021). Another implicit theory of origin is that American raising only affects F1 in the nucleus of /aɪ/; I describe this as an implicit theory because to my knowledge, it's never been directly stated as a claim of origin, just that the analyses using the midpoint of the nucleus only consider the nucleus in F1 in their analysis, not F2 in the offglide (e.g., Berkson, Davis & Strickler 2017). I find evidence of both of these types of change in my data. 8 speakers appear to exhibit offglide fronting and/or raising first, with the split between the pre-voiced and pre-voiceless formants in /aɪ/ moving backwards towards the nucleus. However, six speakers seem to show a difference between pre-voiced and pre-voiceless /aɪ/ originating in the nucleus, rather than in the offglide. And five speakers demonstrate a combination of these two patterns, with nucleus raising in F1 and offglide fronting in F2. These detailed analyses are possible due to the GAMM analysis, which allows for a detailed analysis over the course of the vowel.

These data show that “raising” is occurring in these speakers’ production of /aɪ/ in monosyllabic words, and that the sound change is not only raising in F1, but also fronting in F2, for the majority of speakers.

Chapter 3

Clear speech

Introduction

This chapter considers how Fort Wayne speakers produce /aɪ/ in clear speech. The goal of this research is to see how speakers change their production of /aɪ/ in clear speech compared to citation speech, in order to approximate speakers' mental representation of /aɪ/ by seeing what target (raised or unraised) speakers move toward. Just as speakers produce variations of /aɪ/ in different contexts depending on what stage of the sound change they're experiencing, I expected that there would be variation in how speakers respond to the task of producing /aɪ/ in clear speech. My predictions for this experiment were that if participants appeared to intensify the raising process, or create more difference between the pre-voiced /aɪ/ and pre-voiceless /aɪ/ productions, it would indicate that they are targeting a raised variant of a word during a clear speech task, and therefore indicate that a raised and/or fronted /aɪ/ preceding a voiceless consonant is a part of their representation of /aɪ/. In other words, it would indicate that for that speaker, there are actually two separate phonemes—/aɪ/ and /ʌɪ/—which would also indicate phonologization of the sound change for that speaker.

Another possible scenario is that the sound change is a phonetically motivated centralization process triggered by shortened vowel lengths, that may eventually become a phonologized change but hasn't yet. In that case, the flip side of my predictions were that if participants reduced raising in a clear speech task compared to their relaxed speech, it would seem that it's a phonetically driven process, or that the raised variant of the vowel is not the target of the speakers, and that there is only one phoneme represented for those speakers, /aɪ/. These results were predicted to, and do, vary by speaker; but for the most part, the speakers in this chapter do not produce pre-voiceless /aɪ/ in clear speech with an obviously raised and fronted target.

Background

Previous research has used clear speech as a method of analyzing the representation of a sound. Asking participants to speak clearly typically pushes speakers to produce a more hyperarticulated vowel. Hyperarticulated vowels have been suggested to be beneficial to perception, and also to most closely approximate phonemic representation; when speakers were asked to choose a best example of a vowel, they chose vowels so hyperarticulated that they were outside the range of human production (Johnson, Flemming, & Wright 1993). Whalen et al. (2004) similarly determined that participants preferred hyperarticulated vowels, and that vowel targets can be determined by speakers' productions. Inspired by these studies, the assumption for the clear speech produced in this experiment will be that whatever variation is produced in clear speech is the hyperarticulated target for that sound, and is a closer approximate to the representation of the sound than the non-clear speech.

Hyperarticulated speech as realized in high neighborhood density words as opposed to low neighborhood words has also been used to study speakers' varying representations of a vowel undergoing a sound change (Zellou and Scarborough, 2019). /æ/ typically raises in pre-nasal context, compared to pre-oral context; Zellou and Scarborough find that pre-oral /æ/ is hyperarticulated in clear speech by lowering in high neighborhood density words, compared to pre-oral /æ/ in low neighborhood density words, which have fewer competitors to distinguish against. Similarly, pre-nasal /æ/ is hyperarticulated in high neighborhood density (ND) words with more nasalization and increased diphthongization, furthering the distinctions between pre-nasal and pre-oral /æ/. The results show that speakers enhance the difference between pre-oral and pre-nasal /æ/ in hyperarticulated high ND words compared to less hyperarticulated low ND words, which signals that the speakers increased the split when they were trying to be better understood. I predict that my Fort Wayne speakers will do the same if the change is phonologized, i.e., will raise /aɪ/ in hyperarticulated contexts rather than lower. This is likely if speakers feel that enhancing the split will increase their listeners' chances of understanding them, like they do in the Zellou and Scarborough study.

Clopper, Mitsch, and Tamati (2017) also studied dialectal variation in speaking styles, describing findings in previous literature that dialectal features are emphasized in contexts normally associated with reduction, including in high frequency, low density, and high predictability targets. In their own study, the authors considered two regional US English dialects, Northern and Midland dialects, in the way that they differed in contexts of lexical competition, second mention, and careful speaking style. They find that there are reductions in the form of less peripheral productions in casual speech in the vowels that they were targeting as regional dialect features. Clopper, Mitsch, and Tamati (2019) describe /æ/-lowering in clear speech for Northern dialect speakers, for whom /æ/-raising is a salient dialect feature. They found that participants *reduced* dialectal features in instances of more careful speech, which is the opposite of Zellou and Scarborough's (2019) results, in which speakers *increased* features that separated allophones from one another in clear speech. These competing results could reflect a difference between the salience or dialectal markedness of /æ/-raising in each dialect: for Northern dialect speakers in Clopper et al.'s study, /æ/-raising is a marked feature of their regional dialect; for Colorado speakers in Zellou and Scarborough's study, /æ/-raising pre-nasally is a phonetic feature without any particular markedness. It could be that speakers reduce these marked dialectal features in clear speech, but increase unmarked phonetic features in clear speech, both processes with the end goal of increasing perceptibility of speech.

As these two sets of results could apply to /aɪ/ raising in Fort Wayne, it is not obvious what the results of hyperarticulation will be in the /aɪ/ raising context. /aɪ/ raising doesn't seem to be part of any Atlas of North American English dialect region, like those examined in Clopper et al.'s study, and is instead more specific to a small area and seems to occur in such spread places that it's not a characteristic of one specific regional dialect. This makes /aɪ/ raising seem more similar to Zellou and Scarborough's study, and would lead to the prediction that Fort Wayne speakers will increase raising and/or fronting in clear speech relative to citation speech, because it is not a marked dialectal feature.

Methods

Ten participants completed both speech styles in the clear speech task, in which participants read a wordlist in both citation speech and in clear speech. All of these speakers were analyzed in Chapter 2; they are Subjects 8, 9, 10, 11, 12, 13, 14, 15, 16, and 19. In a counterbalanced design, seven of these participants produced clear speech first and citation speech second, and three of these participants produced citation speech first and clear speech second. The word list (in Appendix 4) includes monosyllabic /aɪ/ vowels in both pre-voiced (N=14) and pre-voiceless (N=16) contexts, with the added dimension of having both clear and citation speech to compare. There will be two parts to this clear speech analysis: first, there will be an analysis of monophthongs in the data set, which include /i/, /u/, and /ɑ/. This will give a basic indication of whether participants produced clear speech at all, as well as provide comparisons for the targets of the nucleus and glide components of the diphthong /aɪ/; second, there will be an analysis of /aɪ/ on its own, to see whether participants increase or decrease raising and/or fronting in /aɪ/ in the pre-voiceless context.

The outcomes of this experiment are predicted to reflect speakers' representation of /aɪ/: if participants increase raising in clear speech, it would indicate that the speakers' representation of a hyperarticulated /aɪ/ includes the raised variant; if participants decrease raising in clear speech, it would indicate that the /aɪ/ raising in pre-voiceless context is not included in their representation of /aɪ/. It is likely and even expected that there will be some individual variation in how participants respond to this task, possibly depending on how far along a speaker is in the sound change process.

Monophthongs

There are relatively clear predictions for monophthongs with regards to clear speech. In clear speech, monophthongs are predicted to peripheralize (Uchanski, 2005; Smiljanic and Bradlow, 2007; Ferguson and Kewley-Port, 2007, as cited in Zellou and Scarborough, 2019). An analysis of the monophthongs /i/, /u/, and /ɑ/ was used to determine whether the participants produced clear speech in the context of the online experiment. These vowels were analyzed using one midpoint measurement. In this study, three vowels were under consideration, /i/, /u/, and /ɑ/. In keeping with predicting peripheralizing of monophthongs in clear speech, /i/ is expected to be both higher and fronter in clear speech, /ɑ/ is expected to be lower and backer in clear speech, and there aren't strong predictions for /u/ in terms of frontness because it often undergoes /u/-fronting and speakers are inconsistent in whether they increase or decrease the /u/-fronting in clear speech, but we do expect it to be higher in clear speech (Clopper, Burdin, & Turnbull, 2019).

These three vowels were analyzed based on midpoints, measured at point five of ten. Separate models were run for each vowel and formant because the prediction of the direction of change in clear speech for each formant in each vowel was different; because the relationship between F1 and formant height is inversely related, vowel lowering is associated with a high F1 and vowel fronting is associated with a high F2.

All of the following formant models in this section predict effects by speech style with random intercepts by speaker. In a model predicting F1 in /i/, speech style predicted a decrease in F1 in clear, relative to casual speech style (estimate=-13.43, t=-2.309, p=0.02). In a model predicting F2 in /i/, speech style significantly predicted an increase in F2 in clear speech (estimate=209.13, t=3.84, $p<0.05$). A lower F1 indicates a higher vowel and a higher F2 indicates a fronter vowel, so these results support the hypothesis that speakers are producing a more peripheral /i/ in the clear condition of this experiment.

In a model predicting F1 in /ɑ/, speech style marginally predicted an increase in F1 in clear speech (estimate= 22.742, t=1.94, p=0.05). In a model predicting F2 in /ɑ/, speech style did not predict an increase in F2 in clear speech (estimate= 30.01, t=1.68, p=0.09). A higher F1 indicates a lower vowel so these results support the hypothesis that speakers are producing lower /ɑ/ in the clear condition of this experiment.

In a model predicting F1 in /u/, speech style did not predict a significant change in F1 in clear speech (estimate= 0.925, t=0.1, p=0.92). In a model predicting

F2 in /u/, speech style did not predict a significant change in F2 in clear speech (estimate= -48.32, t=-0.79, p=0.43). These results are not unexpected in /u/ in particular, as speakers typically aren't consistent in whether they produce a fronter or backer /u/ in clear speech, due to the /u/-fronting process (e.g., Clopper, Burdin, & Turnbull, 2019). This vowel was also only in four tokens for each speaker, so it has less data than the other two vowels and that would make it more difficult to find an effect if there was one.

The main takeaway from this analysis is that speakers seem to be producing clear speech as expected in /i/ and /ɑ/. So, noting that these trends of /i/ and /ɑ/ follow the expected direction of change in clear speech, we can move forward with the analysis assuming that speakers are producing clear speech.

Predictions about /aɪ/ in clear speech

/aɪ/ is an interesting vowel to consider in clear speech because there are not obvious predictions about how it will change. Without phonological /aɪ/ raising, it might be predicted that the diphthong would behave as both individual vowels do in clear speech, with /ɑ/ lowering and /i/ raising and fronting. However, given that the /aɪ/ is raised in pre-voiceless contexts, it is possible that the clear version would be extra raised in those contexts rather than peripheralized. Thus, the clear speech can be used as a window into the representation of the /aɪ/ vowel for speakers. If speakers increase raising in /aɪ/ in pre-voiceless contexts in clear speech (lower F1), it could indicate that the raising is part of the representation of the vowel for that speaker, such that increasing raising makes that vowel a “better” version of /aɪ/ preceding voiceless consonants, and that increasing raising produces a better, more hyperarticulated vowel. However, if raising decreases in clear speech, it would indicate that raising is not a part of the hyperarticulated representation of /aɪ/ for that speaker.

In this chapter, I will first explore group trends in production of /aɪ/ in clear speech, and then the main analysis will consist of modeling individual speakers using generalized additive mixed models, as in Chapter 2. Because I expect differences in the data by speaker, due to differences in representation of /aɪ/, the analysis needs to allow for individuals to have separate outcomes in the experiment.

Group description of /aɪ/ in clear speech

Figure 15 below shows the average formant tracks in Bark for both pre-voiced and pre-voiceless /aɪ/ in both clear and citation (casual) speech. In Figure 15, F2 is the top 4 lines and F1 is the bottom 4 lines. Clear speech is shown in orange, and citation speech is shown in blue. The lighter blue and lighter orange colors show the pre-voiceless /aɪ/ formant tracks, and the darker blue and darker orange colors show the pre-voiced /aɪ/ formant tracks. Formants are shown and analyzed in Bark to make comparisons between differences in F1 and F2 more equitable.

The most obvious differences in the graph are between pre-voiced and pre-voiceless /aɪ/ in both formants, with slightly more consistent differences in F2. These are the differences between the lighter shades and the darker shades in the graph below with all four lines. But the comparison that's really of interest is the comparison between pre-voiceless clear speech and citation speech and also (to a slightly lesser extent) the comparison between pre-voiced clear speech and citation speech.

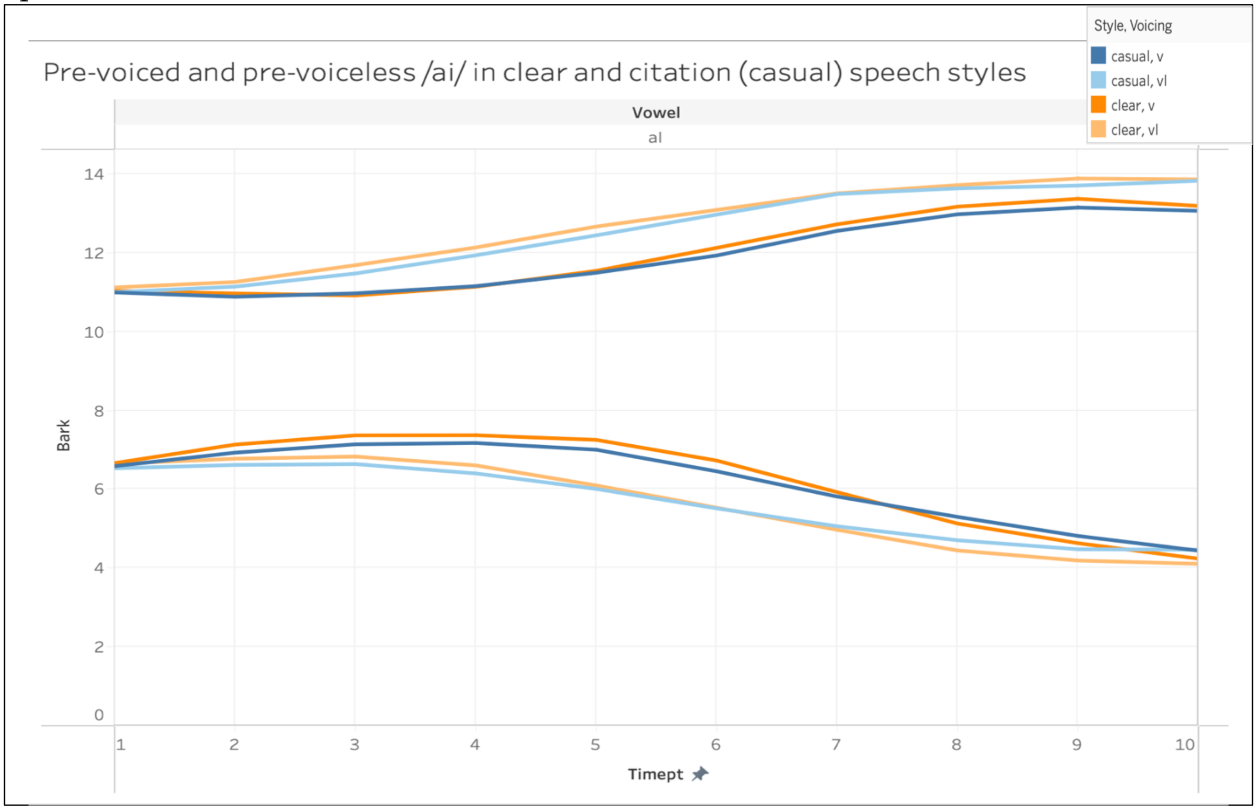


Figure 15 Pre-voiced and pre-voiceless /aɪ/ in clear and citation speech styles

Below, in Figure 16, is a simplified graph of **pre-voiceless** F1 and F2 in /aɪ/ in clear and citation speech. This is the most interesting case to think about clear speech in /aɪ/ because it is the environment with /aɪ/ raising, so it will be the environment in which there could be a difference between clear and citation speech. Clear speech has a higher F1, indicating a lower vowel, i.e., less raised, in the group of speakers overall in the first half of the diphthong; in the second half of the diphthong, clear speech has a lower F1, indicating a higher vowel. Vowel raising in the second half of the diphthong parallels the production of /i/ in clear speech, as described earlier in this chapter. In clear speech, participants are also producing a slightly higher F2 across the entire diphthong, although it is slightly more pronounced in the first half of the diphthong; this indicates vowel fronting across the diphthong as a whole.

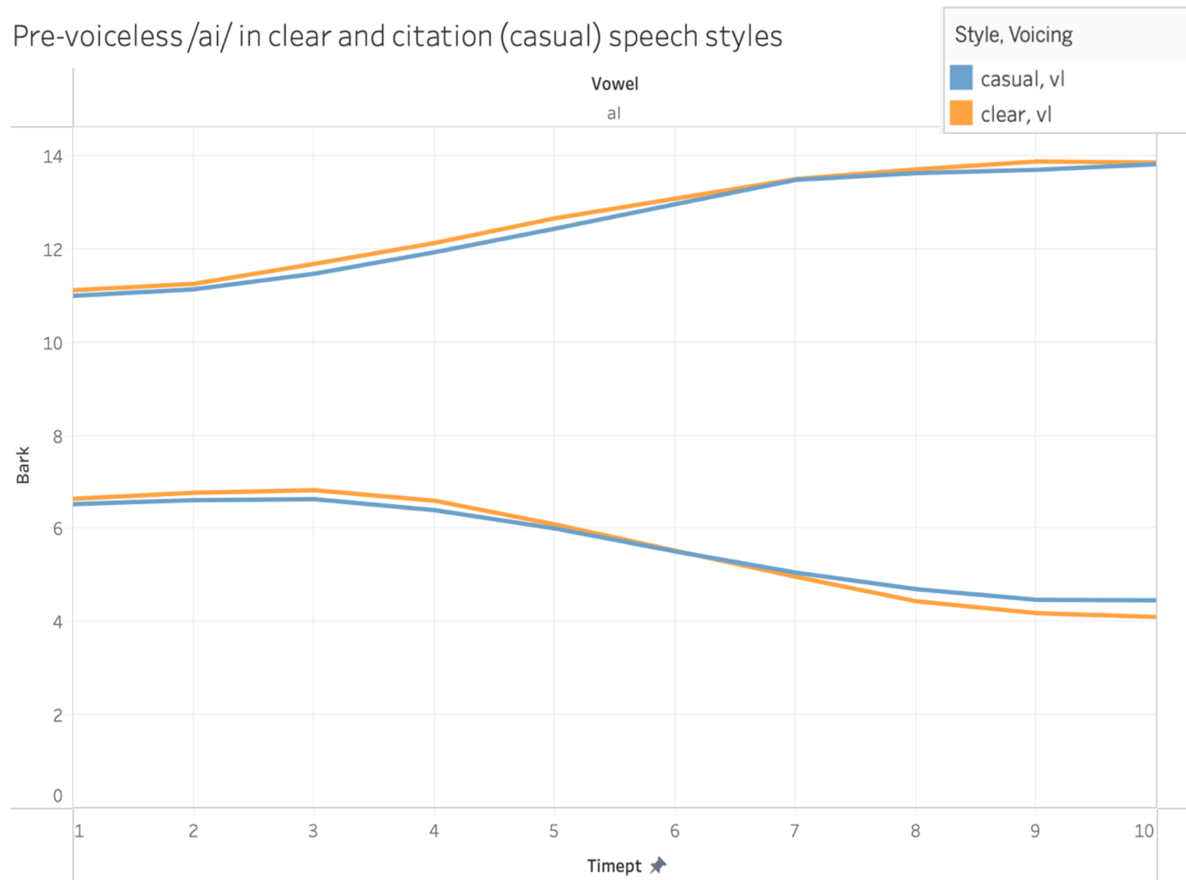


Figure 16 Pre-voiceless /aɪ/ in clear and citation speech styles

Pictured in Figure 17 is a graph of only **pre-voiced** /aɪ/ in both clear and citation speech styles. In pre-voiced /aɪ/ produced in clear speech, F1 is higher in the first half of the diphthong and lower in the second half of the diphthong, again indicating a lower vowel in the first half of the diphthong and a higher vowel in the second half of the diphthong. Although these differences appear to be less than the differences between F1 in clear and citation speech in pre-voiceless context. F2 seems to show no difference in the first half of the vowel, and is higher, indicating a fronter vowel, in the second half of the vowel. Speakers are doing mostly the same thing with /aɪ/ in both clear and citation speech, and the diphthong seems to roughly follow the patterns of /ɑ/ and /i/ in the previous section. The first half of the diphthong (the /ɑ/ part) is backer in clear speech than in citation speech. In the second half of the diphthong (the /i/ part), the vowel is both higher and fronter in clear speech. These differences all pattern with the clear speech of the monophthongs and show peripheralization. In pre-voiced vowels, the first half of the diphthong, /ɑ/ does not change in terms of frontness and backness in clear speech. This pattern also fits that of /ɑ/ in the analysis of monophthongs above, where F2 did not change significantly in clear speech.

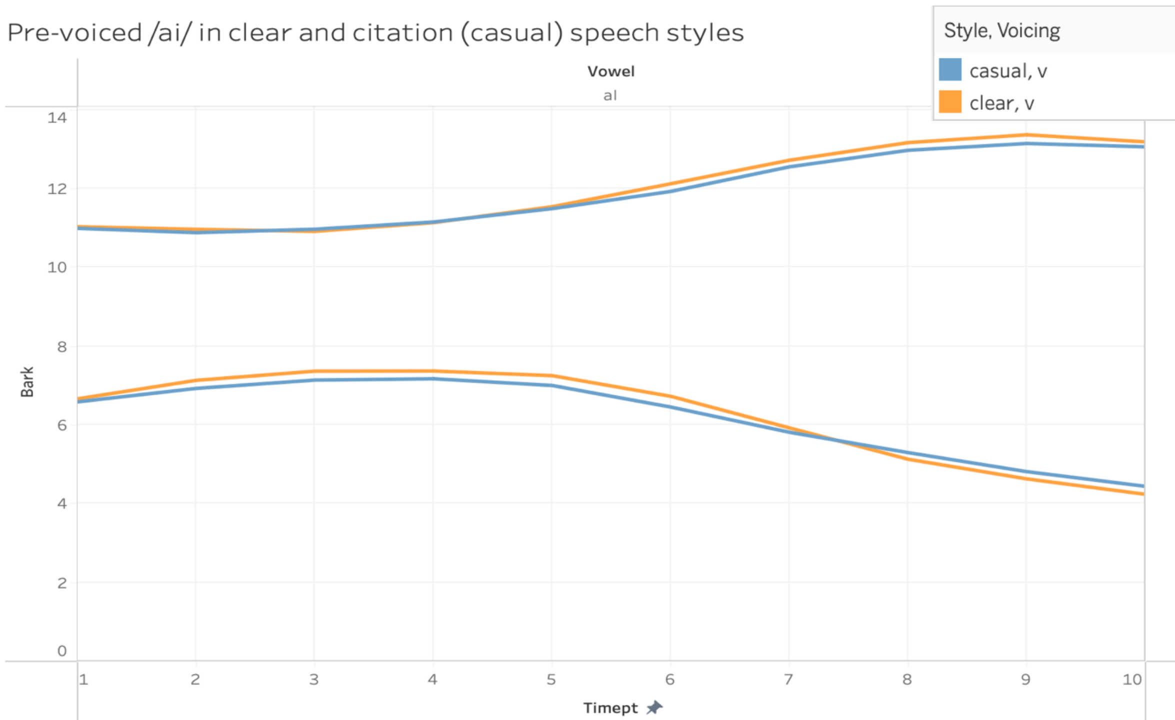


Figure 17 Pre-voiced /aɪ/ in clear and citation speech styles

Considering the group of speakers as a whole, then, /aɪ/ appears to be produced as its two component parts in clear speech, with peripheralization that is produced as /ɑ/ in the first half of the diphthong (lowering) and /i/ in the second half of the diphthong (raising and fronting).

Individual analysis of /aɪ/

Similar to the previous chapter, the interesting piece of analysis will be the individual differences that speakers produce, because speakers are experiencing and producing different parts of the sound change. As in the previous section, GAMMs were run using each individual speaker's data, to statistically identify individual differences in pronunciation between clear and citation speech. For each speaker, each formant (F1 Bark and F2 Bark) were predicted with an interaction of following consonant voicing and style with random smooths for timepoint by voicing and word. I chose the interaction term because in the predictions section, I thought that for at least some speakers, we might expect them to do different things for pre-voiced /aɪ/ than for pre-voiceless /aɪ/ in clear speech.

The plots of these GAMMs and their difference plots are included in Appendix 3 and they depict the predicted formant values with error bars, in both clear and citation speech; significance of these speech style differences are determined by the difference plots also included in Appendix 3, where points of significant difference between the two predictions are shaded red on the x axis. All of the significant differences were significant across the entire vowel, because the speakers varied their formant production in clear speech compared to citation speech across the entire vowel, not for a portion of the vowel like as in the previous chapter with regards to the raising and fronting results. Speakers seem to follow four different patterns, either no difference in either F1 or F2 (N=2), a difference in F1 (N=2), a difference in F2 (N=4), or a difference in both (N=2).

These results indicate that although individuals are doing slightly different things when producing /aɪ/ clearly, half of the speakers are fronting (raising F2) in both the nucleus and the glide. Five of the ten speakers produced significantly fronter /aɪ/. Two of the speakers who produced a fronter nucleus in clear speech also produced a higher diphthong. One speaker manipulated only F1 in clear speech and produced higher vowels.

One speaker did the opposite of the above five speakers and significantly backed and lowered /aɪ/ in clear speech. One additional speaker manipulated only F1 and also produced a lowered /aɪ/ in clear speech, without the fronting in F2. This speaker produced seemingly no difference between clear and citation speech in F2, so there's no trending direction to talk about. These two speakers seem to be undoing the raising and fronting processes in clear speech. This could indicate that these two speakers are moving away from the raising and/or fronting sound change in clear speech, rather than just peripheralizing each half of the diphthong as though it were /ɑ/ and /i/. It's difficult to draw strong conclusions from these two speakers though, so while I will put forth the possible analysis that they're producing a less extreme version of the sound change in clear speech compared to citation speech, I won't put much weight on it at this time.

Discussion

Overall, the ten speakers as a group are mostly treating both halves of the diphthong as its component parts in clear speech; a majority of the participants are producing lower vowels in the nucleus of the diphthong (as in /ɑ/) and higher and fronter vowels in the offglide (as in /i/). However, within this group trend, there are individual differences that indicate that speakers are doing slightly different things in terms of the sound change. Six out of ten speakers generally followed the above-mentioned trend of peripheralizing the diphthong as though it were /ɑ/ and /i/, in either F1, F2, or both formants. Two speakers didn't produce significant differences in either formant in clear speech compared to citation speech. However, the most interesting cases are the two speakers who didn't produce the expected trends for /ɑ/ and /i/ in /aɪ/, but rather moved their formants in the opposite directions from the other speakers; instead of raising and fronting, for one of those two speakers, the diphthong was lowered, and for the other speaker, the diphthong was both lowered and backed. This could possibly indicate that these speakers are undoing their raising and fronting processes that are part of the sound change, and seem to be behaving as Clopper et al.'s (2019) subjects do in clear speech, by undoing or lessening the dialectal features in clear speech compared to citation speech. However, with only two speakers, it's difficult to draw strong conclusions from these data.

Although speakers in the previous chapter produced fronting in addition to raising as part of the sound change, it's not really possible to tease out whether the fronting the speakers in this chapter are producing in clear speech is peripheralizing /i/ or the same type of fronting that's part of the sound change, in conjunction with raising. In this case, as the fronting is paired with lowering at the same time as raising in clear speech, it appears that the speakers are actually just enhancing the diphthong in the same ways that they enhanced /i/ and /ɑ/ in clear speech.

While it's interesting that those two speakers are trending *away* from increasing raising and fronting in clear speech, the main result of this experiment is that in clear speech, Fort Wayne speakers are simply treating /aɪ/ like /ɑ/ and are enhancing the diphthong by increasing the peripherality of the two component vowels, by lowering and fronting the entire vowel. There is no evidence that speakers are enhancing the pre-voiceless diphthong in clear speech by increasing the raising process, and therefore no evidence that the raised and fronted /aɪ/ that is part of the sound change is present in these speakers' representations of pre-voiceless /aɪ/.

Chapter 4

Perception

The ongoing change in the Fort Wayne community is an opportunity to study individual representation change as the sound change progresses, during this brief transitional period in which speakers are producing different patterns of the change. Since we know that speakers are producing variation in /aɪ/, that means that listeners must perceive language while hearing that same variation in /aɪ/, whether or not they use that variation in any meaningful way. This is a unique opportunity to tap individuals' perception during a sound change, to better understand the process of sound change for perception. The exploration of perception of /aɪ/ will contribute to the study of how sound changes for a listener, from the perspective of a listener rather than the perspective of a speaker.

This chapter consists of three experiments. The first task is a formant sensitivity task, designed to determine listeners' sensitivity to the small differences in formants that are produced by speakers in Fort Wayne. I predict that listeners will be able to accurately differentiate these differences because this task asks participants to decide which word out of four is different from the others, and is an acoustically-focused task compared to the other two tasks which will recruit linguistic perception. The second task is a preference task, where participants will be asked whether they prefer a raised or unraised version of /aɪ/. This task is designed to see whether participants' representation of /aɪ/ is raised or unraised, by analyzing which version of /aɪ/ they prefer and tapping into how they actually store /aɪ/. The third experiment is an identification task, in which participants heard one word and responded which of a minimal pair they heard. This task was intended to investigate how participants used raised /aɪ/ in pre-voiceless position, and whether raised /aɪ/ improved perception of words when it preceded voiceless consonants.

The results of all three of these perception tasks reveal that listeners are able to hear and accurately perceive both raised and unraised /aɪ/ in pre-voiceless position, without much preference for either the raised or the unraised version, and without much evidence that the pre-voiceless raised version of /aɪ/ helps listeners perceive words any better than the unraised version of /aɪ/. The most prominent result is that listeners do seem to be sensitive to types of vowel and consonant combinations that don't occur normally, such as a raised vowel paired with a voiced consonant. This indicates that listeners do have a level of awareness of acceptable versus unacceptable or non-occurring vowel-consonant combinations, they just don't have a preference or receive any perceptual benefit within those acceptable vowel-consonant combinations.

Methods

This chapter reports three perception experiments, including a 4IAX task, an identification task, and a preference task. Each task was designed to test a certain feature of representation: the 4IAX task was intended to investigate acoustic sensitivity to raising and fronting in /aɪ/, the identification task was intended to investigate perceptual usability of /aɪ/, and the preference task was intended to investigate individuals' representation of /aɪ/. More details about the specific methods for each experiment will be discussed within each section.

Subjects for this experiment were thirteen listeners from Fort Wayne, Indiana, recruited to participate online via a local college or online community board posting. The experiments were completed online using the experiment platform Gorilla; participants were instructed to sit in a quiet place with a stable internet connection, and to use headphones or earbuds to listen to the stimuli.

The stimuli for these three experiments were created using the speech of thirteen participants who had produced citation speech in Chapter 2. I included all of the citation speech that was available to create the tokens, to accurately reflect the range of variation in /aɪ/ productions that Fort Wayne speakers produce; in other words, I didn't choose which speakers to include and which to cut, I just included them all.

The stimuli were created by splicing the vowels of voicing minimal pairs (e.g., *Xait* and *Xaid*) and swapping the vowels from one word into the other consonant context, within a single speaker; this resulted in stimuli in which the vowel from *write* ended up with the consonants from *ride*. This was done via script that used the hand-labeled vowels marked in a textgrid, and used the nearest 0 crossing to the vowel boundary. A duration script was then run to adjust the length of the vowel to match that speaker's average vowel duration of pre-voiceless or pre-voiced monosyllabic words, in order to match the expected vowel duration preceding the voicing of the following consonant. Versions of words that kept their original vowels were also generated by splicing using the same splicing script: the vowel was just dropped back into the same in-word consonant context that it came from. Those "unedited" words were then run through the same duration adjustment script, so that their vowels were normalized to also equal the pre-voiced or pre-voiceless average vowel duration for that speaker. This also ensured that the comparisons that listeners were making weren't on the basis of edited versus unedited—both the splicing script and the duration manipulation script were used on all stimuli tokens used in the experiment. Thus the differences came in whether the same vowel was put back into the word or the vowel from the minimal pair word was dropped into the word. The tokens in which the vowel matches the consonant context will be referred to as "matched" and the tokens in which the vowel came from the opposite consonant context by voicing will be referred to as "unmatched." Table 2 clarifies the types of tokens used in these studies.

Word	Matched	Unmatched
write	Vowel from *write* + consonants from *write*	Vowel from *ride* + consonants from *write*
ride	Vowel from *ride* + consonants from *ride*	Vowel from *write* + consonants from *ride*

Table 2 Types of tokens in perception experiments, matched and unmatched

Seventeen sets of minimal pairs (both monosyllabic and bisyllabic) were used in the stimuli creation. These words are listed in Table 3.

advice	advise		price	prize
bite	bide		sight	side
biting	biding		sighting	siding
bright	bride		strife	strive
device	devise		tight	tide
dice	dies		title	tidal
height	hide		write	ride
ice	eyes		writing	riding
lice	lies			

Table 3 List of words used as stimuli

Approximately 20 tokens were not used because they were not produced accurately or with high enough quality recoding to use as stimuli. Overall, there were 440 functional tokens used as stimuli in the two versions of the three experiments. The thirteen speakers who produced stimuli includes two participants who were also subjects in this chapter, who heard their own spliced speech as approximately 12% of their stimuli in the perception experiments.

As described in Chapter 2, of the thirteen speakers, ten both raise and front in pre-voiceless /aɪ/ compared to pre-voiced /aɪ/ in monosyllabic words. Of the other three, two just raise and one just fronts /aɪ/ in pre-voiceless position in monosyllabic words. Four of the ten raisers + fronters also raise in flapped words.

Each participant received either an A version or B version of the study, which included only 8 of the 13 total speakers across the three experiments. The stimuli were split in this way to keep experiment times manageable, and to also make sure that all of the tokens were heard and could be included in analysis. In each experiment version, there were 6 or 7 raisers + fronters in monosyllabic words, 1 raiser, and one version had 1 fronter as well. Each version also had two pre-flap raisers, so listeners generally heard the same types of production of /aɪ/ in terms of fronting and raising.

Experiment 1—formant sensitivity, 4AIX

Predictions and Methods

The first perception experiment investigated formant sensitivity, specific to the /aɪ/ vowel. This experiment was designed to investigate whether the differences that people are producing can be heard by listeners and how sensitive Fort Wayne area speakers are to differences in F1 and F2 in /aɪ/.

Trials consisted of listening to four tokens, grouped in pairs with a pause in the middle. Participants were told that one of the pairs in each trial would be identical and one of the pairs would be different, and were asked to respond whether the first pair or the second pair was *different*. The stimuli in this experiment were created as described above, with vowels spliced from minimal pairs and durations normalized to match the voicing of the following consonant. The position of the "different" token in the set of 4 stimuli was rotated so the "different" token appeared in each position (1-4) approximately 25% of the time. The "different" token was always an unmatched token with the unmatched vowel that came from its minimal pair counterpart (i.e., a raised vowel in a voiced coda word or an unraised vowel in a voiceless coda word), compared to the 3 matched but still vowel-edited, “same” tokens.

Participants heard three basic types of tokens: matched consonant + vowel tokens, with both pre-voiced and pre-voiceless consonant context (these were edited in vowel duration to approximate the editing done to the other tokens); unmatched vowels in an acceptable context (i.e., unraised pre-voiced vowels spliced into pre-voiceless context); and unmatched vowels in an unacceptable context (i.e., raised pre-voiceless vowels spliced into pre-voiced context). I describe the two types of unmatched tokens as acceptable and unacceptable because in the pre-voiceless context, listeners hear both raised and unraised /aɪ/ vowels because there is variability in the speech they hear others produce. Therefore, in pre-voiceless consonant context, any vowel from either context should sound acceptable to the listeners, whether it came from a pre-voiceless or pre-voiced consonant context. However, listeners never hear a raised vowel in a pre-voiced context, so the unmatched tokens that have a pre-voiceless vowel spliced into a pre-voiced consonant context should sound unacceptable to listeners.

I predict that it will be the most difficult for listeners to distinguish small differences that are acceptable to them, such as in the pre-voiceless context in both matched and unmatched vowel cases. I expect that perceiving more difficult sounds will mean longer response times from listeners. Listeners hear a large variety of different vowels in the pre-voiceless case, including both raised and unraised, and therefore they will be slower to identify differences in that context. On the other hand, because the pre-voiced context is relatively more constrained and includes less variation, I expect that listeners will be faster to identify the differences in vowels in this context, because they're not used to hearing the pre-voiceless vowel in the pre-voiced context, or the matched case. Another prediction is that listeners will be slower to recognize the different tokens in the pre-voiceless monosyllabic words compared to both underlying voicing type disyllabic words and the pre-voiced monosyllabic words. While there are differences between the monosyllabic and disyllabic words in both the pre-voiced and pre-voiceless context in some speakers, fewer speakers produce raising in the pre-flap context. Because a speaker who raises in Fort Wayne will produce the raised vowel first in the monosyllabic context (Berkson et al., 2017), we expect listeners to experience the most variation in the monosyllabic context in their regular interactions, and therefore be less attuned to hearing differences in that monosyllabic context, and more used to a wider range of variation.

Results

Surprisingly, accuracy in identifying the different pair was at 100%. There were likely several contributing factors: the main one being that participants were untimed, to allow for any technical difficulties or loading time with the remote experiment environment, and so had as much time as they wanted to respond to each trial. The average response time across all trials was 817 ms, and response times longer than two standard deviations longer than an individual's mean response time were filtered out of the analysis; this amounted to approximately 100 responses.

The basic predictions in the previous section turn out to be true; overall, participants were able to correctly identify the different pair in 763 milliseconds in the pre-voiced context, and in 866 milliseconds in the pre-voiceless context. This difference in reaction time is likely due to the matched and unmatched cases both being acceptable in the pre-voiceless case, and therefore more difficult to distinguish between, versus the pre-voiced context, where only the matched vowel cases are acceptable, and therefore it's easier to distinguish between the matched and unmatched tokens.

In keeping with the above predictions, it is expected that it would be more difficult for listeners to identify small differences in vowels in monosyllabic, pre-voiceless /aɪ/, because they often hear a lot of variation in these vowels, and it all sounds acceptable to them. This set of stimuli indeed garnered the longest reaction times, averaging 932 milliseconds. On the other hand, listeners don't hear as much—or any—variation in the pre-voiced monosyllabic words, so it is likely easier for participants to identify differences in that context; listeners identified these differences in an average of 751 milliseconds.

Disyllabic, pre-flap words yield slightly different predictions, as listeners overall hear less variation in them because only the most advanced participants in the sound change produce raising in pre-flap context. These words (such as *writing* and *riding*) are also generally more confusable, because they have fewer distinguishing cues than the monosyllabic words, which differ in vowel duration and final consonant voicing, as well as vowel quality (raised or unraised). In the pre-flap context, the vowel duration and post-vowel consonant voicing are neutralized, and for non-raising dialects, the vowels are also the same; listeners must rely on sentence context to determine the word. All this said, listeners are also likely used to hearing a lot of variation in /aɪ/ in the pre-flap context, but not quite as much as in monosyllabic words because only speakers far along in the sound change are producing a difference in the pre-flap /aɪ/ corresponding to the underlying voicing of the following consonant. And as predicted, the reaction times for correctly identifying different tokens when comparing pre-flap, pre-voiced /aɪ/ and pre-flap, pre-voiceless /aɪ/ are very similar, 826 milliseconds and 801 milliseconds.

To model the expected longer reaction times for more difficult words to perceive, reaction times were modeled using predictors of word type with random effects for individual speakers, to account for individual differences. In a linear mixed effects regression model predicting reaction time by an interaction between voicing of following consonant and number of syllables (monosyllabic vs pre-flap disyllabic) as main predictors with random intercepts for listener and stimuli speaker, the voicing prediction was significant (estimate=86.33, $t= 3.57$, $p<0.05$), the binary number of syllables was significant (estimate= 77.84, $t=2.38$, $p=0.02$), and the interaction between the two main predictors was not significant (estimate=-75.35, $t= -1.65$, $p=0.1$). Response times were predicted to be significantly longer in the voiceless cases compared to the voiced case. Response times were also predicted to be significantly longer in the disyllabic words compared to the monosyllabic words.

Experiment 2—Preference AX task

Predictions and methods

The second perception experiment asks participants to respond which of two tokens sounds better to them. Stimulus pairs are a matched and unmatched token of the same word as described above. A matched token includes the vowel that was produced in the given word context with minor editing done, and an unmatched token is the vowel taken from the minimal pair word and spliced into the same consonant context as the matched token. Both tokens are duration-normalized to fit the expected vowel duration based on the voicing of the following consonant. In other words, participants heard the same word twice, with the same consonants and different vowels, one from the given word and one spliced in from the minimal pair. This was done to give participants a chance to hear a raised and unraised vowel to directly compare them and let us know which version they prefer. Previous experiments investigating the relationship between speech production, perception, and representation have relied on similar preference experiments. Whalen et al. (2004) and Johnson, Fleming and Wright (1993) had similar results in experiments where they investigated representation by asking participants to choose the best example of a particular vowel using synthesized speech. In both experiments, participants generally selected a hyperarticulated version of a vowel as the "best" example of that vowel.

Predictions for this experiment are tricky, because the outcome is expected to vary by participant. This experiment was designed to investigate representation of pre-voiceless /aɪ/, and presumably, there are some subjects in this study who have a raised pre-voiceless representation of /aɪ/ and others that don't. So, with the caveat that this experiment will be discussed more in the section on individual results in the perception tasks because that's how it was designed to be interpreted, there aren't really predictions about the outcomes of this experiment on the group level.

Results

Participants heard two tokens, a matched and unmatched vowel with the same consonant environment and were asked to choose which token they preferred. Participants generally preferred the matched token in both pre-voiced and pre-voiceless context compared to the unmatched token which included the vowel from the minimal pair word of the opposite voicing context. The preferences were fairly evenly distributed between matched and unmatched tokens for both pre-voiced and pre-voiceless words, with an approximately 60% preference for matched tokens and an approximately 40% preference for unmatched tokens. In the pre-voiced monosyllabic words, participants preferred the matched token 70% of the time. In the pre-voiceless monosyllabic words, participants preferred matched tokens 61% of the time. However, in the pre-flap case, participants showed little preference for either matched or unmatched tokens. They did not prefer either matched or unmatched pre-flap token in either voicing condition, with pre-voiced-flap matched tokens preferred 51% of the time, and pre-voiceless-flap matched tokens preferred 56% of the time. These results demonstrate that participants are actually attending to the vowel height and fronting differences in the tokens, rather than just preferring the matched tokens over the unmatched tokens. Participants greatly preferred the matched token in the pre-voiced monosyllabic cases (70%), where the unmatched token is an unacceptable vowel height with regards to the voicing of the following consonant. In the pre-voiceless monosyllabic cases, participants showed slightly less preference for the matched tokens over the unmatched tokens (61% preference for matched tokens), because the unmatched token is acceptable to listeners in that case. Participants showed little preference for matched or unmatched tokens in the pre-flap cases, because all variations of /aɪ/ height should be acceptable in the pre-flap context.

Similar to the previous experiment, longer reaction times are interpreted to indicate more difficulty in perception, or in this case, less of a preference for either the matched or the unmatched tokens. In a linear mixed effects regression model predicting reaction time by an interaction between voicing of following consonant and number of syllables (monosyllabic vs pre-flap disyllabic) as main predictors with random intercepts for listener and stimuli speaker, there were no significant results (voicing prediction: estimate=-37.916, t=-0.948, p=0.34, the number of syllables prediction: estimate= 5.608, t= 0.104, p=0.9, and the interaction between the two main predictors: estimate=-16.419, t=-0.22, p=0.8). Response times were not part of the predictions for this experiment, because they weren't expected to vary much by stimuli type, so this result is not unanticipated. The main variable of interest in this experiment was the preference for matched versus unmatched tokens within particular stimuli categories, and there were general trends that participants preferred matched tokens over unmatched in the voiced monosyllabic words, and show little preference for either matched or unmatched tokens in the disyllabic words. The results to the preference task were predicted to vary by participant, and will be analyzed by individual in Chapter 5.

Experiment 3: Word Identification Task

Predictions and methods

The third experiment in the perception portion of the dissertation investigates whether raising actually benefits processing of words with raised and fronted /aɪ/. This experiment explores whether listeners are better able to identify words with raising compared to words without raising. The experiment was a word identification task in which participants identified which of two words they heard, given the option of two minimal pairs on the screen to choose from. For example, listeners heard "write" and saw the options *write* and *ride* on the screen. The tokens used in this experiment are the same as in the previous two experiments, with half using a spliced vowel from its minimal pair word, and half using the original vowel from the word, and all with edited vowel length to match the average for that speaker, for consistency.

In this experiment, listeners heard stimuli from the same group of stimuli described above—matched and unmatched vowels in both pre-voiced and pre-voiceless context, in monosyllabic and bisyllabic syllables. Participants were asked to identify the word they heard, from the options of a minimal pair on-screen. For example, listeners heard "write" and were given the options of *write* and *ride*. Participants' responses were able to correctly identify all of the words with 100% accuracy, with differences in reaction time as again the variable to analyze.

Predictions for this experiment differ by syllable structure. For monosyllabic words, we might predict that raising could help identify pre-voiceless /aɪ/ compared to pre-voiced /aɪ/ because it's an additional cue to the voicing of the final consonant, but there are also other cues such as vowel length and the consonant voicing that the listener can rely on to identify the word, so effects will be small in this case. In the disyllabic words, these additional cues of vowel length and consonant voicing are neutralized in the flap, so the vowel quality will be the clearest means to identification.

Results

Participants were able to accurately identify all of the words in this experiment again, with slightly longer reaction times compared to the 4IAX task. Response times longer than two standard deviations above a listener's average response times were again filtered out; 239 trials were excluded by this standard. Data were also split into two analyses to consider responses to monosyllabic words and disyllabic words separately because they're different conditions and different results were expected by syllable structure, as discussed in the previous section.

Overall, the reaction times to trials in this task were longer than the reaction times to trials in the 4IAX task, indicating that it was a more difficult task. Similar to the previous two models, a longer reaction time was expected to indicate more difficulty in perception. In a linear mixed effects model predicting reaction time to monosyllabic stimuli by an interaction of main effects between voicing of final consonant and matched versus unmatched tokens, with random intercepts for subject and stimuli talker, both main effects and the interaction between them significantly predicted reaction time. Voiceless tokens significantly predicted slower reaction times (estimate= 29.33, $t=1.972$, $p=0.05$), unmatched tokens significantly predicted slower reaction times (estimate=104.84, $t=6.694$, $p<0.05$). The interaction between voicing and matched versus unmatched also significantly predicted reaction time, such that voiced matched tokens are faster to perceive than voiceless unmatched tokens (estimate=-75.63, $t=-3.579$, $p<0.05$). So, the overall results for the identification of monosyllabic words are that while participants were able to accurately identify the words, matched voiced tokens were the fastest, or easiest to identify.

In a linear mixed effects model predicting reaction time to disyllabic stimuli by an interaction of main effects between voicing of final consonant and matched versus unmatched tokens, with random intercepts for subject and stimuli talker, neither the main effects (voicing predictor: estimate= 36.55, t=1.271, p=0.204; matched versus unmatched predictor: estimate=39.37, t=1.370, p=0.171) or the interaction between them significantly predicted reaction time (estimate= -65.30, t=-1.604, p=0.109). This is a surprising result that indicates that for these speakers, the raised and fronted /aɪ/ makes no significant difference in their ability to identify words, in both the matched and unmatched, and the underlyingly voiceless and underlyingly voiced, flap cases.

Individual analysis of perception results

To see how listeners' performances across the perception tasks correlated, I considered individual speakers' perception results. If there's a path through sound change for production, with evidently different stages, it's possible that a series of stages of perceptual change exist too, such that individuals' performance on the various tasks should show their stage in the perceptual change. To investigate this possibility, I considered individual results in each of the three perception tasks, and determined where each participant ranked in each task. In other words, I ranked each participant by their performance in each task, and compared these rankings to see if the same listeners were top performers across all three tasks, indicating that some listeners were further along in the perceptual sound change.

I chose to approach this problem of comparing individual results across the three different perception tasks for a few reasons. First, the three tasks are fundamentally different, and had different average reaction times across the speakers. For example, participants performed the 4IAX task fastest, at an average of 817ms per trial; the preference task next fastest, at an average of 868ms per trial; and the Identification task slowest, at an average of 914ms per trial. Within individual listeners, participants all followed this pattern of being quickest at the 4IAX task, followed by the preference task, and slowest at the identification task; this reflects the relative difficulties of the tasks rather than participants' perception of the stimuli. In addition to these patterns reflecting the differences in difficulty of the tasks, the preference task was designed to investigate a separate part of perception altogether. Recall that the 4IAX task was designed to investigate whether listeners could simply hear the small acoustic differences produced in /aɪ/ in pre-voiced and pre-voiceless context, and the identification task was designed to investigate whether listeners actually made use of the patterned variation in /aɪ/ to perceive words with voiced or voiceless codas. Both of those tasks were intended to reflect perception of /aɪ/. On the other hand, the preference task was designed to investigate the listeners' representation of /aɪ/, just like the clear speech task; it was the intention that in forcing listeners to choose between the matched and unmatched tokens, especially in the pre-voiceless monosyllabic cases and in the pre-flap cases where both the matched and unmatched vowel-consonant combinations are attested in the sound change, would indicate whether the raised /aɪ/ or unraised /aɪ/ most closely resembled their own representation of the sound. The time it took listeners to choose their preferred token was informative because it indicated their confidence in their choices and was therefore analyzed, but in the interest of documenting the progression of a sound change through a listener's perceptual system and representation, their overall choice of matched or unmatched /aɪ/ in those pre-voiceless monosyllabic cases and disyllabic cases matters more in determining their progression in the sound change than the speed in which they made that choice. So, the first reason I chose this ranking analysis was to be able to compare the reaction times in the 4IAX and identification tasks to the results of the preference task, which were measured in percentage of matched tokens chosen.

The second reason that I chose to use the ranking system rather than raw or transformed time measures is because the reaction times are arranged on a continuum rather than clustered together, without many outliers or particular obvious groups. Figure 18 below shows box and whiskers plots of the reaction times to the three tasks, where each colored dot represents the average response time from a participant to certain stimuli in that task. The colors of the dots are consistent across the three experiments, and the participants that are slowest at one task are not necessarily slowest at the other two. Response times for the preference task are included for completeness, but as mentioned in the previous paragraph, the measure that best reflects the results of that task in capturing participants' representation is the proportion of "matched" tokens chosen, not reaction time. Because there are no obvious clusters of participants using either raw or log transformed reaction times, the rank analysis was the best way to compare participants to one another within the confines of the experiment.

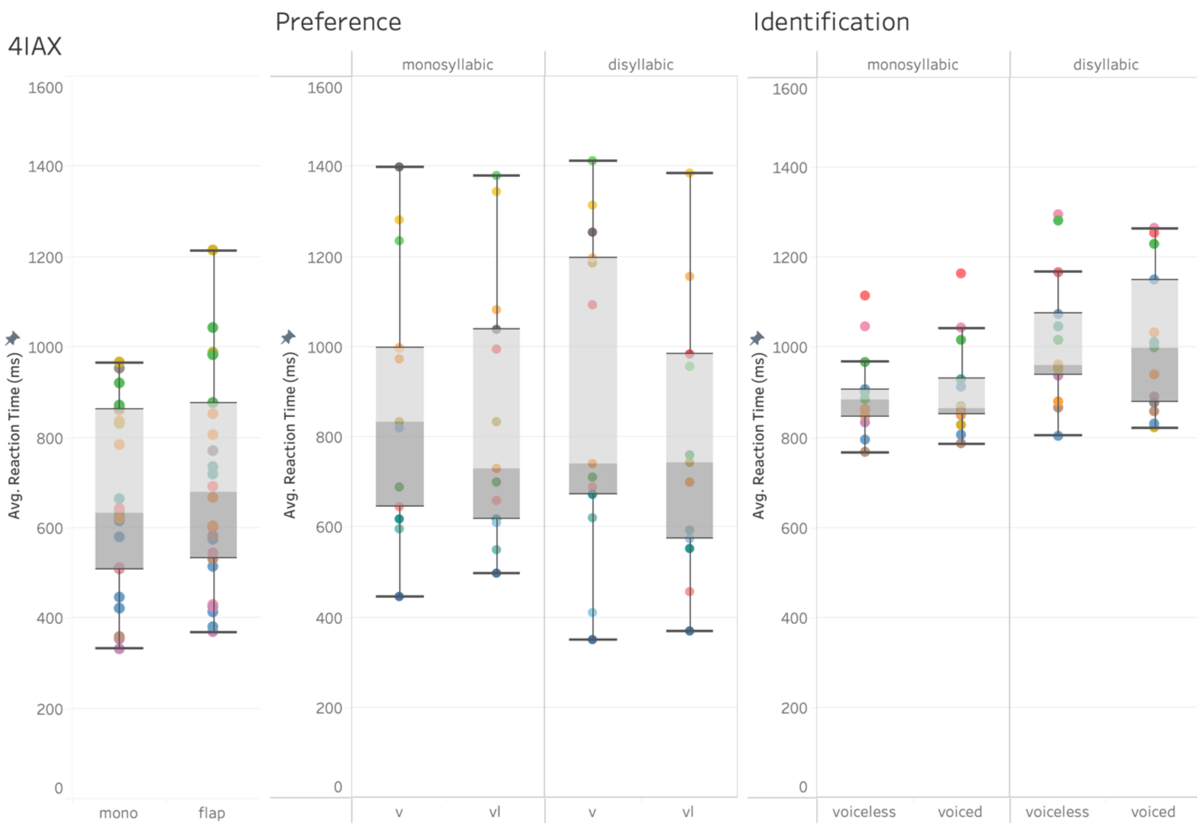


Figure 18 Box and whiskers plots showing the reaction times to each of the three experiments

To recap, in order to best compare two different types of results (reaction times and proportions) and to magnify potential differences among listeners that may have very similar average reaction times, the results to the perception tasks across individual speakers will be analyzed as ranked rather than raw results.

To perform the analysis of individual perception results relative to one another within the study, 4IAX and identification tasks were ranked by slowest to fastest reaction time. So, the participant with the fastest reaction times in the 4IAX task and in the identification tasks was ranked #1, and the participant with the slowest reaction times in the 4IAX and identification tasks was ranked #13, out of 13 total subjects. The preference task results were ranked on proportion of choice of the matched voiced tokens over the unmatched voiced tokens, because this was the category in which participants were predicted to have the strongest and most meaningful preferences. These rankings are shown below in Figure 19 (for monosyllabic words) and Figure 20 (for bisyllabic words), with the subject numbers along the x axis, and subjects' ranks in each experiment illustrated by the three bars; for all tasks, better performance yields a lower ranking number, so the best performers have the shortest bars.

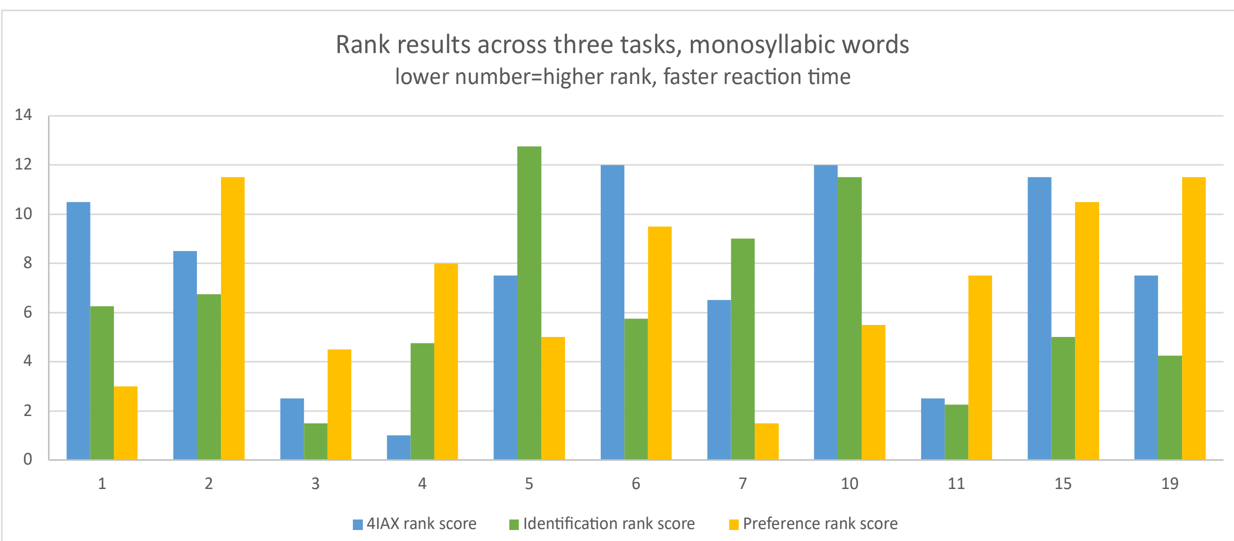


Figure 19 Ranking score by subject in each perception task, monosyllabic words; a shorter bar/higher rank means that participant did well in that task, relative to others

If there were consistent patterns of perception suggesting a systematic perceptual path through sound change, I'd expect to see subjects with similar ranks for the three tasks. That's not really what these rankings show. Subject 3, for example, ranks relatively highly in all three tasks, but most of the other speakers don't have similar standings where they're either near the top or near the bottom across all three tasks.

However, the preference task is fundamentally different from the other two tasks, both in that it's designed to access representation rather than perceptibility and in that the participants are ranked based on proportion of responses rather than reaction time. While considering only the 4IAX (blue) and Identification (green) results, the ranking results look slightly more similar per participant, but there's still not a lot of consistency across subjects in terms of how they compare to one another in response time to the monosyllabic stimuli. The rankings by each participant's responses for the three experiments' disyllabic words don't look much different from the monosyllabic rankings, and are below in Figure 20.

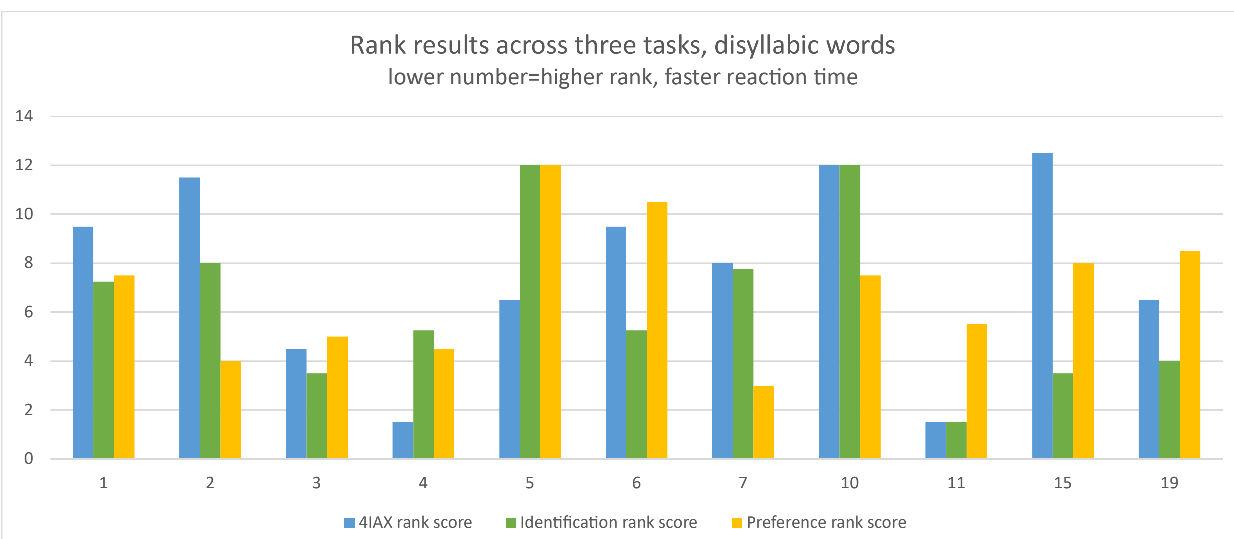


Figure 20 Ranking score by subject in each perception task, disyllabic words; a shorter bar/higher rank means that participant did well in that task, relative to others

Just as in the monosyllabic words, the participant rankings by responses to disyllabic don't show clear patterns that any speakers are better than the others across tasks. I included the rankings based on responses to disyllabic stimuli separately, because in my own previous research, I found that speakers who produced flap raising were better able to perceive a raised flap compared to speakers who did not produce flap raising in a word identification task (Strickler, 2019). I didn't find similar results here, although one encouraging thing to come from this individual analysis is that within individual speakers, their rankings across the monosyllabic stimuli analysis and the disyllabic stimuli analysis are consistent. This indicates that the ranking analysis is comparing the speakers to one another to try to reveal any patterns among them; it reassures me that the data aren't totally unpatterned, but rather, individual speakers are better at particular tasks than others, but that their performance in one task doesn't seem to correlate with their performance on any of the other tasks or their production patterns.

At this time, it doesn't seem that listeners in the Fort Wayne area demonstrate clear patterns in performance within individuals across the three perception tasks. This is an unexpected finding; the predicted results were that subjects would fall into predictable patterns in terms of perception, as they do in production. I predicted that the same listeners who were top performers in the 4IAX task would also be top performers in the identification task. If the results didn't match in that way, I expected that there would be a clear pattern and reason why such as, maybe a few top performers were good at 4IAX and identification, but some others were only good at the 4IAX task because it's easier and doesn't require as much evidence of a linguistic representation of a raised and fronted pre-voiceless /aɪ/ to complete compared to the identification task. I will discuss these results and their implications further in the following chapter, with additional information about the correlation between production and perception in the subjects of this study.

Discussion

In this chapter, I painted a picture of the perceptual aspects of /aɪ/ raising in Fort Wayne. Of the three experiments, the 4IAX task was designed to test sensitivity to small changes in formants, and the results were that listeners could easily pick out the different token in all cases, but were fastest to recognize the unmatched voiced token, a raised and fronted vowel paired with a voiced consonant, which is the type of token that is unattested in terms of the sound change because it never happens. In the preference task, listeners preferred matched voiced tokens, or unraised and unfronted /aɪ/ preceding voiced consonants compared to raised and fronted /aɪ/ preceding voiced consonants, but didn't show much preference for matched or unmatched in the disyllabic case. The lack of preference in the disyllabic case indicates that for these speakers, a raised /aɪ/ preceding an underlyingly voiceless flap is not part of their representation, or at least shares equal prominence with the representation of unraised /aɪ/ preceding an underlyingly voiceless flap. Finally, in the identification task, listeners were quickest to identify matched voiced tokens were the fastest in monosyllabic words, with no differences in response times to disyllabic words due to underlying voicing of the flaps or the matching status of the vowel.

In terms of results of individual speakers in perception, I set out to show that individual speakers are progressing in the sound change in Fort Wayne in terms of perceptual sensitivity, and to be able to identify the point that each individual is at in integrating the sound change into their mental lexicon. What the results actually turned out to be are more complex—listeners vary in their performances across the three perception tasks, seemingly without correlation between the three tasks by individuals. These results indicate that these subjects don't have a firm representation of a raised /aɪ/ preceding voiceless consonants. This discussion point leads into the next chapter, where I argue that at this stage in the sound change, it's actually disadvantageous for participants to have a firm representation of /aɪ/ that strictly matches their own production, because they must perceive other pronunciations of /aɪ/ when listening to other speakers who are at different points in the sound change.

Chapter 5

Discussion

While the various locations, patterns and features of American raising are described in previous studies, in this dissertation, I have expanded and elaborated on that production work, including thinking more carefully about what constitutes "/aɪ/ raising" and contributed work that describes perception in a sound change. In this chapter I combine a description of the processes prevalent in production and perception of /aɪ/ raising in Fort Wayne with a probe into the internal shift in representation of those participating in the change. This moment of variability in the grammars of speakers in Fort Wayne provides a unique opportunity to study representation and its relationship to production, due to the real-time variability in the production of the same sound. The representation is revealed (at least in part) through perception experiments and clear speech described in the previous chapters. It was predicted that the way that speakers utilize a raised and fronted /aɪ/ in perception could lend an interesting perspective to the question of why the production of /aɪ/ raising seems to be evolving in many dialects of English over time; however, the relationship between these two processes seems to be more complicated.

The current studies revealed a lot of variation in both production and perception. In the previous analysis chapters, I focused on the patterns that revealed themselves in the data. These patterns are convincing enough both in production and in the perception experiments taken as a group that it's striking that there aren't more clear patterns within individuals in the perception experiments. In this chapter, I will first discuss the potential relationship between /aɪ/ production, clear speech, and perception that was predicted. Then, I will present the evidence for the lack of correlation and propose an account for this process of /aɪ/ raising in Fort Wayne that accommodates the attested patterns in both production and perception, and an explanation for that lack of correlation.

Production of /aɪ/

Of the eleven participants for whom we have both production and perception data, eight fall into the pattern of raising and fronting /aɪ/ before voiceless consonants in monosyllabic words, as described in the Chapter 2. These eleven speakers are separated into four groups based on these categorizations of their production of /aɪ/ preceding voiceless consonants, voiced consonants, and flaps: participants who both raise and front /aɪ/ preceding voiceless consonants in monosyllabic words and also raise preceding underlyingly voiceless flaps in bisyllabic words (2 participants), participants who both raise and front /aɪ/ preceding voiceless consonants in monosyllabic words but do not raise in bisyllabic words (6 participants), fronting of /aɪ/ in pre-voiceless contexts in monosyllabic words only (2 participants), and no raising or fronting in /aɪ/ in pre-voiceless context (1 participant). These categorizations are summarized below in Table 4.

Category of production	Number of speakers
Raise + front /aɪ/ in prevoiceless monosyllabic words; also raise in pre-voiceless flap /aɪ/	2 (Subjects 10, 14)
Raise + front /aɪ/ in prevoiceless monosyllabic words	6 (Subjects 3,5,6,7,11,15)
Front /aɪ/ in prevoiceless monosyllabic words	2 (Subjects 1, 4)
No raising or fronting in pre-voiceless /aɪ/	1 (Subject 2)

Table 4 Categories of speakers who also have perception data

Perception experiments

As discussed in Chapter 1, I'm specifically interested in how sound change affects perception, or how sound changes for a listener. This section explores three aspects of perception of /aɪ/ raising for Fort Wayne listeners: sensitivity to formant changes in the 4IAX task, preference as a way to consider representation of /aɪ/ in the preference task, and the potential benefit of a raised /aɪ/ as a perceptual cue in the identification task. Generally, in the previous chapter, the results of these three experiments were that listeners were able to complete the tasks, but with longer reaction times and presumably more difficulty identifying and perceiving a raised and fronted vowel in the pre-voiced position, where it is unnatural. Otherwise, listeners didn't demonstrate a large preference or gain measurable benefit from perceiving a raised and fronted vowel in the pre-voiceless position.

However, there is quite a bit of variation in reaction times between the four groups of speakers outlined in the section above, which doesn't seem at all correlated to their production group. We might have predicted, based on Strickler (2019), for example, that speakers who produce raising + fronting would be able to perceive it "better," or show signs of advanced perception abilities, compared to speakers who aren't as far along in the change. Even with only one non-raiser/fronter to compare to, the data in the current study don't indicate any correlation between advancement of production and performance in the perception experiments, because they don't seem to behave much differently from the other speakers who do produce some version of raising and/or fronting. So, although the group-level results in the perception experiments indicated that listeners are sensitive enough to a raised + fronted /aɪ/ to be able to accurately perceive it, the results beyond that don't show any correlation between the perception tasks by individuals, but rather seem to reflect unpatterned variation by individuals between the three tasks in terms of performance in the tasks.

4IAX results by production groups

The lack of correlation between production and perception results is especially apparent in the 4IAX task. This task was designed to investigate acoustic sensitivity to raised + fronted /aɪ/. This task was expected to be the easiest of the three for speakers who did not raise or front /aɪ/ in pre-voiceless context, because it reflects acoustic perception rather than linguistic perception; listeners only had to hear acoustic differences in /aɪ/, not use the differences to differentiate between words (as in the identification task) or to decide between two potentially acceptable versions of a word (as in the preference task).

As described in the previous chapter, in this task, listeners were asked to choose one unmatched token out of four, by choosing the pair with unmatched vowels when presented with another pair with matched vowels. Unmatched vowels came from the minimal pair counterpart of the target word (i.e., listeners heard four tokens of "write," but in one token, the /aɪ/ vowel from "ride" was spliced into the consonants from "write"). In this design, the easiest token to hear as "different" is likely the unmatched voiced token, because that vowel and consonant combination is the one out of four that isn't produced normally, since naturally occurring words never have a raised vowel in a voiceless context, regardless of a speakers stage in the sound change. Otherwise, the matched and unmatched voiceless monosyllabic tokens and the matched and unmatched disyllabic tokens are all variations on types of tokens that listeners regularly hear in normal speech in Fort Wayne, and are therefore likely more difficult to hear as "different." Participants were able to complete this task at 100% accuracy. Within those accurate responses, there is variation in response times that can be analyzed by production group.

As we might predict, the participant who does not produce a difference in /aɪ/ preceding voiceless consonants correctly identified the unmatched token in an average of 861 ms for the pre-voiceless monosyllabic /aɪ/ words (like *write* and *bite*), 982 ms for the pre-underlyingly voiced (e.g., *riding*), and 984 ms for the pre-underlyingly voiceless flap words (e.g., *writing*). This participant was much faster at choosing the unmatched token for pre-voiced monosyllabic words (like *ride* and *bide*), at an average of 636 ms (averages summarized below in Table 5). These results indicate that this listener is good at hearing the unmatched variety of token that doesn't occur in real speech, which is a pre-voiceless vowel in the pre-voiced context. However, they were equally good at determining the unmatched token in the contexts where variation naturally occurs, as in pre-voiceless monosyllabic words, and preceding flaps that are underlyingly either voiced or voiceless.

	Reaction times to recognize monosyllabic unmatched token		Reaction times to recognize pre-flap unmatched token	
	Voiced	Voiceless	Underlying /d/	Underlying /t/
Raise + front monosyllabic /aɪ/, raise /aɪ/ flapped words (avg. RT 2 participants)	745 ms	852 ms	734 ms	856 ms
Raise + front monosyllabic /aɪ/ only (avg. 6 participants)	633 ms	702 ms	701 ms	689 ms
Front monosyllabic /aɪ/ (avg. RT 2 participants)	584 ms	605 ms	626 ms	618 ms
No difference between pre-voiced and pre-voiceless /aɪ/ (1 participant)	636 ms	861 ms	982 ms	984 ms

Table 5 4IAX results summarized from Chapter 4

However, even though it appears that the single non-raiser/fronter is slowest to choose the unmatched token among all of the groups of speakers, other speakers' results don't seem to extend this pattern. As mentioned in the previous chapter, all of the participants were quickest to choose the unmatched tokens in this experiment when they were voiced monosyllabic tokens; this result is expected because the unmatched voiced vowel-consonant combination is unattested in everyday speech and sounds incorrect to them. But within that result, it was further predicted that speakers with more advanced raising patterns would be able to choose the unmatched token more quickly than other speakers. In Table 5, the fastest responders overall are the participants who only front /aɪ/ in monosyllabic, pre-voiceless context, and the second slowest group of responders are the participants who raise in disyllabic, pre-flap context in addition to raising + fronting /aɪ/ in pre-voiceless, monosyllabic words. These pre-underlyingly-voiceless-flap raisers should be the most advanced speakers in the sound change, but are nearly the slowest to choose which token is unmatched in the 4IAX task—which again, was designed to be the least difficult of the three tasks. In other words, the subjects are not demonstrating a clear correlation between production and perception like we might expect, because speakers who are more advanced in the sound change are not performing better in this formant difference sensitivity task than speakers who are less advanced in the change or who are not producing a raised and/or fronted /aɪ/ in pre-voiceless position at all.

Identification results by production group

The identification task was designed to investigate whether listeners could actually use a raised + fronted /aɪ/ in perception of pre-voiceless /aɪ/ in monosyllabic and disyllabic words. The outcome predicted for this experiment was again that there would be some correlation between production group and performance in this task, specifically that speakers farther along in the sound change would perform better, or faster. Just as in the 4IAX task, accuracy was at ceiling and participants' response times do not seem to correlate with their production group in this experiment. In this section, I'll walk through the analysis of reaction times, and show that there are not patterns in the results by production group.

In the identification task, participants heard one word, and were asked to determine which word they heard of a minimal pair (e.g., participants heard "write" and were asked whether they heard *ride* or *write*). Participants heard four types of tokens: matched voiced, unmatched voiced, matched voiceless, and unmatched voiceless words. The unmatched tokens were the vowel from the minimal pair word with opposite voicing spliced into the word with the appropriate consonant voicing. For example, an unmatched voiced token would have the consonants of *ride* with the vowel from *write*. Mean reaction times by production group are shown in Table 6 for monosyllabic stimuli and analyzed below.

Reaction times to identify monosyllabic stimuli				
	Matched tokens		Unmatched tokens	
	Voiced	Voiceless	Voiced	Voiceless
Raise + front monosyllabic /aɪ/, raise /aɪ/ flapped words (avg. RT 2 participants)	887 ms	892 ms	982 ms	936 ms
Raise + front monosyllabic /aɪ/ only (avg. 6 participants)	849 ms	886 ms	951 ms	897 ms
Front monosyllabic /aɪ/ (avg. RT 2 participants)	793 ms	816 ms	926 ms	863 ms
No difference between pre-voiced and pre-voiceless /aɪ/ (1 participant)	860 ms	873 ms	877 ms	896 ms

Table 6 Reaction times to identify monosyllabic tokens summarized from Chapter 4

Once again in these identification data, unmatched voiced tokens, or the abnormal tokens that don't fit the types of variation that participants hear in Indiana, are slightly different than the other types of stimuli. This time, they're using these atypical stimuli to identify words, and so the response times to correctly identify the unmatched voiced tokens are slowest of the four types of tokens to identify. This parallels the results in the previous experiment, in which listeners were faster to identify the voiced unmatched because they were salient. In this case the abnormality is making the task harder, rather than easier, so the reaction time to identify the unmatched voiced tokens are slower than the rest. However, that's where the pattern in reaction times seems to end, and the pattern doesn't really seem to even hold for the non-fronter/raiser in the last row. Within groups, the reaction times to the other three types of stimuli (matched voiced, unmatched and matched voiceless) are very similar.

Reaction times to disyllabic stimuli with underlyingly-voiced and underlyingly-voiceless flaps are similarly without much pattern within speaker groups. The lack of overall pattern average response times to the disyllabic stimuli are not unexpected because there were no significant predictions of response times to these stimuli in the group analysis in Chapter 4. These averages are summarized for completeness below in Table 7, but it should now be clear that there aren't any particular trends that speakers who produce more advanced raising and fronting are performing substantially better or even much different than their counterparts who only front /aɪ/ in monosyllabic words or the participant who produces no fronting or raising in /aɪ/.

.

Reaction times to identify disyllabic stimuli				
	Matched tokens		Unmatched tokens	
	Underlying /d/	Underlying /t/	Underlying /d/	Underlying /t/
Raise + front monosyllabic /aɪ/, raise /aɪ/ flapped words (avg. RT 2 participants)	1109 ms	1094 ms	1048 ms	1060 ms
Raise + front monosyllabic /aɪ/ only (avg. 6 participants)	921 ms	976 ms	956 ms	939 ms
Front monosyllabic /aɪ/ (avg. RT 2 participants)	968 ms	945 ms	960 ms	953 ms
No difference between pre-voiced and pre-voiceless /aɪ/ (1 participant)	1038 ms	955 ms	959 ms	1079 ms

Table 7 Reaction times to identify of disyllabic tokens summarized from Chapter 4

Preference results

As mentioned above in the results section of the Preference section, this task was designed to investigate individual speakers' representation of /aɪ/, by asking whether they prefer a raised and fronted vowel in the pre-voiced position. I'll start by discussing the preference results in the disyllabic words because the results are a little clearer, then move on to the preferences in the monosyllabic words.

Overall in the pre-flap context, participants did not show a preference for either the matched or the unmatched token. This is not unexpected, because in the pre-flap environment, the words all had a minimal pair and without text identifying the word, it's difficult to know whether you're hearing *writing* or *riding*. If it's difficult to know which word you're hearing, much less be able to decide which vowel sounds better with the consonants. These preferences in disyllabic words are visualized below in figure 20. In this figure, the words with underlyingly-voiced flaps (e.g., *riding*) are shown together in one column; the percentage of matched, pre-underlyingly-voiced-flaps are shown in purple compared to unmatched pre-underlyingly-voiced-flaps, which are shown in blue. The underlyingly voiced flaps are depicted in the left column for each speaker with darker colors, and the underlyingly voiceless flaps are depicted in the right column for each speaker with lighter colors; subjects are ordered by production group. The "flap" group refers to speakers who raise + front pre-voiceless monosyllabic /aɪ/ and also raise /aɪ/ in underlyingly-voiceless flapped words, "raise + front" refers to speakers who raise + front in pre-voiceless /aɪ/ monosyllabic words, "front" refers to speakers who front monosyllabic pre-voiceless /aɪ/, and "none" refers to the speaker who produces no difference between pre-voiced and pre-voiceless /aɪ/. These speakers' preference percentages are shown in Figure 20.

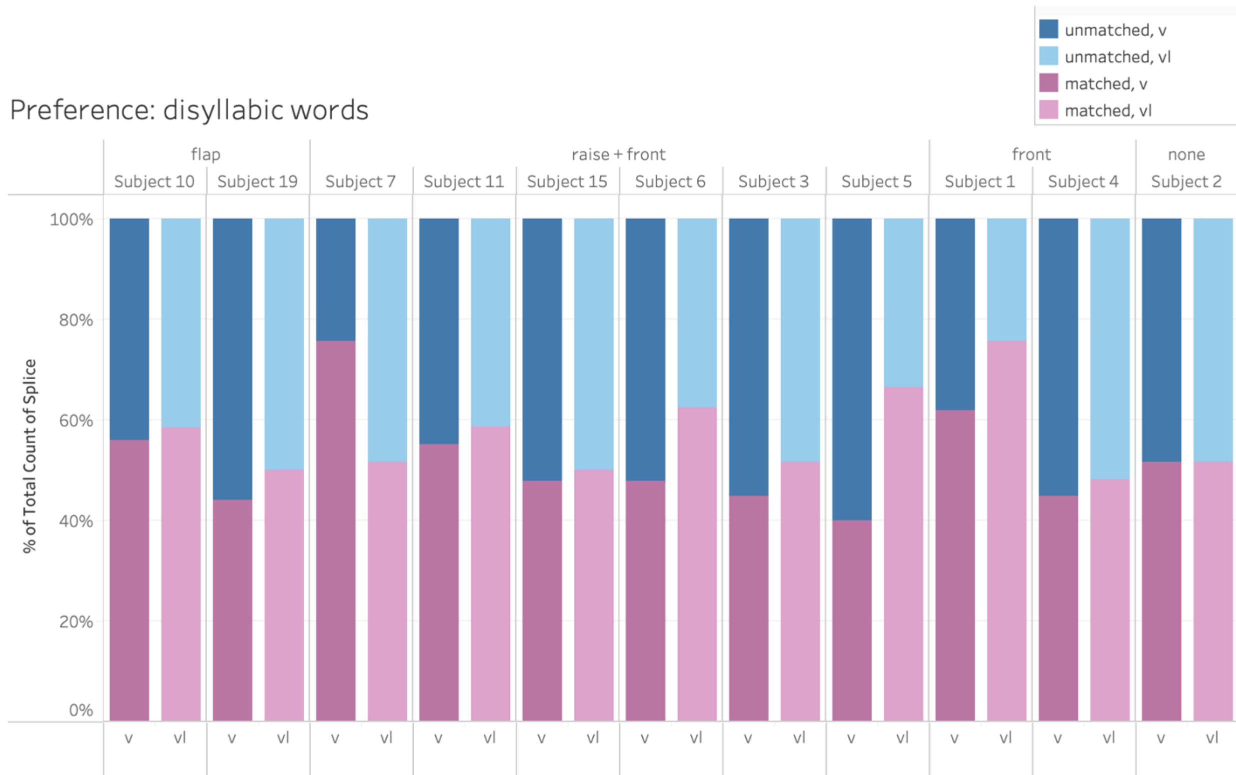


Figure 20 Percentage of preference of matched tokens in disyllabic words

As expected, there's not a clear preference for matched over unmatched tokens for any speaker in the disyllabic case, because it's difficult to tell which version of a flap you're hearing out of context. Likely either the matched or unmatched vowel sounded fine to listeners in context, and the preference rates average around 50% for the matched tokens. The preference proportions for the monosyllabic words are shown in Figure 21, and are more interesting because they're less evenly distributed between matched and unmatched tokens.

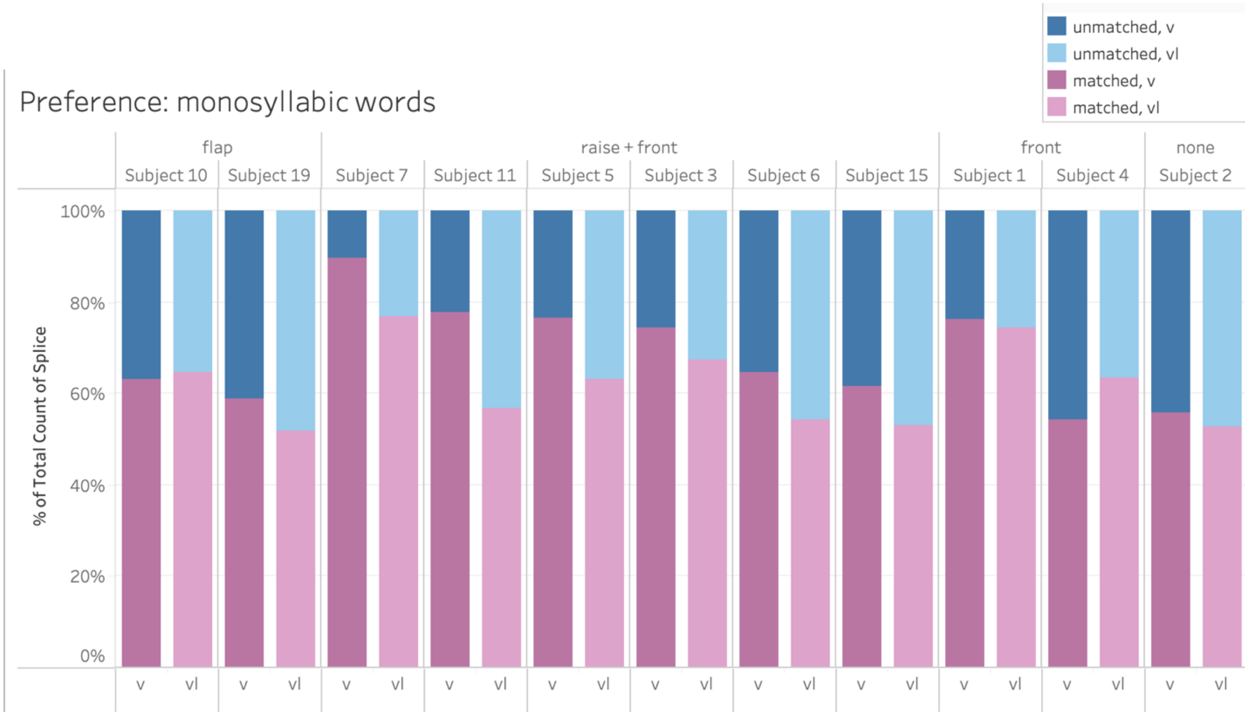


Figure 21 Percentage of preference of matched tokens in monosyllabic words

In the monosyllabic words, listeners had a higher percentage of preference for the matched tokens, especially in the voiced cases. Recall that I was interested in whether listeners preferred the matched tokens over the unmatched tokens in the pre-voiceless words, both monosyllabic words and underlyingly-pre-voiceless-flaps in disyllabic words. While a few subjects seem to show a clear preference for the matched tokens, especially in the voiced monosyllabic case where it's really expected, these speakers are distributed across the production groups, and the speakers in the "flap raising" group, who we'd expect to be most advanced and have the strongest preference of the matched tokens, don't. There again aren't any noticeable patterns by speaker group in these data, nor any notable results within individual speakers that clearly point to a raised representation in /aɪ/, which would surface as an obvious preference for matched voiceless tokens over unmatched voiceless tokens.

Individual perception results analysis

In the previous chapter, I presented an analysis of the perception results of individual speakers, compared across the three perception tasks showing that individuals did not appear to perform in patterned ways across the three perception tasks. Individuals who were fastest at hearing the unmatched tokens in the 4IAX task were not necessarily also fastest at identifying the word in the identification task, and vice versa—listeners who were good at identifying tokens quickly in the identification task were not necessarily quick to choose the unmatched tokens in the 4IAX task, as would be expected if perceptual change is following a systematic progression. Listeners' preferences for matched tokens in the pre-voiced environment did not seem to correlate with the results of the 4IAX or the identification task either.

The explanation for these unpatterned results can perhaps be found in the varied production of /aɪ/ in Fort Wayne. In Chapter 2, it was demonstrated that speakers produce a variety of raising, fronting, and both in /aɪ/ preceding voiceless consonants, and preceding underlyingly-voiceless flaps. Given this variability, it wouldn't be efficient for listeners to perceive this variety of speech based only on their own pronunciation, requiring additional effort to perceive other pronunciations of /aɪ/ that they hear all the time. If there are at least four acceptable variations of /aɪ/ preceding voiceless consonants, including the same /aɪ/ as preceding voiced consonants, a raised /aɪ/, a fronted /aɪ/, or both raised + fronted /aɪ/ that are all being produced regularly in Fort Wayne, listeners need to be able to accurately perceive these different versions of /aɪ/ for what they are.

This could also explain why the effect that was most prevalent in the 4IAX and the identification tasks was the salience or markedness of the unmatched voiced tokens; in the 4IAX task, the noticeability made the task easier, as was reflected in faster reaction times across the board, and in the identification task, the incorrectness made the task more difficult, because listeners had a harder time identifying the words, which was reflected in longer reaction times. But beyond those two results due to the salience of the incorrect tokens, no other results in the perception tasks were interpretable as reflecting stages of an ongoing sound change, or a representation change for the listeners, because there weren't obvious groups to be made by perceptual behavior like there are for production data.

The lack of consistency across perception tasks does not point to a changing mental representation of /aɪ/ (or perhaps [aɪ] and [ʌɪ]) that listeners are using in day-to-day life. The clearest result from the three perception tasks is that to the extent that there are any patterns in perception, it's that the unnatural token type is the easiest to perceive, and being able to perceive something that doesn't occur normally isn't actually useful for these listeners. These results instead point to an ongoing acceptance of the variation that's present in everyday speech, without perceptual preference for one version (that's actually produced by real speakers) over another. This lack of evidence for a perceptual representation of a phonologized form of this raised vowel is also reflected in the preference task results, which may more directly reflect an individual's representation of /aɪ/; speakers again only showed really clear preferences for the matched voiced tokens, over the saliently incorrect unmatched voiced tokens. The rest of the results did not support a preference for either a raised or unraised /aɪ/; for the most part, even within a speaker there was not a real preference for matched or unmatched /aɪ/ in the pre-voiceless context.

So, I propose that listeners don't demonstrate evidence of a changed representation because it's not in their best interest to do so at this time. They must continue to perceive a variety of /aɪ/ in pre-voiceless position, and their perception and representation results reflect that variety.

Clear speech

The clear speech experiment results can be tied in with the rest of the results linking production and perception of /aɪ/. There were four speakers who participated in both the clear speech task and the perception experiments, and with such a small number of participants, I refer to the results given in Chapter 3 to support my claim that Fort Wayne listeners don't seem to have adopted a raised and fronted /aɪ/ into their representation.

Recall, in the clear speech experiment, participants were asked to read a word list twice, once as though they were talking casually to a friend, and once clearly, making special effort to be easily understood. In comparing the more casual citation speech to hyperarticulated clear speech, participants did one of two things with /aɪ/: either they enhanced the two component parts of the diphthong as they would with monophthongs, by making the first half backer and the second half higher and fronter in clear speech (ten speakers); or, they actually undid what would be the raising and fronting processes, by making the diphthong lower and backer in clear speech (two speakers).

This experiment was designed to investigate the speakers' target /aɪ/, which should be more closely achieved in clear speech; I predicted that if a raised and fronted pre-voiceless /aɪ/ was part of a speaker's representation, they would enhance raising and fronting in clear speech. Not only do we not see that in the clear speech study, we actually see some evidence of the opposite effect, that speakers are undoing the raising and the fronting. This is evidence that speakers are not striving towards a raised and fronted pre-voiceless /aɪ/ as a target in clear speech, and further evidence that these speakers do not have a raised and fronted representation for /aɪ/.

A return to sound change theory

Although I didn't find evidence of a stepwise progression through sound change for perception in Fort Wayne speakers, nor a direct correlation between their progress in production and in perception, I don't believe these results are outside the predictions of theories of sound change like Ohala (1981), Blevins

(2004), Lindblom et al. (1995), or Beddor (2009). As discussed in Chapter 1, these theories of sound change describe how perception allows for sound changes. Of the multiple theories discussed, sound change originates several ways: a misperception of the speaker's intention in the case of Ohala and Blevins, speaking with intention to make oneself easily understood for Lindblom et al., and simplifying production for the speaker when the listening conditions allow for it for Beddor (2009). The results here don't contradict those theories; the sound change in Fort Wayne could have originated as any one of the mechanisms proposed by Ohala, Blevins, Lindblom et al., or Beddor, and there's no clear evidence one way or another presented in this study.

These theories of sound change all focus on the initial mechanism of sound change, whether that be a misperception of a phoneme, a gradual drift towards hypoarticulated speech, or a merging of cues onto a single phoneme due to coarticulation. Any one of those explanations could apply to /aɪ/ raising: to speak directly to Ohala's theory of sound change as it includes the listener, there is no evidence at this point to either prove or disprove the theory that /aɪ/ raising could have begun as a consequence of shortening the vowel in pre-voiceless position and concurrently "raising" the nucleus in order to reach the target of the glide, which was then misperceived as "intentional" and then reproduced by other speakers—although, to maintain this theory, the speakers will eventually need to develop some sense of awareness and control over the raising, which the clear speech results do not support at this time. It also appears that listeners are not actively using /aɪ/

raising to identify words at this point in time, and this taken with the results of the clear speech experiment demonstrate that speakers are not increasing /aɪ/ raising to try to improve perceptibility, either intentionally or unintentionally. Beyond that, it is difficult to compare a snapshot of the ongoing progress of an incomplete sound change to theories that focus mainly on the initial mechanism of the change. The mechanism of change in Fort Wayne, which I will define here as the single point in time when one speaker's representation of /aɪ/ shifted to include both raised and unraised /aɪ/, is obviously unobservable and open to speculation at this time. In that sense, nearly any of these previous theories on the mechanisms and dissemination of sound change and the role that perception plays in that process could be applied to the *mechanism* that initiated /aɪ/ raising in Fort Wayne, without changing the current state of the sound change. In other words, the theories discussed previously in this section and in Chapter 1 predict the very beginning stage of the sound change, which has passed in Fort Wayne; the goal of my work was to document the different stages in production that are currently developing in parallel during the progression of the sound change, and to try to document a similar set of stages in the perception of the sound change as well.

Adding to this difficulty in interpreting a single point in time in the process of /aɪ/ raising in Fort Wayne as part of an established theory of sound change is that it doesn't recognizably mimic any many of the common categorizations of sound change. Garrett and Johnson (2011) provide a thorough summary of many types of sound change related to both speaking and listening as have been described over the past century, including theories of Ohala, Blevins and others, organized by type of change. They discuss four types of environments that provide variation that can lead to sound changes including motor planning errors, such as blending two consonants in a cluster together; aerodynamic constraints, which can lead to voiced stop devoicing; constraints in gestural mechanics, in which one gesture overlaps another and renders the other masked in perception; and differences in perceptual parsing, which includes hypocorrection as is key in the theories of Ohala, Blevins, and Lindblom et al. as mentioned above. Garrett and Johnson (2011) use these categories to group together common types of sound change such as assimilation, palatalization, vowel harmony, metathesis, deletion, and more. They note in the introduction of the chapter that not all sound changes can fit into these neat boxes, and the splitting of the /aɪ/ vowel conditioned by voicing of the following consonant and by syllable structure certainly doesn't sound like any category of sound change commonly described around the world; and yet, this specific sound change does happen, repeatedly and independently, in dialects of English in the US and beyond (e.g., Fruehwald, 2016; Hualde, Barlaz & Luchkina, 2022).

To categorize /aɪ/ raising into a category of sound change described in Garrett and Johnson (2011), the best fit may be a combination of gestural constraints and perceptual parsing. It seems both possible and likely that the effort to produce the full gesture /aɪ/ is difficult to complete in syllables where the vowel precedes a voiceless consonant, where there is a shorter time frame in which to complete the movement compared to pre-voiced syllables; from there, it also seems possible that other listeners could interpret that original change due to gesture timing constraints as intentional, and before long it spreads throughout the lexicon. Garrett and Johnson also call this type of change articulatory enhancement, which also describes incremental differences present in variation that slowly become more patterned and phonologized. Theories like those of Ohala (1988), Blevins (2004) are included in the articulatory enhancement, and that of Moreton (2004) as a type of gesture timing constraint as offglide peripheralization in the case of /aɪ/ raising in particular, as discussed in Chapter 2.

Garrett and Johnson (2011) also incorporate word-based exemplar representations into their description of the different types of sound change. They describe sound change in exemplar terms as a cloud of exemplars that exist within a category of perception, and if this cloud shifts over time, that is sound change. They go on to say that the cloud, or number of representations in a particular category, must be larger for perception than for production; they say that this is because listeners are able to perceive many more forms than they are able to produce. In this case, not everything a listener hears becomes a part of their speaking representation, and are also not even necessarily incorporated into their listening representation. They provide evidence for this dual-representation model, including exemplars for both speaking and listening, in the differences between speech and hearing errors. They write that speech errors often include segments that interact with one another in some way, as in two phonemes swapping places or blending together articulatorily; listening errors, on the other hand, often have to do with the meaning of the phrase and do not often involve these articulatory-based errors. This interpretation of representation change as at least partially exemplar-based is also reminiscent of Beddor (2009), who describes coarticulation of nasal consonants and vowels that seems to be slowly moving towards sound change. The description of the dual-representation model nicely reflects the results in both the production and perception experiments: it seems that while speakers are producing /aɪ/ raising, listeners aren't terribly sensitive to hearing it, or willing to manipulate it in clear speech. This difference could either be described via Garrett and Johnson's dual representation model, or my own hypothesis that it's not beneficial at this stage of the change for listeners to narrow their perceptual representation to closely mirror what they themselves produce, when they must perceive other forms of /aɪ/ as well.

Finally, Garrett and Johnson make the very case for multiple strategies or paths to reach the very same goal in sound change, as was my argument for the variation in approaches in the production study in Chapter 2. They write:

> "…language learners may develop different articulatory strategies for realizing the 'same' acoustic target. It may be that two such strategies yield perceptibly different outcomes in some contexts, such as coarticulation; this could be the point of entry of a sound change." (Garrett and Johnson, 2013, p. 38).

Although Garrett and Johnson describe the "entry point" or the mechanism point in time discussed above rather than the progress towards stabilization of sound change as is happening in Fort Wayne, this prediction describes something similar to what I found in the results of the different approaches to /aɪ/ raising. In these varied approaches, some participants raised pre-voiceless /aɪ/, some fronted pre-voiceless /aɪ/, and some both raised and fronted pre-voiceless /aɪ/, but they appear to be participating in the same process overall, which is the process that has traditionally been called /aɪ/ raising but has been documented to also include fronting in several cases (i.e., Hualde, Luchkina, & Eager, 2017; Moreton & Thomas, 2007).

One key difference between my findings and Garrett and Johnson's hypothesis is that I described acoustic differences rather than articulatory differences in approach. I hypothesized that this multiple-approach method to achieving /aɪ/ raising and/or fronting will eventually become more "stable" or standardized, and that may be true for this case of multiple acoustic paths to similar perceptual outcomes. However, their theory suggests that perhaps variance in articulatory approach to reach the same outcome could be a regular occurrence, and so perhaps the variation in approach to raising and/or fronting /aɪ/ will also remain varied, but will produce perceptually similar vowels.

Conclusions

This dissertation set out to compare the progression of sound change in both production and perception within individuals. In the component parts of that analysis, valuable insights were gained regarding both production and perception of /aɪ/ in Fort Wayne, Indiana.

In chapter 2, production data were analyzed using three different techniques to determine the best analysis method to use when studying /aɪ/ raising. I compared the same data as analyzed using a traditional difference of 60Hz difference in F1 between pre-voiced and pre-voiceless /aɪ/ at the midpoint of the nucleus; an analysis to compare the trajectory lengths of the nucleus and coda separately, and included both halves of the diphthong and both F1 and F2; and, finally, I used generalized additive mixed models to predict the formant tracks of pre-voiced and pre-voiceless /aɪ/, to determine where exactly along the formant track significant differences occur. These analyses contribute not only to future analyses of /aɪ/, but also revealed theoretical implications about the origin point within the vowel of /aɪ/ raising in Fort Wayne. There is evidence to suggest that speakers begin to raise and front /aɪ/ in a variety of ways, including starting from the back of the vowel and "unzipping" the formants from one another, raising only at the nucleus midpoint and not from the end of the vowel, and a combination of the two approaches.

In perception of /aɪ/, listeners are sensitive to saliently incorrect tokens, namely unmatched voiced tokens where a raised and fronted /aɪ/ vowel is paired with a voiced consonant. While listeners are sensitive to differences that are perceived as wrong, raising and fronting /aɪ/ in pre-voiceless position doesn't aid perception of tokens with /aɪ/ preceding voiceless consonants.

Finally, there is no evidence that a raised and fronted representation of /aɪ/ preceding voiceless consonants is present in the participants in this study. Not only does a raised and fronted /aɪ/ not aid in perception in the identification and 4IAX tasks, but speakers also don't strive for a raised and fronted target in clear speech, and don't overwhelmingly choose a raised and fronted pre-voiceless /aɪ/ in the preference task; both of these tasks were designed to be indicators of representation of /aɪ/.

Previous work has supported the idea that in the Fort Wayne raising sound change in particular, there's a particular path through sound change in production that speakers follow. In the case of Fort Wayne, it's been proposed that speakers raise first in unstressed syllables, followed by pre-voiceless /aɪ/ in monosyllabic words, and finally preceding underlyingly voiceless flaps in disyllabic words, as the final puzzle piece to complete phonologization of [aɪ] and [ʌɪ] (Berkson, Davis, & Strickler 2017). This process is assumed to proceed in this way for the main reason that speakers who produce raising in pre-underlyingly-voiceless flaps also produce pre-voiceless /aɪ/ raising in monosyllabic words. In other words, there's no pre-flap raising without monosyllabic raising, and that's compelling evidence for a progression of sorts, because there is monosyllabic raising without pre-flap raising. The data in my second chapter support this as well—of the four speakers who raise /aɪ/ in pre-underlyingly-voiceless flaps, they all also produce a raised and fronted /aɪ/ in pre-voiceless monosyllabic context.

However, although this progression of /aɪ/ raising through the lexicon with regards to syllabic word context is well documented and supported, my production data also reveal that there are additional phonetic pieces to the sound change. Although we call the process /aɪ/ *raising*, I and others present evidence that it's more than just a decrease in F1, or a higher vowel, in that first half of the diphthong. I, and other previous research, have shown that for most speakers, it's not only /ɑ/ raising, but /i/ fronting as well (e.g., Moreton & Thomas 2004; Hualde et al. 2021).

I've also demonstrated that speakers appear to reach that progression differently. In the production chapter, I laid out how some speakers are reaching the raising and fronting processes back-to-front throughout /aɪ/, similar to Moreton & Thomas's (2004) theory of the origin point of /aɪ/ raising in Ohio speakers. I also showed that approximately one third of my speakers don't appear to begin the process at the end of the vowel, but rather raise only in the nucleus. Finally, 5 participants exhibited a combination of both strategies in pre-voiceless /aɪ/. And although I don't have diachronic data, I predict that these speakers will eventually reach a point of total separation between both F1 and F2, and reach the same point of raising and fronting no matter how they begin.

My data show that there might be different ways of being a raiser/fronter. These may or may not be related to the degree or stage of raising in that individual, but it seems like they're not. It seems like a difference in approach, or two means to the same end; because it does appear that speakers end up in the same place, or we wouldn't see the pre-underlyingly voiceless flap raising that is the same pattern across speakers, that also encompasses all earlier forms of /aɪ/ raising and fronting in monosyllabic words, and although I didn't include any in my analysis, hyper-short unstressed syllables like *citation*, too.

In perception, I hoped to see a similar path through sound change, and to see patterns in the different experiments that were designed to investigate different features of perception of /aɪ/. Instead what I see is variation in perception, similar to the variation that we see in approach to raising and fronting in production, and similar to the variation in how speakers adjusted pronunciation in clear speech—there are still apparent patterns, that can be described systematically. I imagined that clear speech was going to show that speaker's targets and representations of /aɪ/ were raised and fronted, at least in some cases, but that isn't what happened. Subjects did different things in perception, that aren't seemingly correlated to their patterns of production, or even to their results in other perception tasks. There's also not evidence that anyone is targeting a raised and fronted /aɪ/ in clear speech. There's good evidence that listeners are sensitive to the unmatched voiced tokens that are incorrect in the 4IAX and identification tasks, but there's no evidence that the raised vowel in the voiceless context is aiding in perception either. If speakers aren't producing a pre-voiceless raised and fronted vowel a raised vowel in clear speech, and the raised and fronted vowels in the appropriate context aren't helping participants identify tokens faster, it doesn't seem to be a part of representation.
I propose that this is because perception can't follow set patterns of development in the way that production does, because there are several patterns of production that you have to be able to accommodate as a listener. Perceptual patterns in a sound change don't actually make sense when listeners operate in a world in which they

could hear any combination of raised and fronted /aɪ/ depending on who they're talking to.

In chapter 1 I distinguished this dissertation as attempting to describe what happens *for* perception to change in a sound change, rather than how perception affects production in a sound change; so while the results of the dissertation don't provide opposing evidence to those theories of sound change in perception that affects production, the results also don't provide a clear picture of how sound changes for perception, for a listener.

I predicted that perception would evolve in steps or fairly linearly over the course of the sound change, similar to the steps previously documented in production. Participants who were producing more advanced forms of /aɪ/ raising, like pre-flap raising would be better at perceiving these more advanced forms of raising, compared to other who don't produce raising in that context. While this would've created a nice clear picture of sound change, instead the process of perception undergoing sound change seems to be much murkier. There don't seem to be patterns or steps to perception of raised /aɪ/ within individuals, as discussed at the end of the previous chapter; there also isn't a correlation in these participants between those who produce a raised /aɪ/ and those who are "best" at perceiving it, either acoustically in terms of the 4IAX task or in a useful sense as in the identification task. These results indicate that there is no pattern to acquiring perception of raised /aɪ/ revealed by this research. This lack of a changed representation for perception can actually be beneficial to the listener; while there is still so much variation in production of /aɪ/ in pre-voiceless context, where some speakers are producing raising only in monosyllabic pre-voiceless words, others are producing raising in both monosyllabic and disyllabic pre-voiceless words, and still others are producing no raising at all, it would not be in the listeners best interest to be over-specialized in perceiving a raised /aɪ/ at this time.

This is exactly what the results show—not only do most listeners not show clear preference for either raised or unraised /aɪ/ in pre-voiceless monosyllabic words or in disyllabic words (where either raised or unraised /aɪ/ is attested and acceptable to them), they show a clear preference *against* raised /aɪ/ only in the pre-voiced monosyllabic case, where raised /aɪ/ is unattested. This demonstrates their relative sensitivity to know what is happening around them in the sound change: regardless of what type of /aɪ/ they themselves produce, they show no preference to what type they hear, unless it's not unattested in every day speech. Perception is playing a different role than anticipated for listeners in Fort Wayne. Rather than adjusting their perception of /aɪ/ to match their own production of /aɪ/, they've adjusted their perception of /aɪ/ to accommodate *all* potential productions of /aɪ/ that they encounter on a daily basis, and haven't shifted to prefer one version over the other.

At this time, there's no evidence that for listeners in Fort Wayne, the evolution of perception of /aɪ/ will mimic the evolution of production of /aɪ/. And in fact, the results of the production study in chapter 2 may call into question the evidence of clear steps in production of /aɪ/ raising in Fort Wayne as well. For these speakers, considering both F1 and F2 and both the nucleus and the glide, there were three separate approaches to /aɪ/ raising, with participation in each relatively evenly distributed across the subjects. Not only is this variation in approach to raising and/or fronting another reason for listeners not to lock in a perceptual representation, it also indicates that the path to clearly produced raised /aɪ/ could be more complicated than previously thought.

Future directions

In this chapter, I discussed that there were not clear patterns in the perception data of individual speakers, either indicating more advanced perceivers or parallel development of production and perception of /aɪ/ raising. I proposed that this is because it is more beneficial for listeners to maintain a broad representation that accounts for all variation in /aɪ/ raising at this time, and that reflects the amount of overall variation in the production of /aɪ/ in Fort Wayne. Future research will need to return to this sound change when it is further advanced to determine whether perceptual representation of raised /aɪ/ ever becomes more specialized in the area.

Maybe when the sound change stabilizes a bit and there is less overall variation in production, there will be more predictable patterns in perceptual behavior. Maybe we're at the wrong stage in the sound change to capture a clear path through sound change. Or maybe perception participates differently in sound change altogether, and we'll have to ask different questions to determine what sound change looks like from the listener's experience.

References

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 785-821.

Berkson, K., Davis, S., & Strickler, A. (2017). What does incipient/ay/-raising look like?: A response to Josef Fruehwald. *Language*, *93*(3), e181-e191.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Bybee, J. (2003). *Phonology and language use* (Vol. 94). Cambridge University Press.

Chambers, J. K. (1973). Canadian raising. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *18*(2), 113-135.

Clopper, C. G., Burdin, R. S., & Turnbull, R. (2019). Variation in/u/fronting in the American Midwest. *The Journal of the Acoustical Society of America*, *146*(1), 233-244.

Clopper, C. G., Mitsch, J. F., & Tamati, T. N. (2017). Effects of phonetic reduction and regional dialect on vowel production. *Journal of Phonetics*, *60*, 38-59.

Dailey-O'Cain, J. (1997). Canadian raising in a midwestern US city. *Language variation and Change*, *9*(1), 107-120.

Davis, S., & Berkson, K. (2021). 1. American Raising: An Introduction. *Publication of the American Dialect Society*, *106*(1), 1-12.

Davis, S., Berkson, K., & Strickler, A. (2021). 4. Characteristics of Incipient American Raising and Lifespan Change: Focus on Fort Wayne. *Publication of the American Dialect Society*, *106*(1), 63-94.

Di Paolo, M., & Faber, A. (1990). Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language variation and change*, *2*(2), 155-204.

Farris-Trimble, A., & Tessier, A. M. (2019). The effect of allophonic processes on word recognition: Eye-tracking evidence from Canadian raising. *Language*, *95*(1), e136-e160.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels.

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. The Journal of the Acoustical Society of America, 126(5), 2603-2618.

Fruehwald, J. (2016). The early influence of phonology on a phonetic change. *Language*, *92*(2), 376-410.

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. Origins of sound change: *Approaches to phonologization*, 1, 51-97.. (2022). Acoustic differentiation of

allophones of/aɪ/ in Chicagoland English: Statistical comparison of formant trajectories. *Journal of the International Phonetic Association*, *52*(2), 197-227.

Hualde, J. I., Luchkina, T., & Eager, C. D. (2017). Canadian Raising in Chicagoland: The production and perception of a marginal contrast. *Journal of Phonetics*, *65*, 15-44.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 505-528.

Joos, M. (1942). A phonological dilemma in Canadian English. *Language*, 141-144.

Kisler, T. and Reichel U. D. and Schiel, F. (2017): Multilingual processing of speech via web services, Computer Speech & Language, Volume 45, September 2017, pages 326–347.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. Mouton de Gruyter.

Lindblom, B., Guion, S., Hura, S., Moon, S. J., & Willerman, R. (1995). Is sound change adaptive?. *Rivista di linguistica*, *7*, 5-36.

Moreton, E. (2004). Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics*, *32*(1), 1-33.

Moreton, E., & Thomas, E. R. (2007). Origins of Canadian Raising in voiceless-coda effects: a case study in phonologization. *Laboratory phonology*, *9*, 37-64.

Ohala, J. (1981). The listener as a source of sound change. *Papers from the Parasession on Language and Behavior*.

Ohala, J. (2012). The listener as a source of sound change. *The Initiation of Sound Change: Perception, Production, and Social Factors*, *323*.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. The Journal of the Acoustical Society of America, 32(6), 693-703.

Renwick, M. E., & Stanley, J. A. (2020). Modeling dynamic trajectories of front vowels in the American South. *The Journal of the Acoustical Society of America*, *147*(1), 579-595.

Smiljanic, R., & Bradlow, A. R. (2007, August). Clear speech intelligibility: Listener and talker effects. In *Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrucken, Germany*.

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017.

Strickler, A. (2019). Within-speaker perception and production of dialectal/aɪ/-raising. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia* (pp. 3205-3209).

Thomas, E. R., & Mielke, J. (2021). 3. The Phonetic Development of American Raising in Eastern Ohio. *Publication of the American Dialect Society*, *106*(1), 45-62.

Traunmüller, Hartmut. "Auditory scales of frequency representation." The Journal of the *Acoustical Society of America 88, no. 97 (1990).*

Uchanski, R. M. (2005). Clear speech. *The handbook of speech perception*, 207-235.

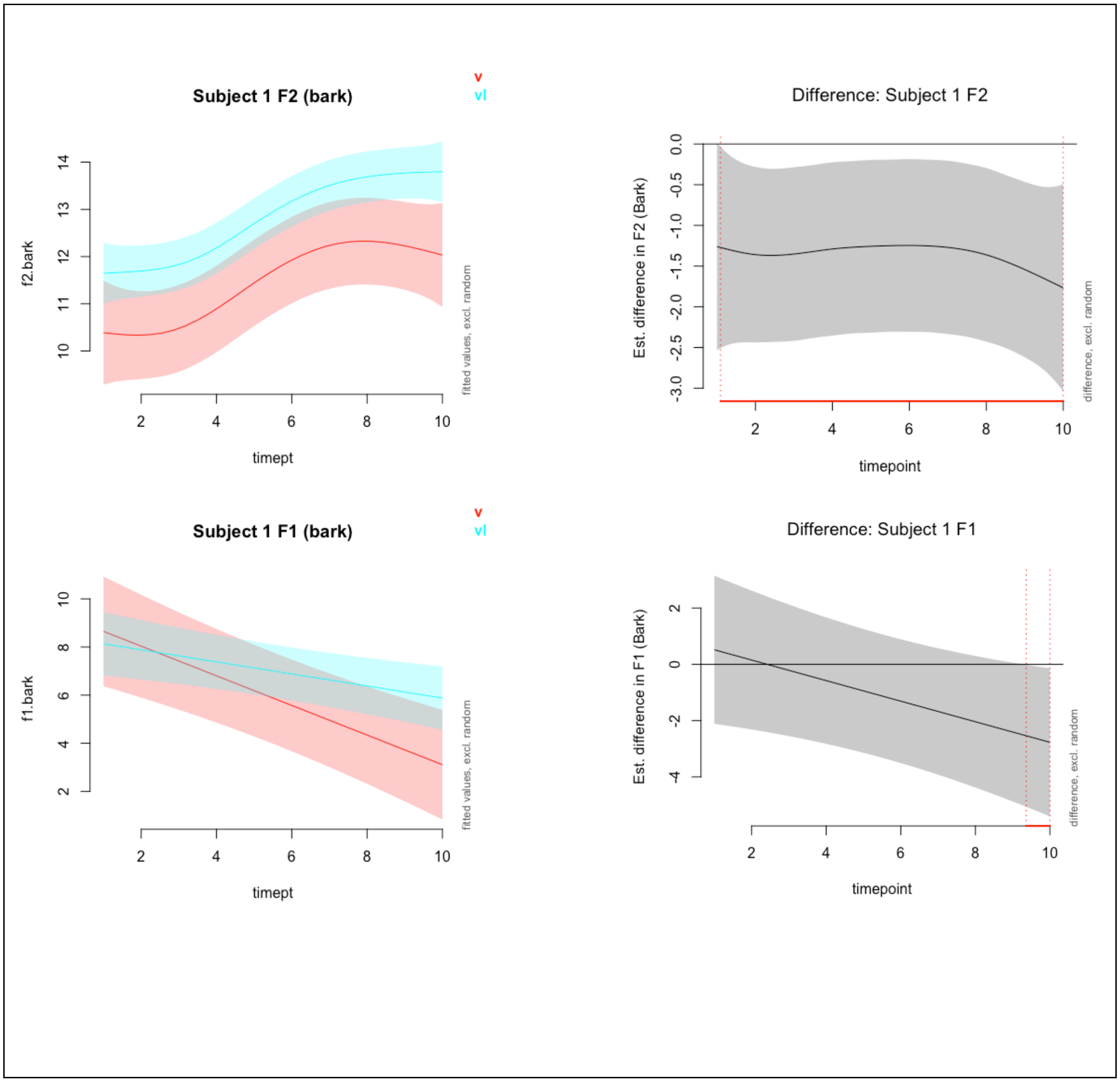
Vance, T. J. (1987). " Canadian Raising" in some dialects of the northern United States. *American Speech*, *62*(3), 195-210.

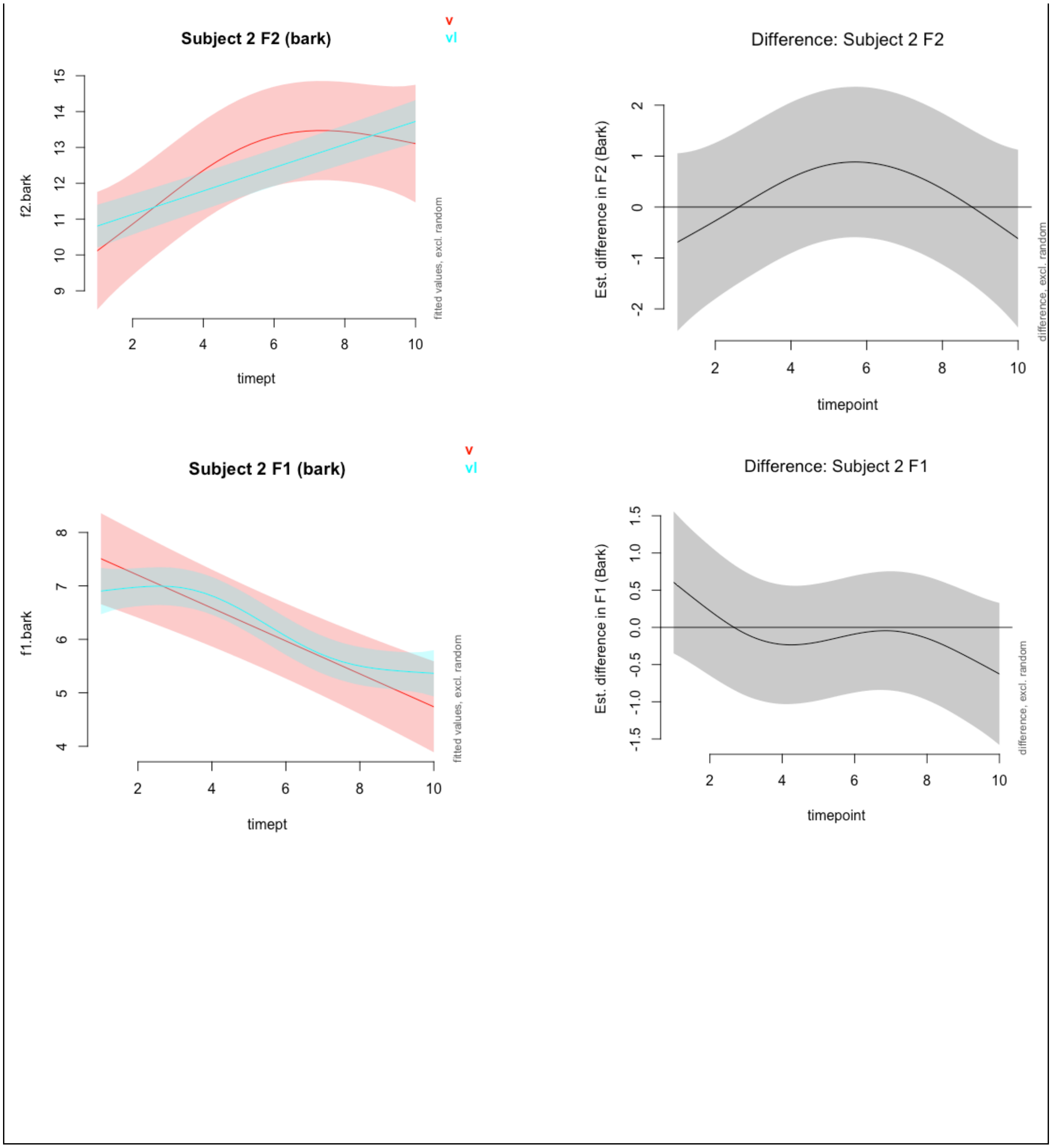
Whalen, D. H., Magen, H. S., Pouplier, M., Min Kang, A., & Iskarous, K. (2004). Vowel production and perception: Hyperarticulation without a hyperspace effect. *Language and speech*, *47*(2), 155-174.

Yu, Alan C. L. 2019. "On the nature of the perception-production link: individual variability in English sibilant-vowel coarticulation." *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1), 2.

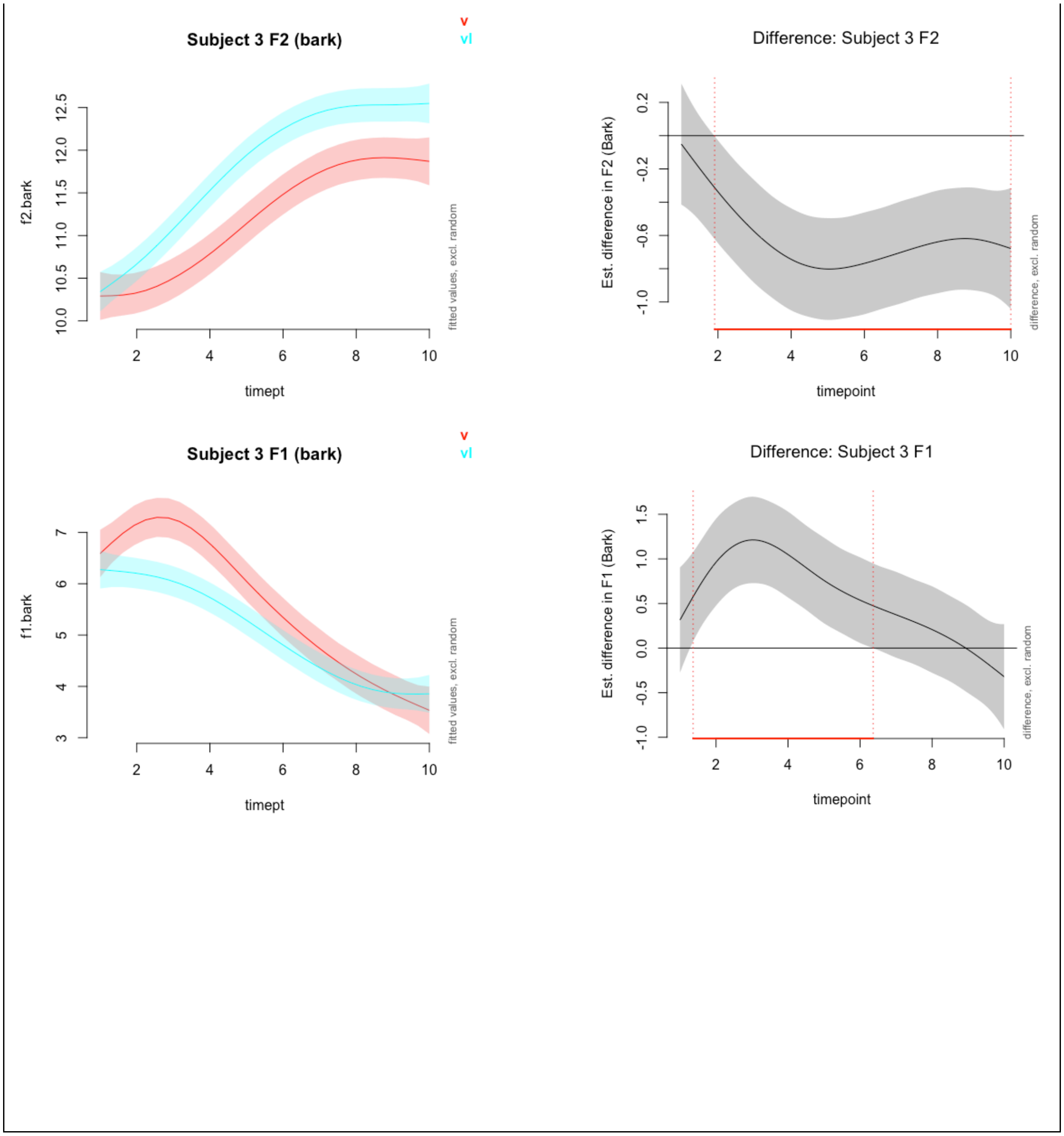
Zellou, G., & Scarborough, R. (2019). Neighborhood-conditioned phonetic enhancement of an allophonic vowel split. *The Journal of the Acoustical Society of America*, *145*(6), 3675-3685.

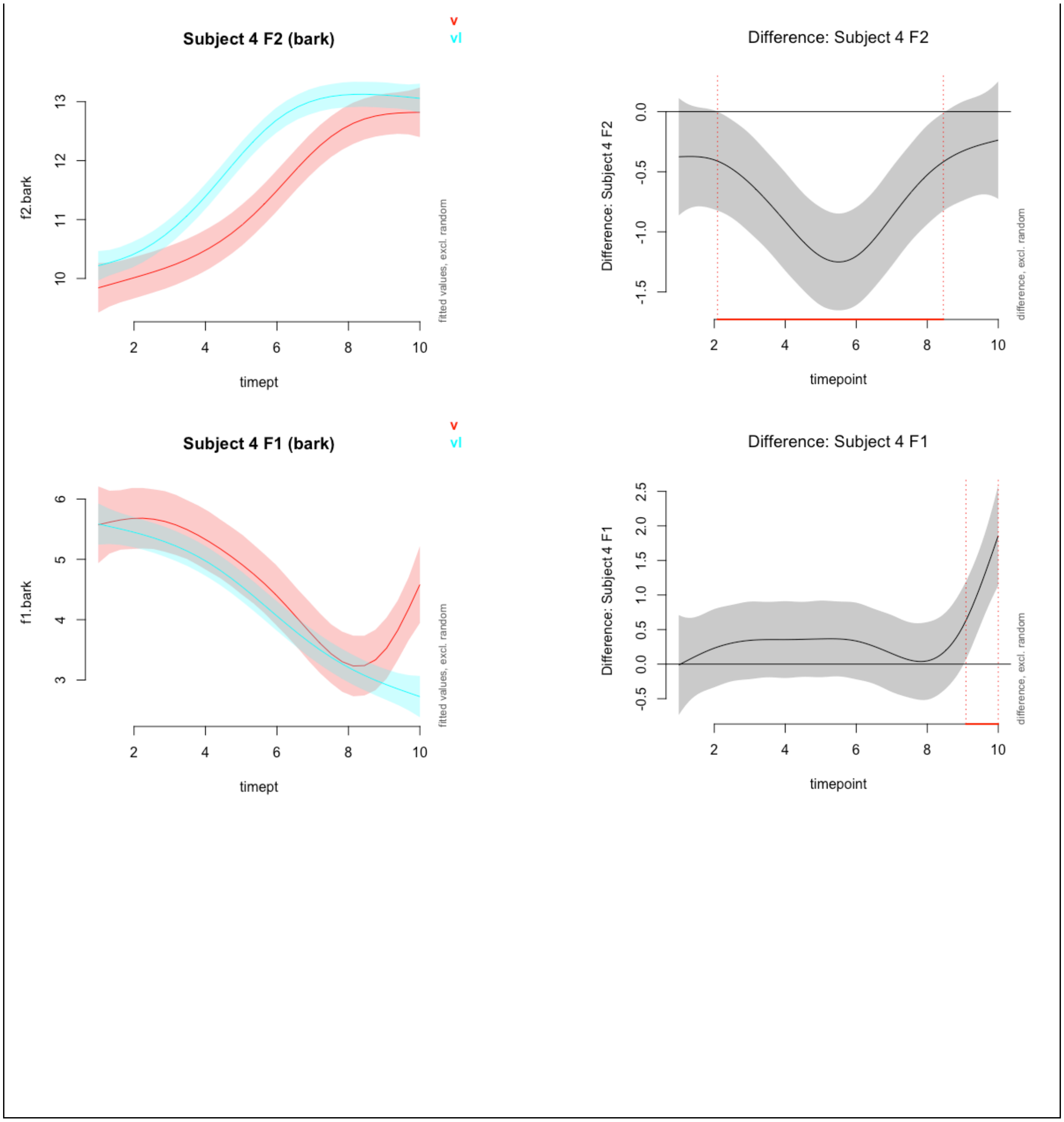
Appendix 1: GAMMs and Difference plots for monosyllabic words, by speaker

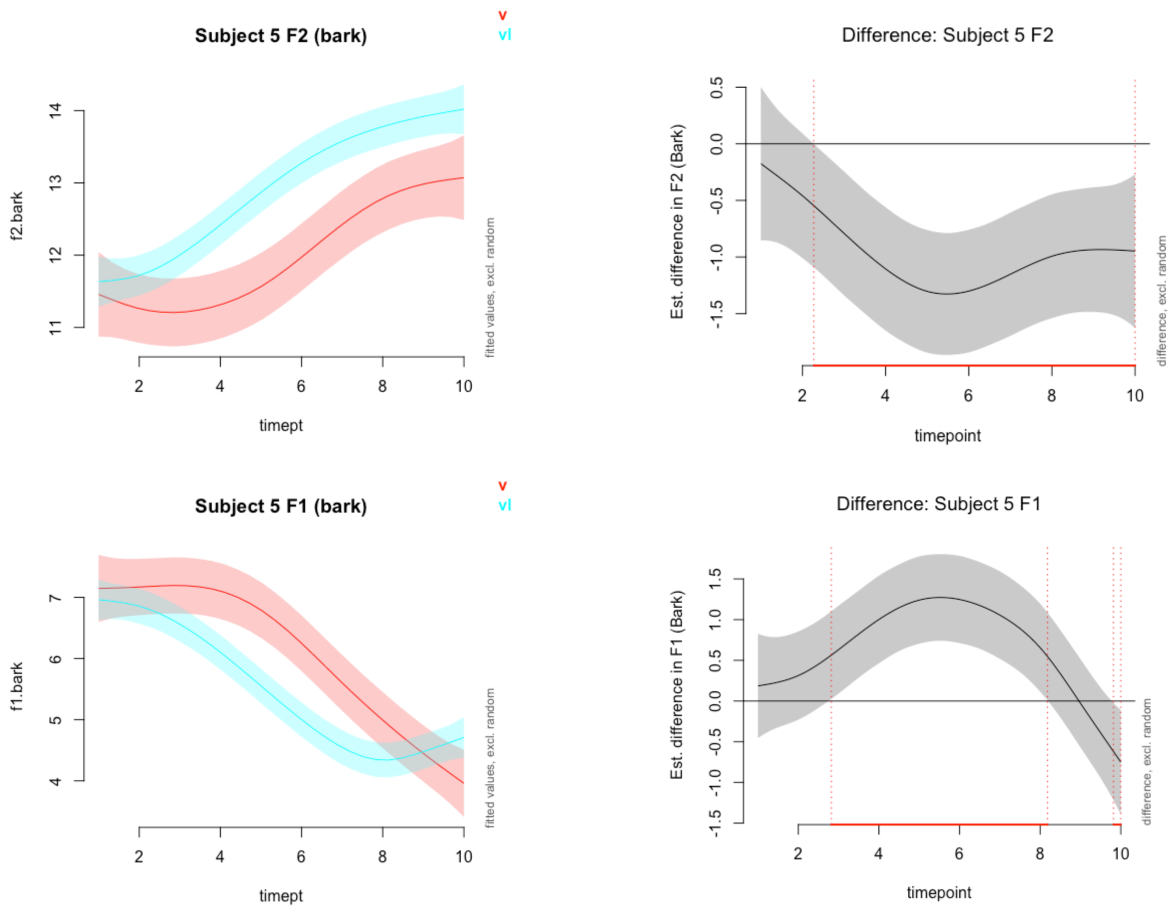


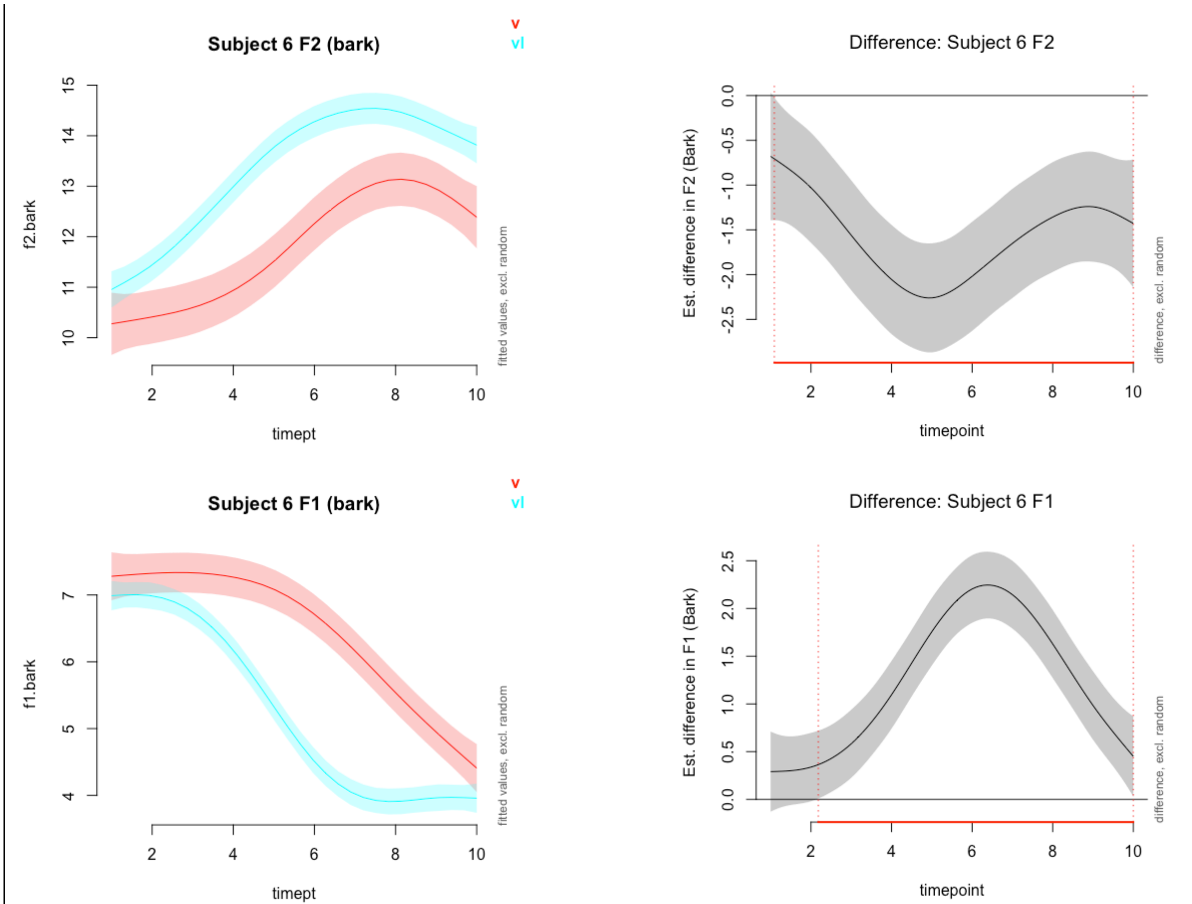


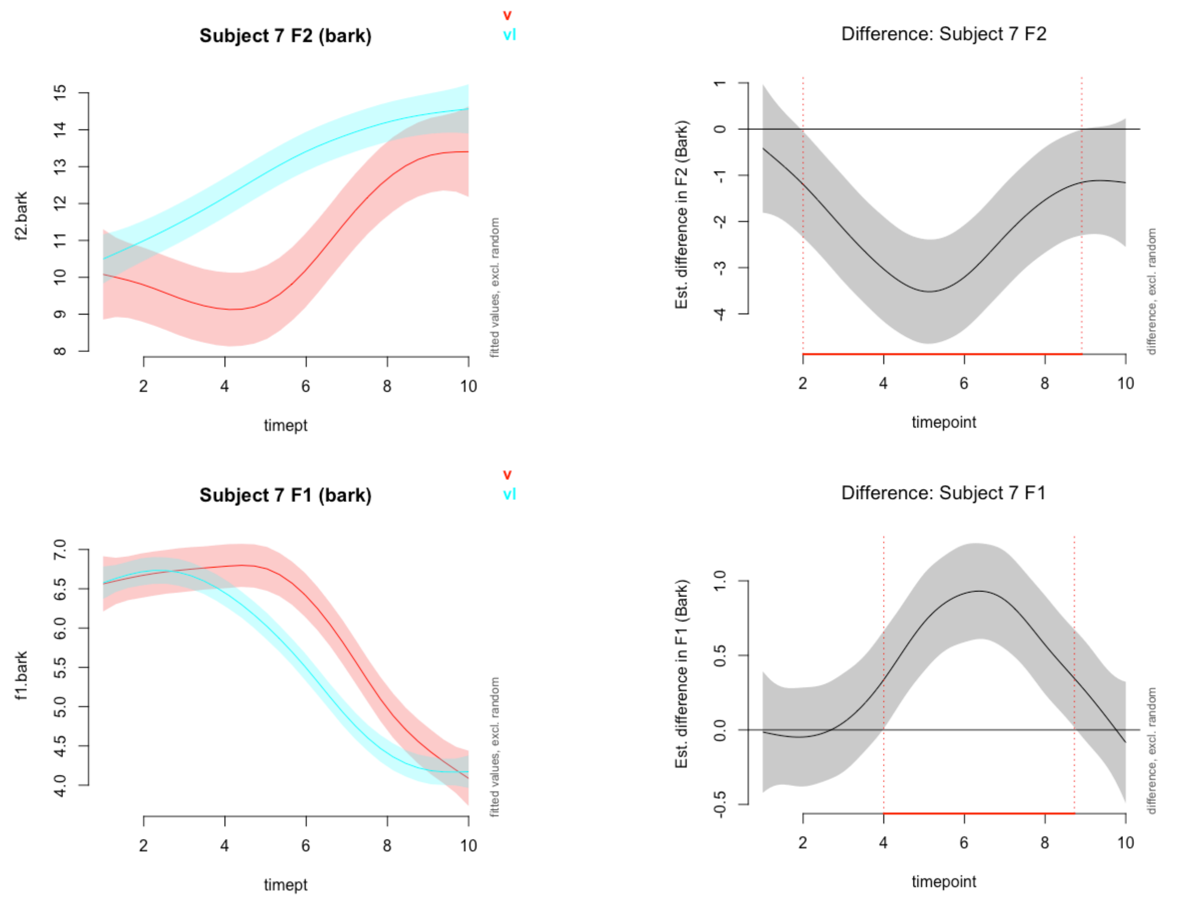
Subject 2 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 2 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 2 F1 (bark)
f1.bark
Difference: Subject 2 F1
Est. difference in F1 (Bark)

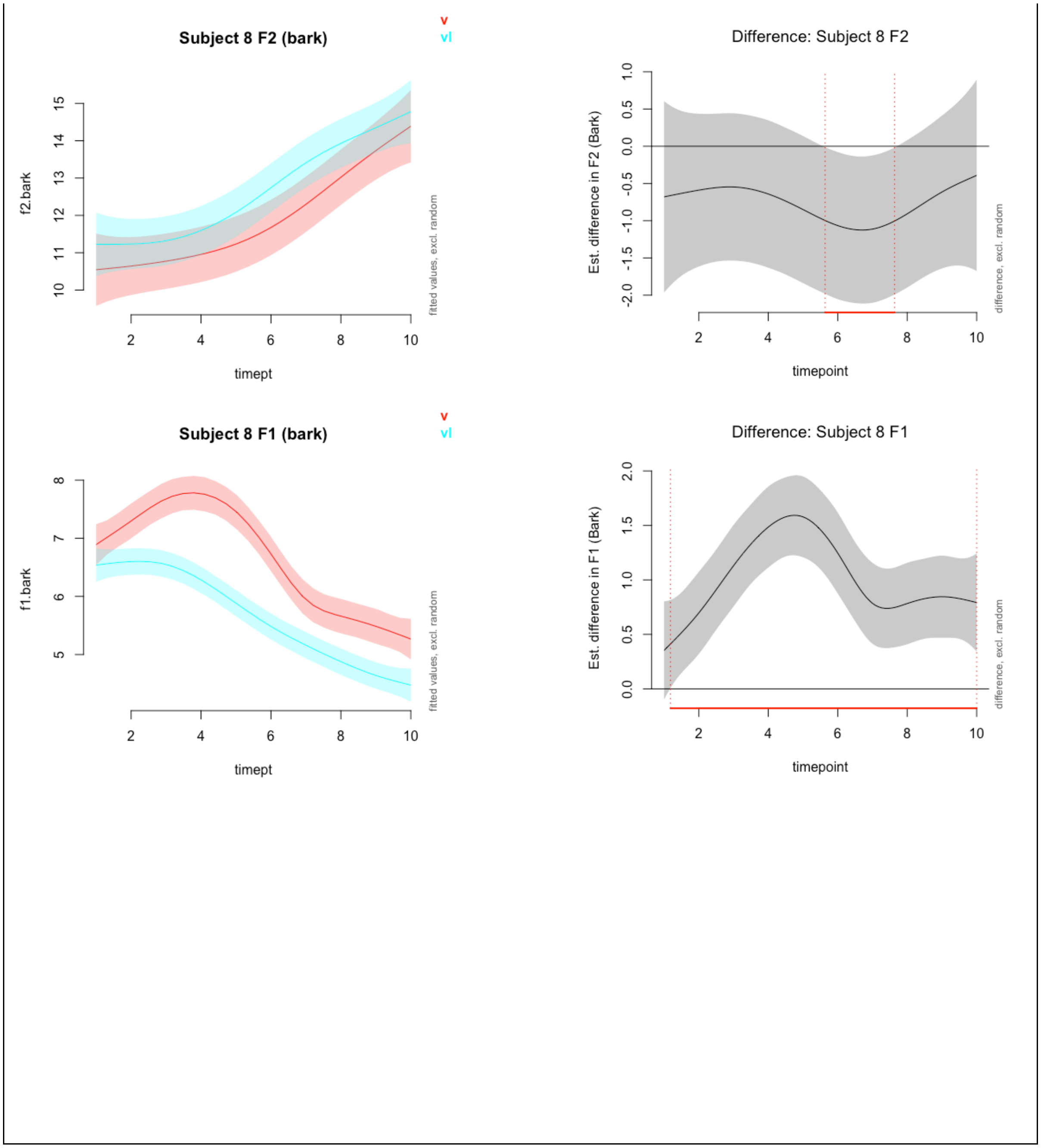

Subject 3 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 3 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 3 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 3 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random


Subject 4 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 4 F2
Difference: Subject 4 F2
timepoint
difference, excl. random
Subject 4 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 4 F1
Difference: Subject 4 F1
timepoint
difference, excl. random

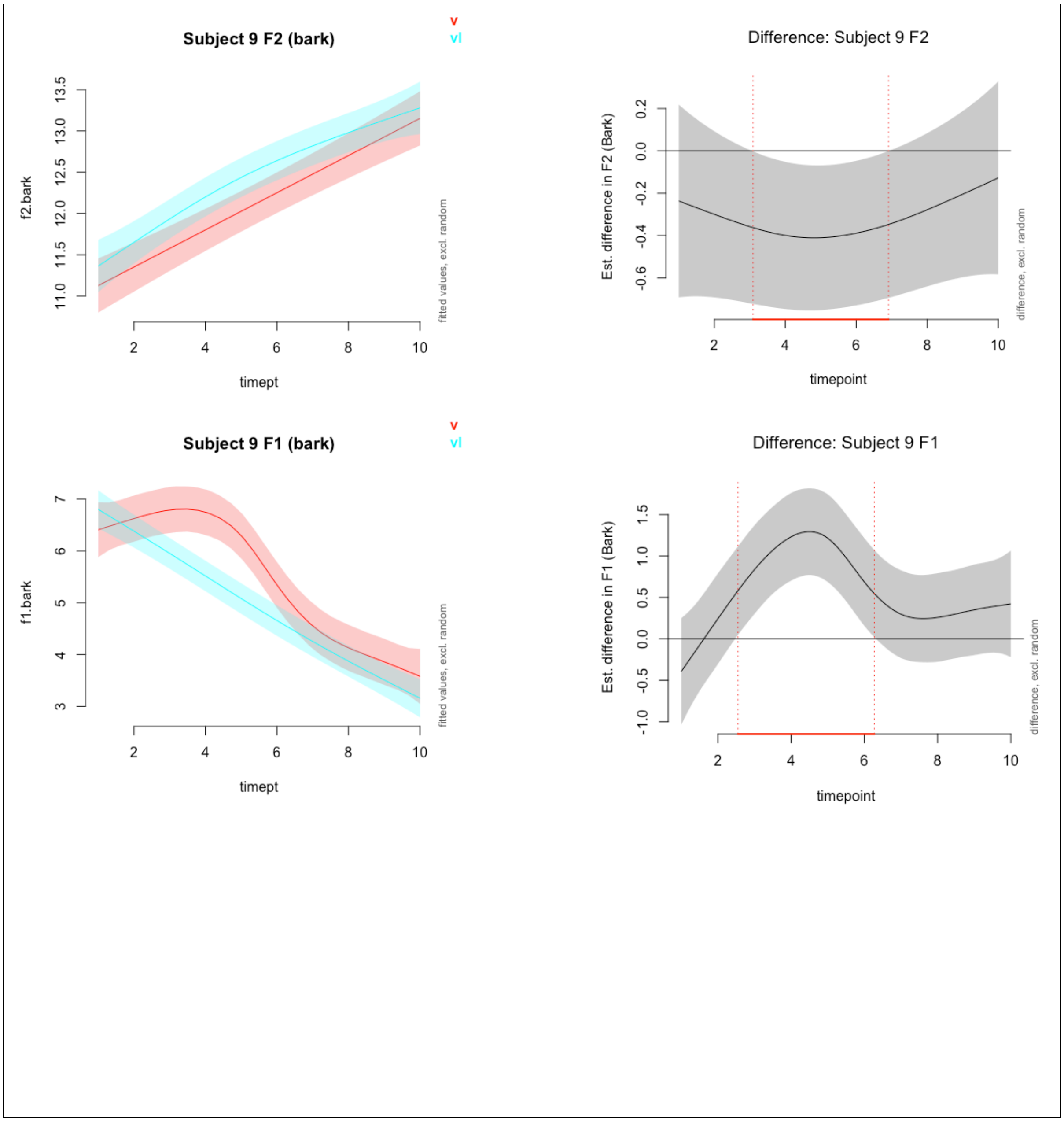

Subject 5 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 5 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 5 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 5 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

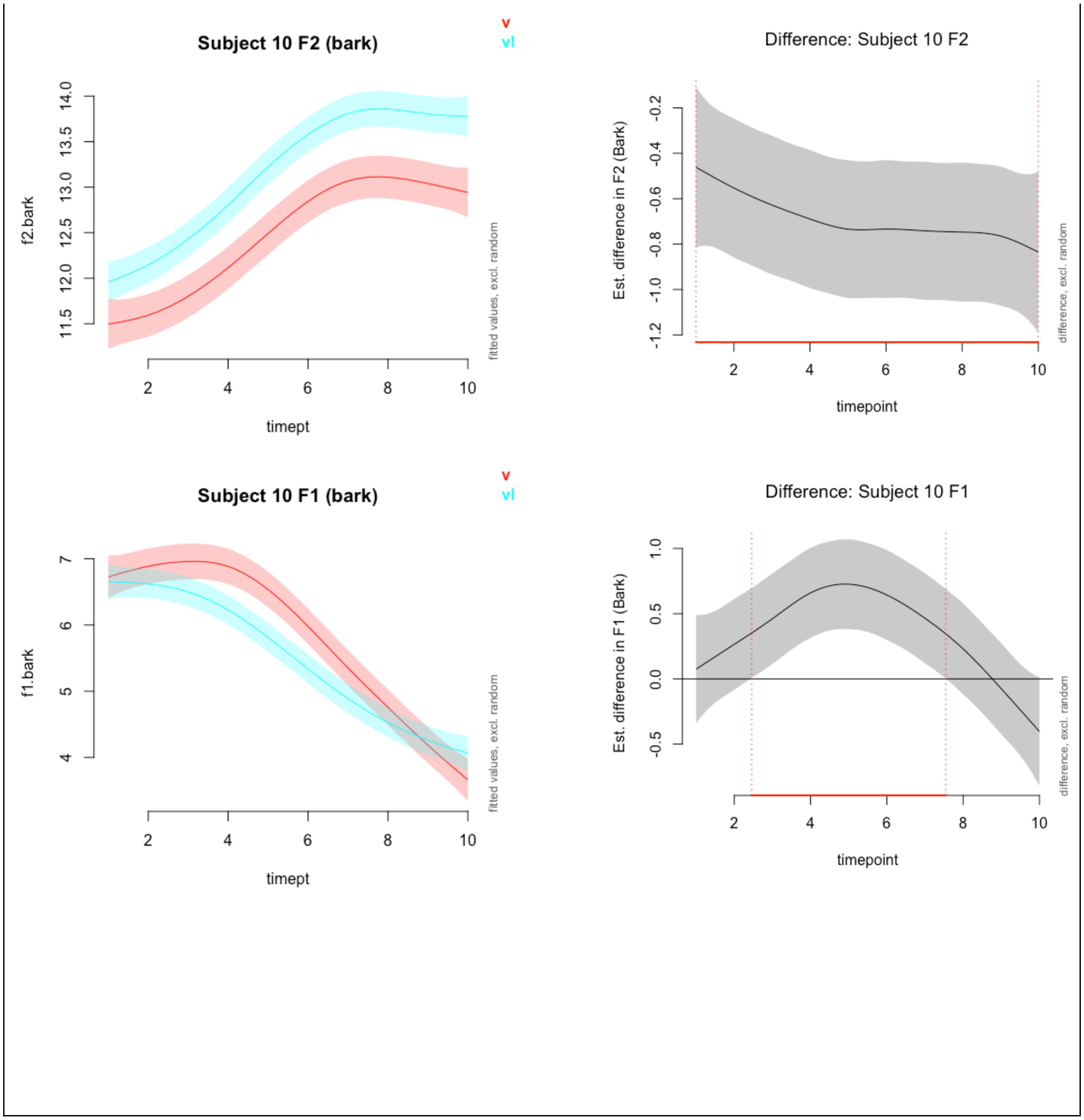

Subject 6 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 6 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 6 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 6 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

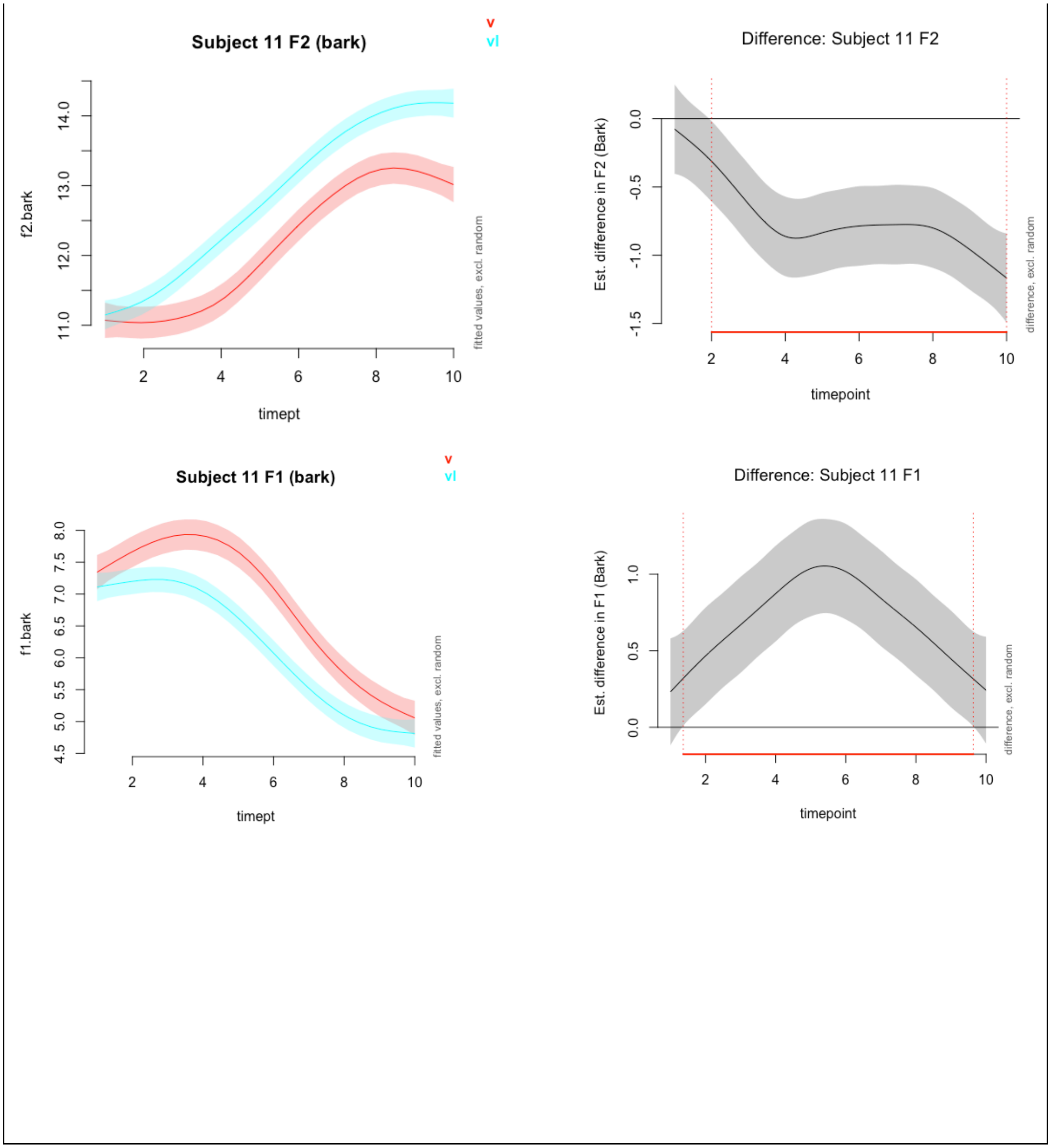

Subject 7 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 7 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 7 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 7 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random



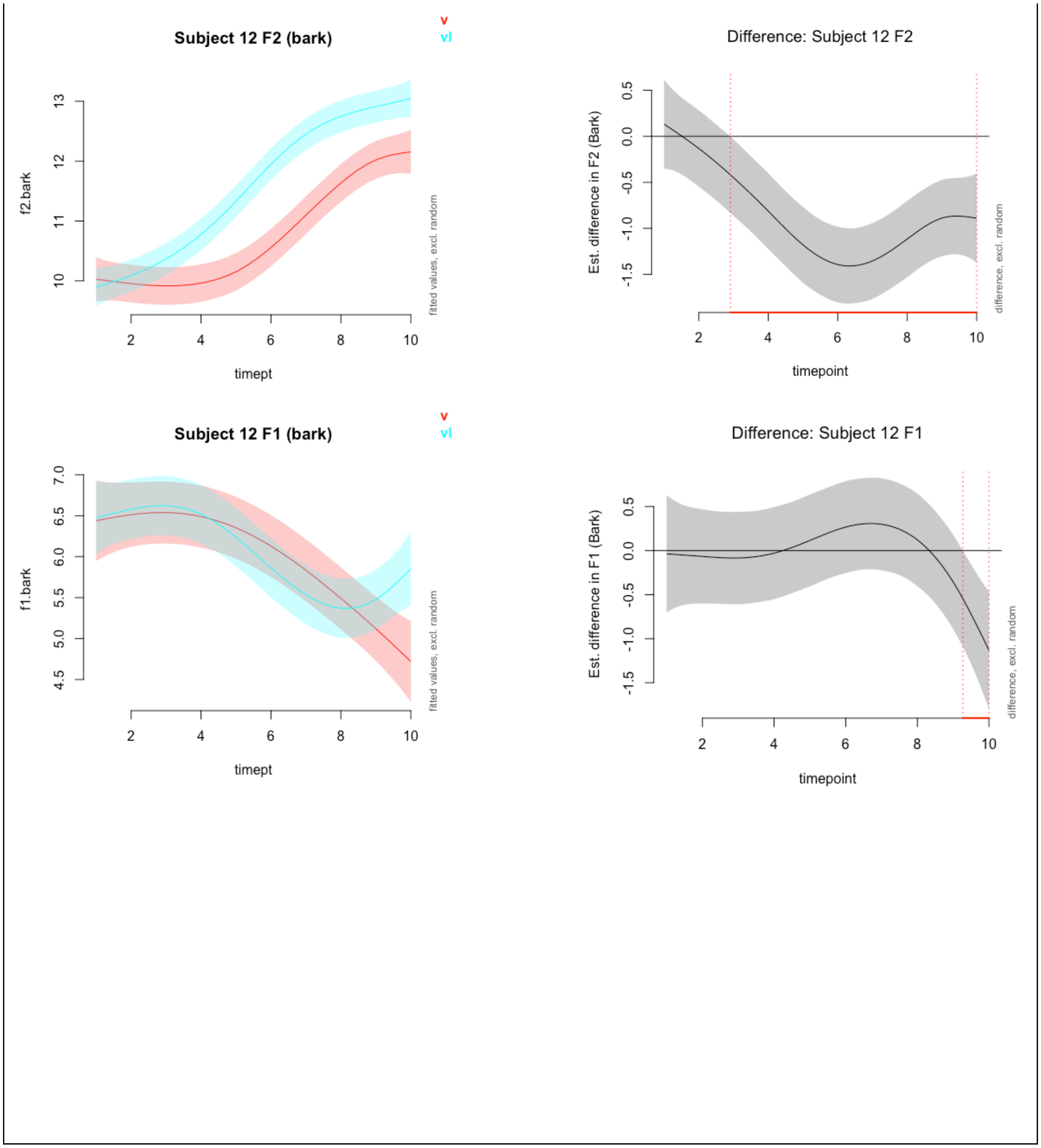
Subject 8 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 8 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 8 F1 (bark)
f1.bark
Difference: Subject 8 F1
Est. difference in F1 (Bark)

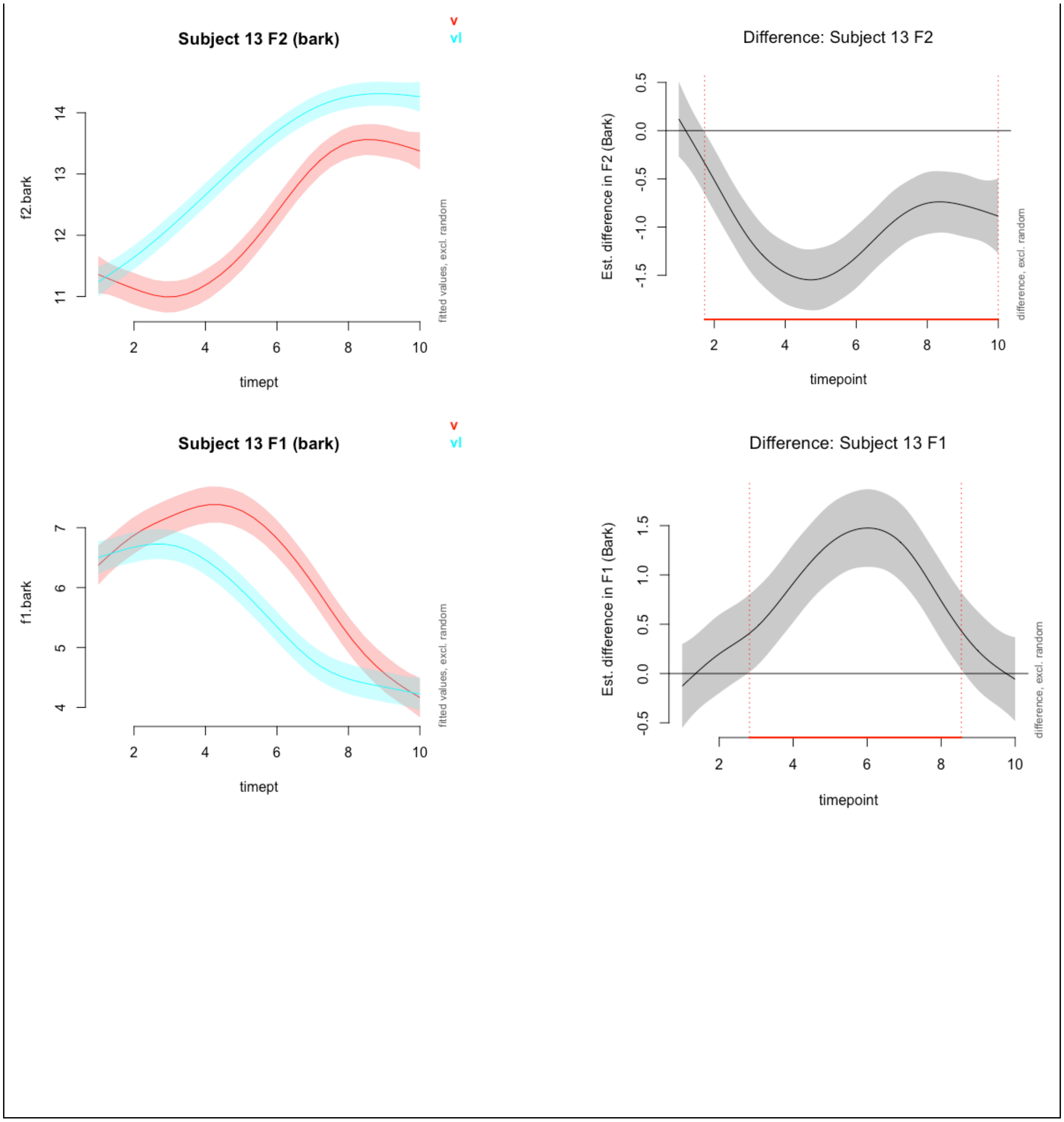

Subject 9 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 9 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 9 F1 (bark)
f1.bark
Difference: Subject 9 F1
Est. difference in F1 (Bark)

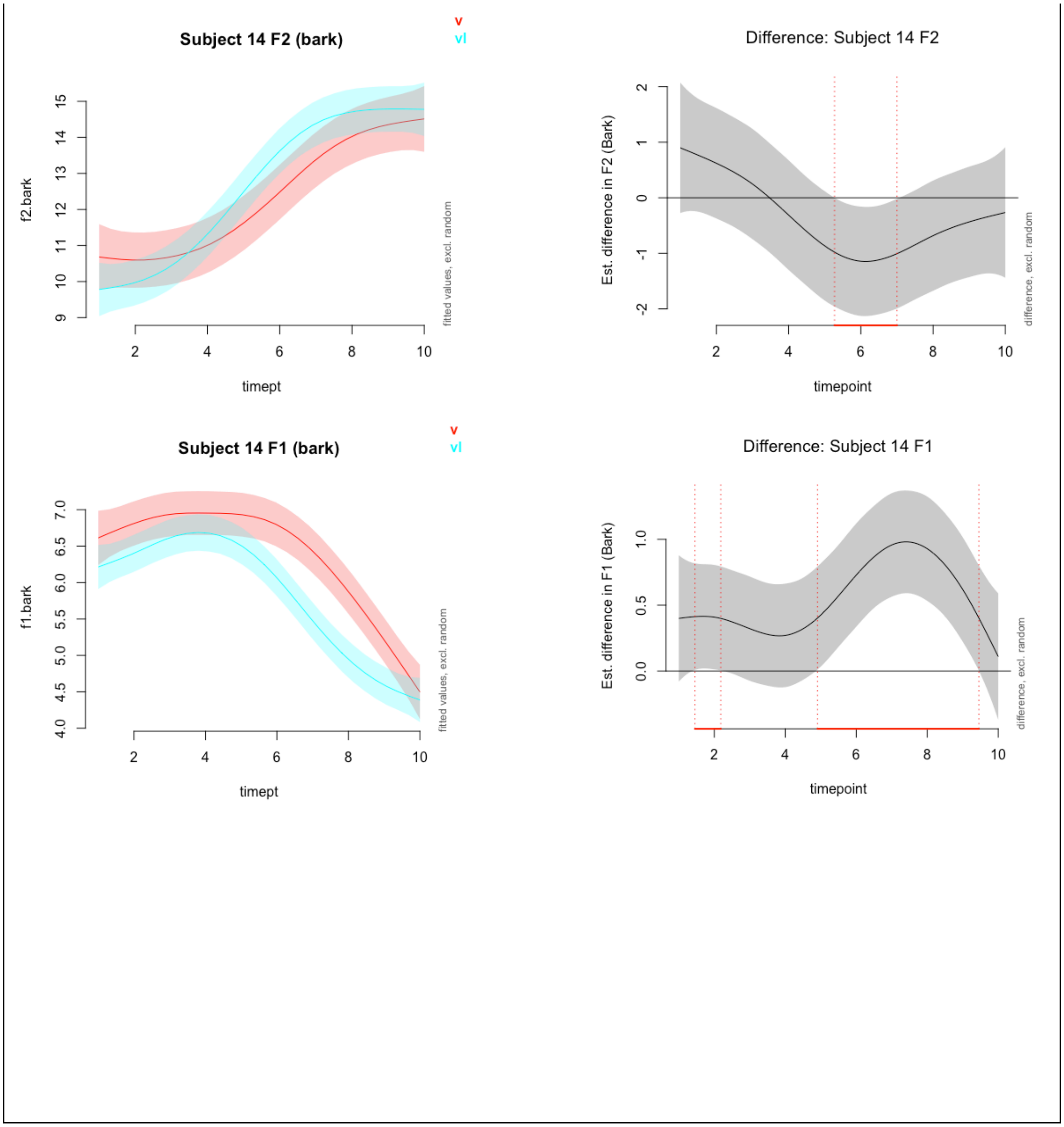

Subject 10 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 10 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 10 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 10 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

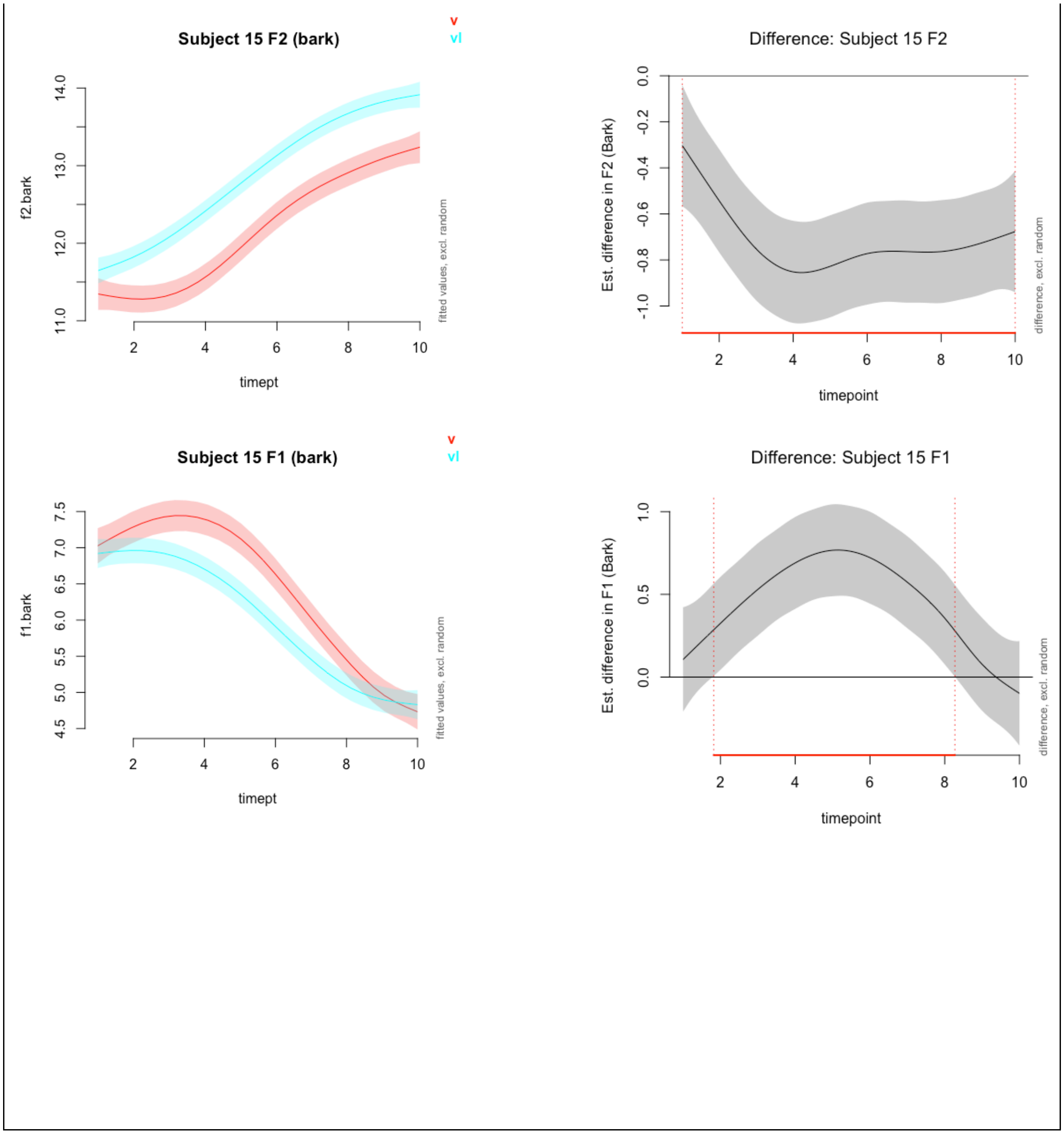


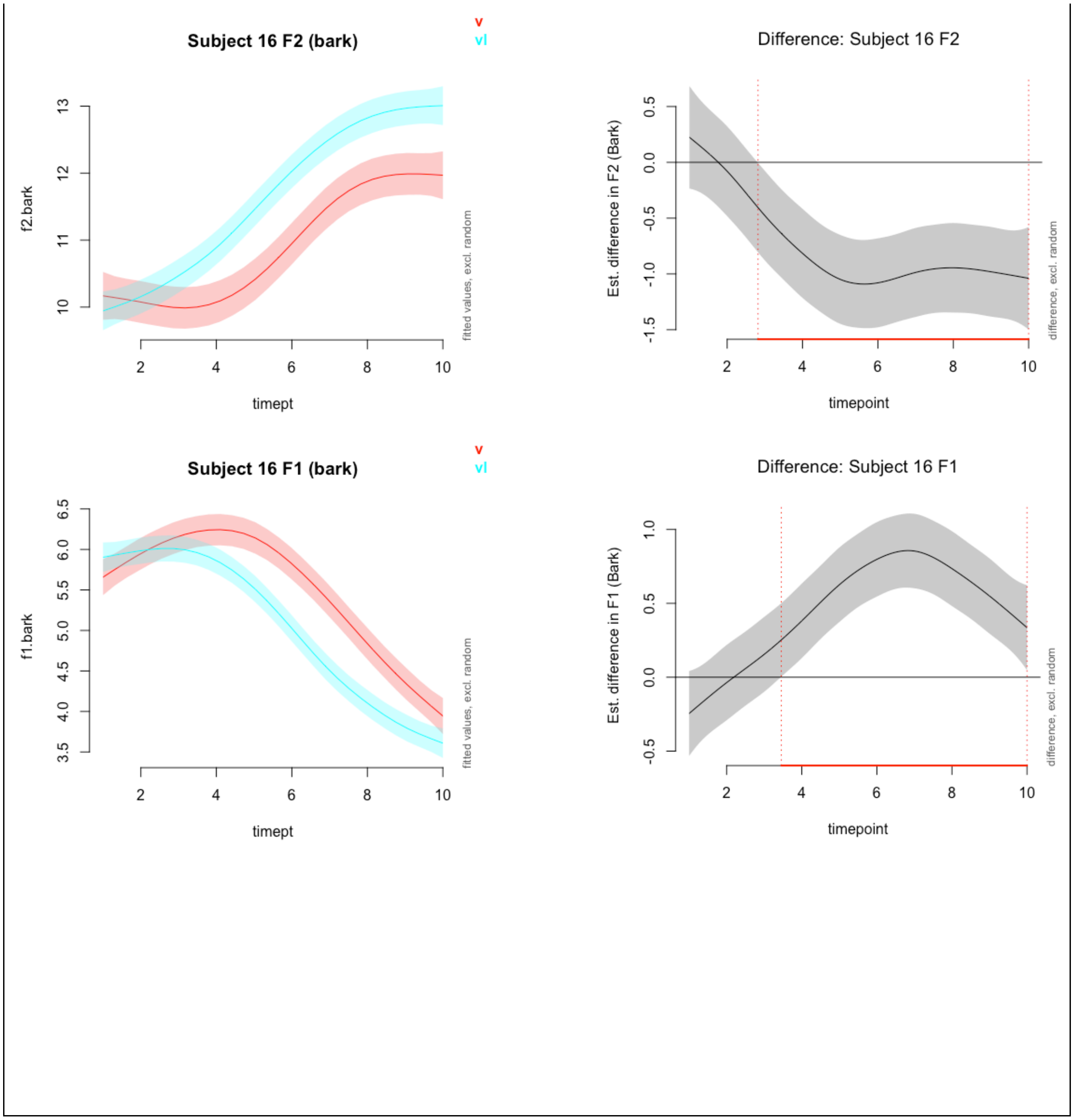
Subject 11 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 11 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 11 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 11 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random


Subject 12 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 12 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 12 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 12 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

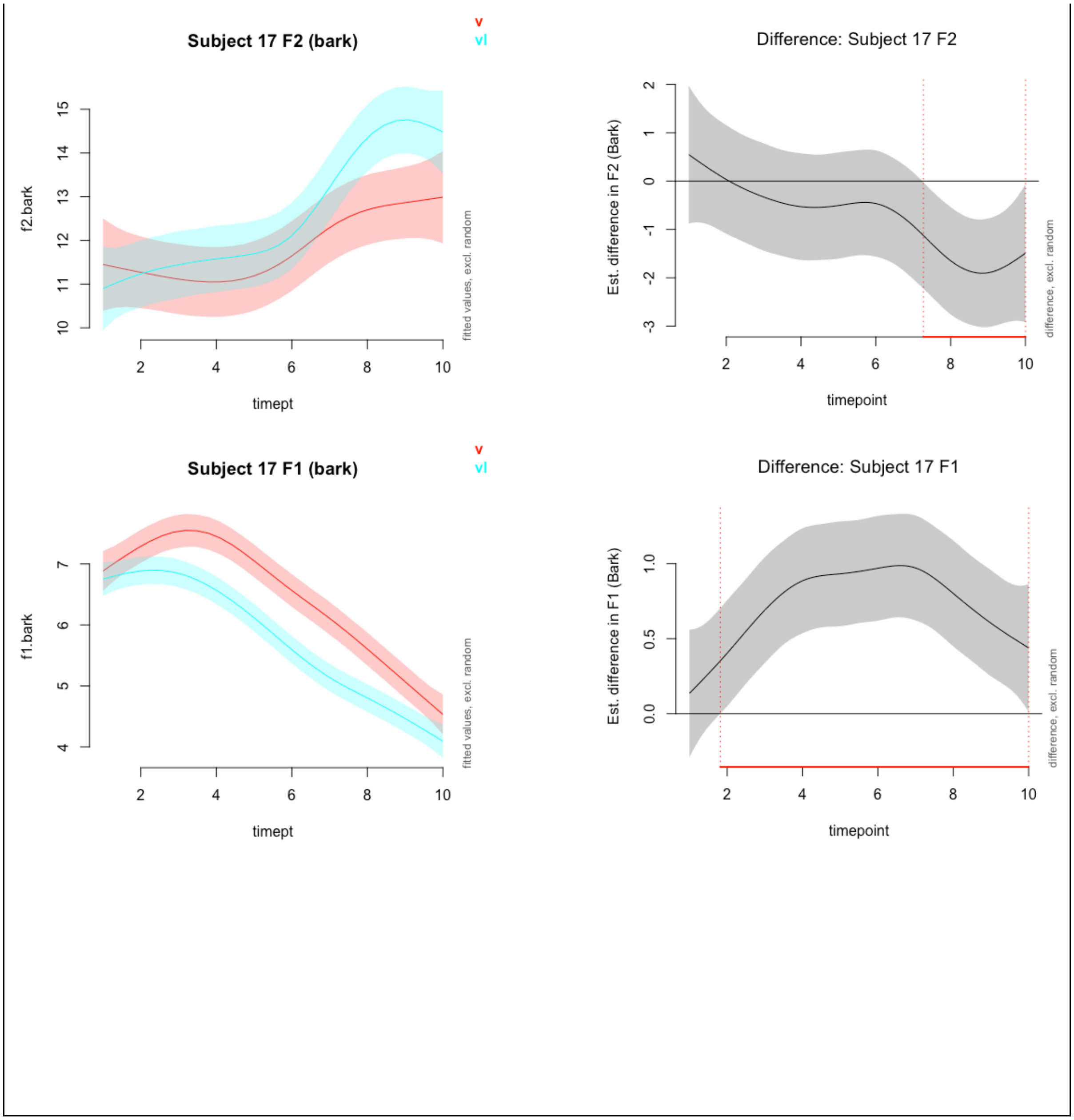

Subject 13 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 13 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 13 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 13 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

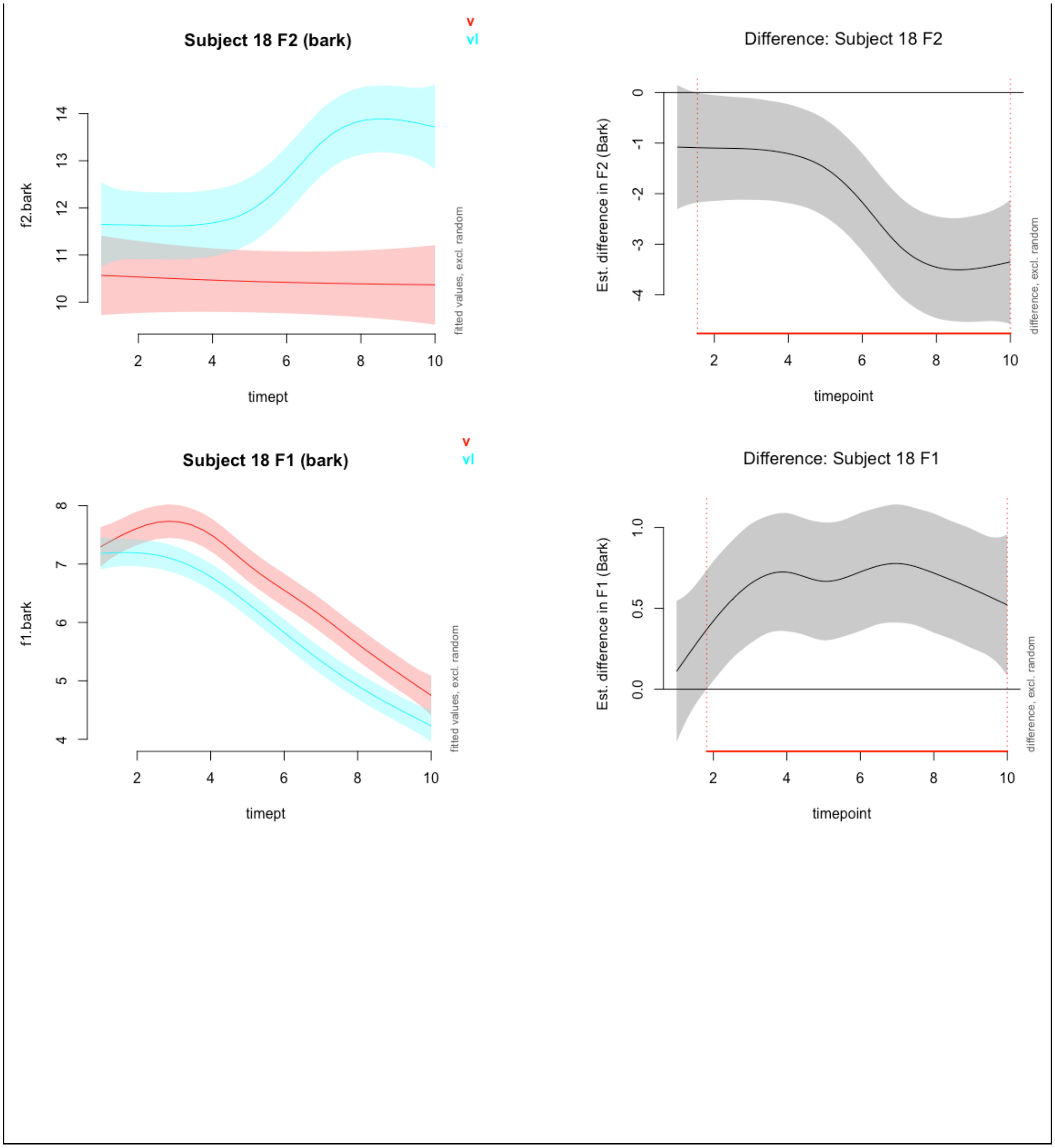

Subject 14 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 14 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 14 F1 (bark)
f1.bark
Difference: Subject 14 F1
Est. difference in F1 (Bark)

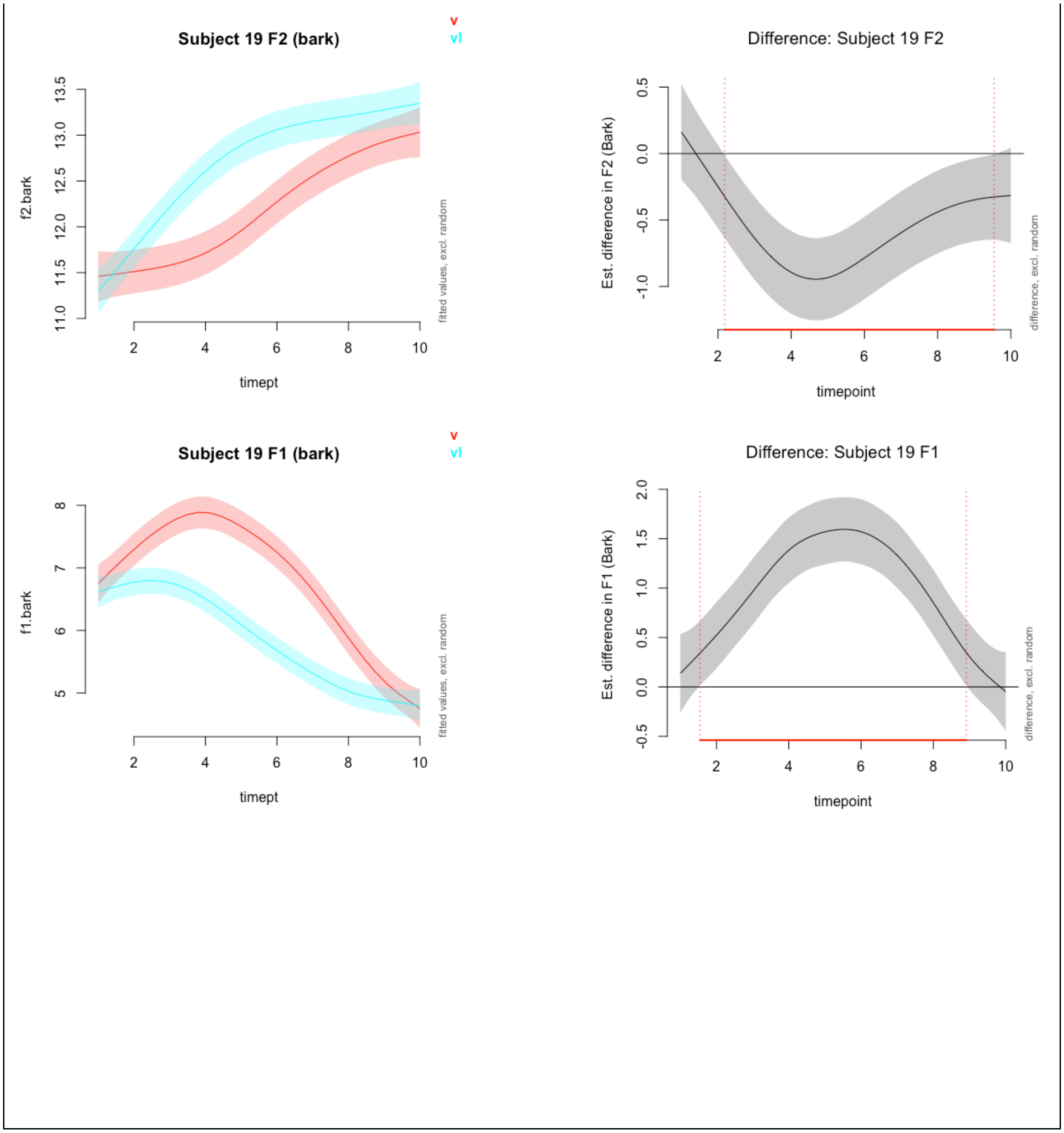

Subject 15 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 15 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 15 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 15 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

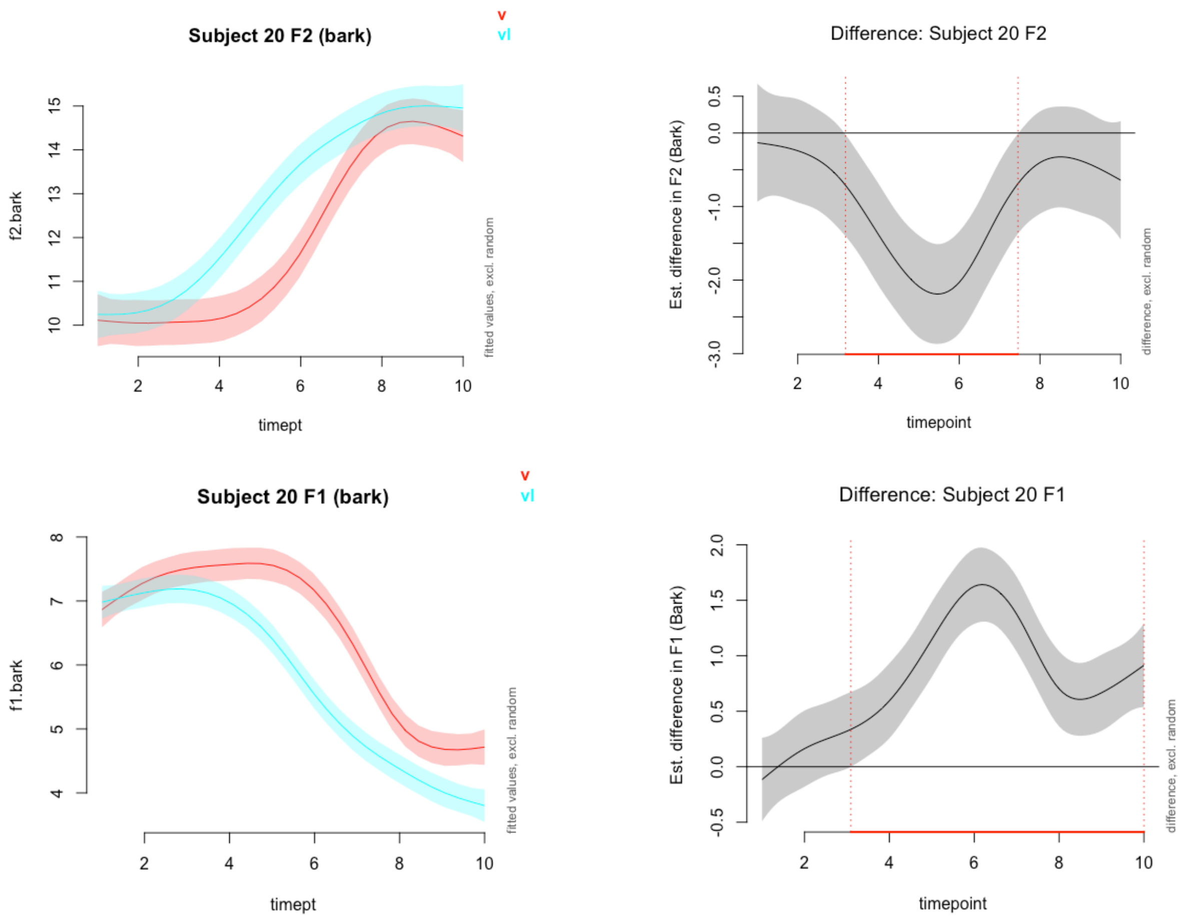


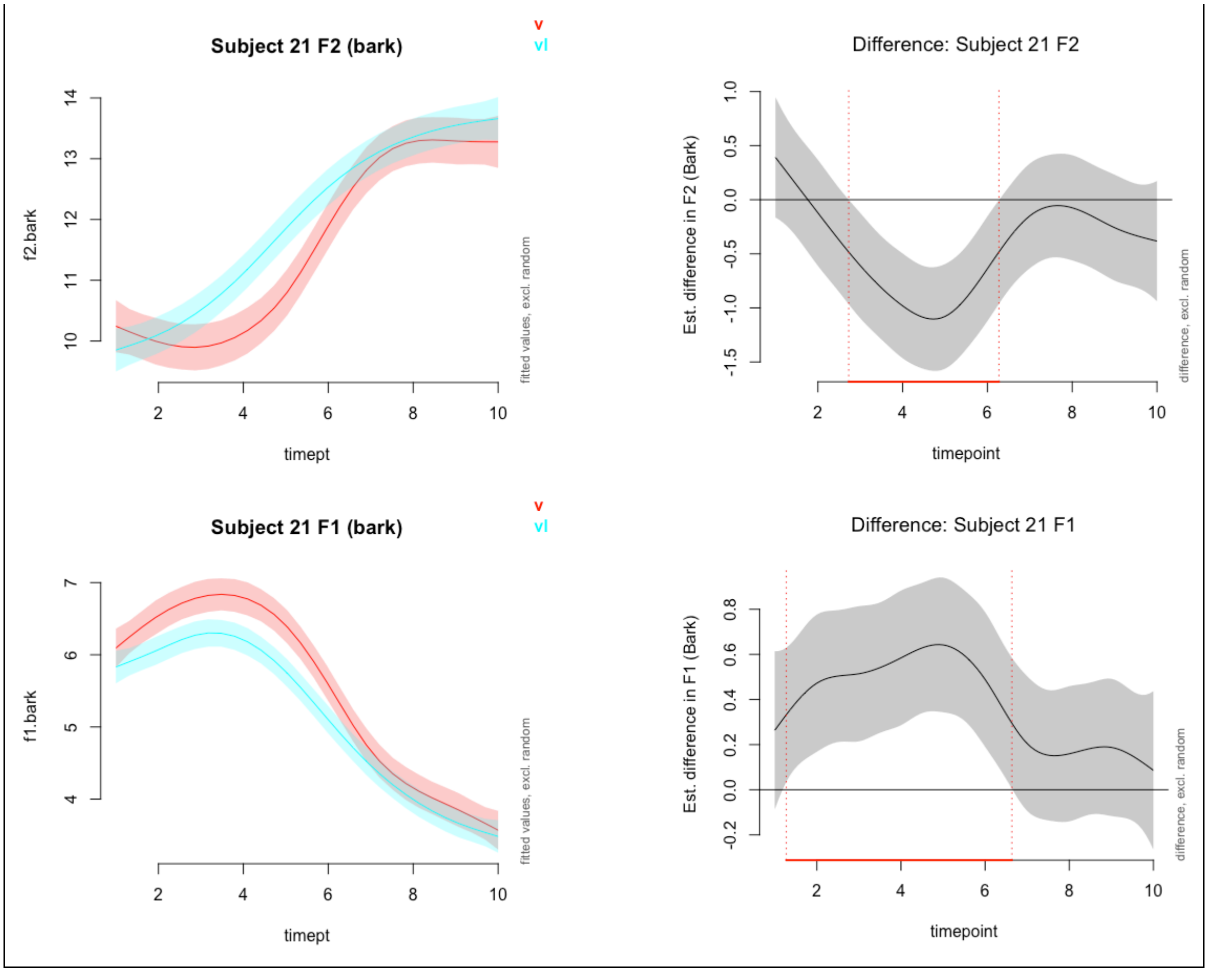
Subject 16 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 16 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 16 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 16 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random


Subject 17 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 17 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 17 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 17 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

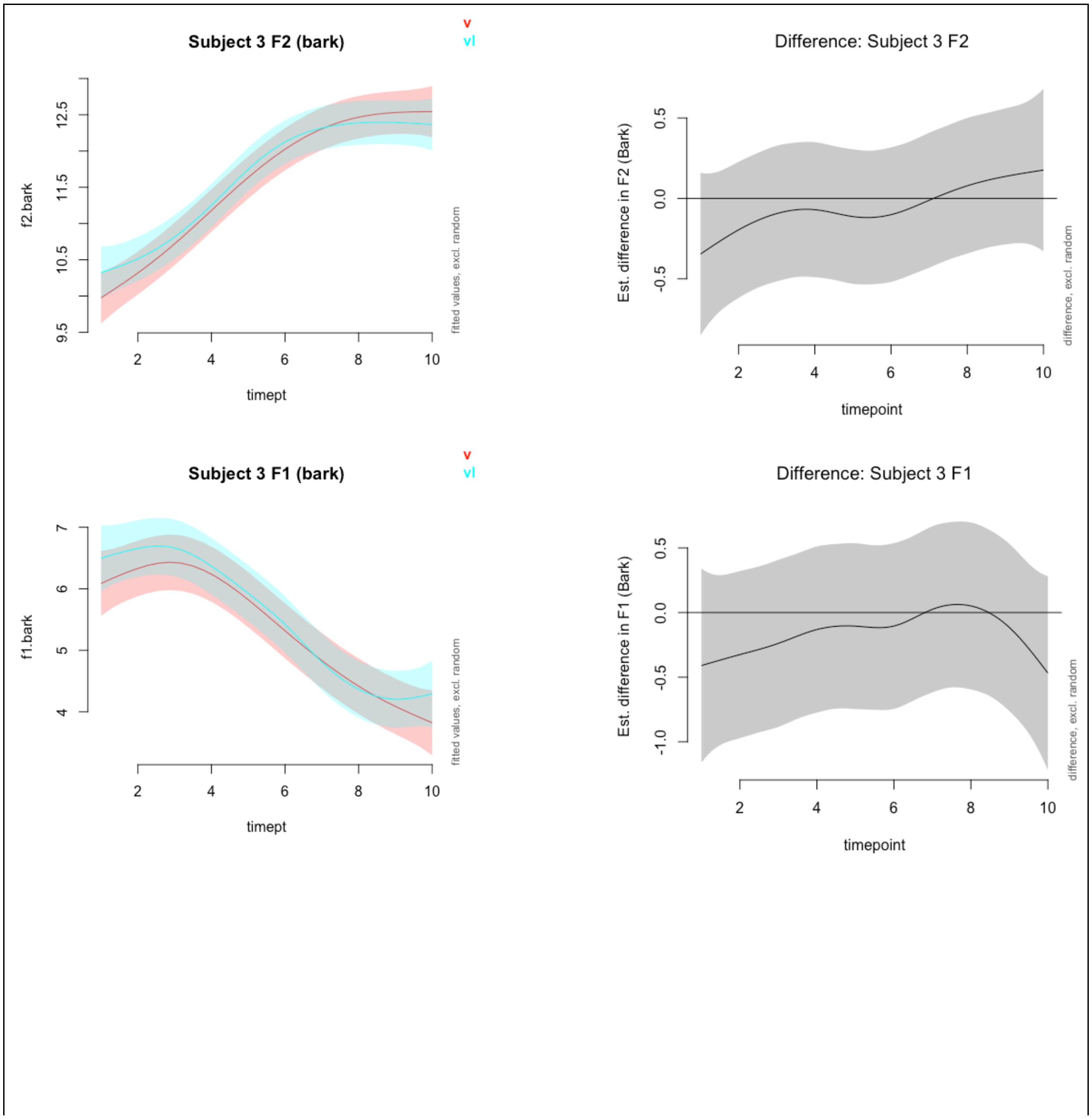

Subject 18 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 18 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 18 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 18 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

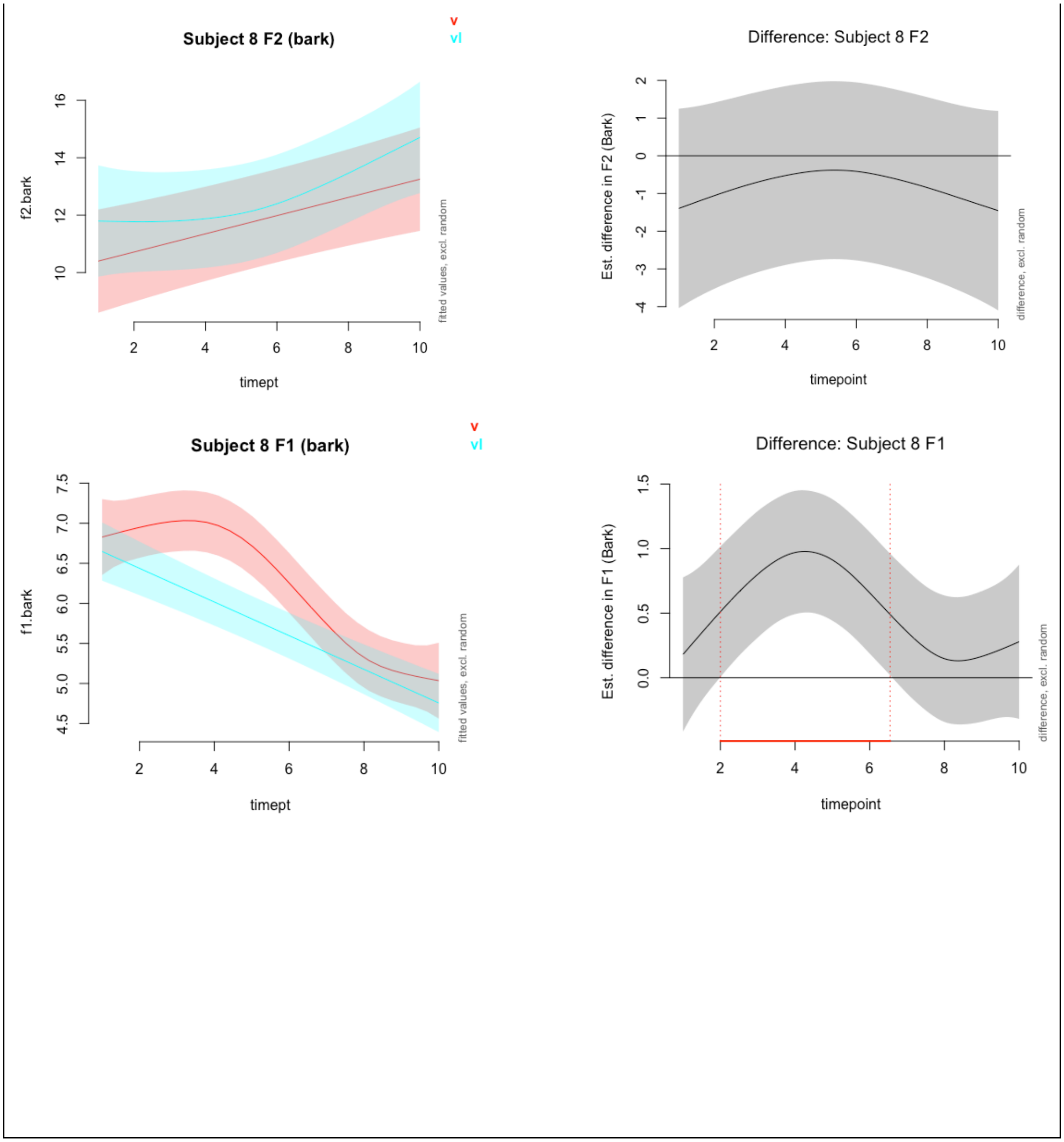

Subject 19 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 19 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 19 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 19 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random


Subject 20 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 20 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 20 F1 (bark)
f1.bark
Difference: Subject 20 F1
Est. difference in F1 (Bark)

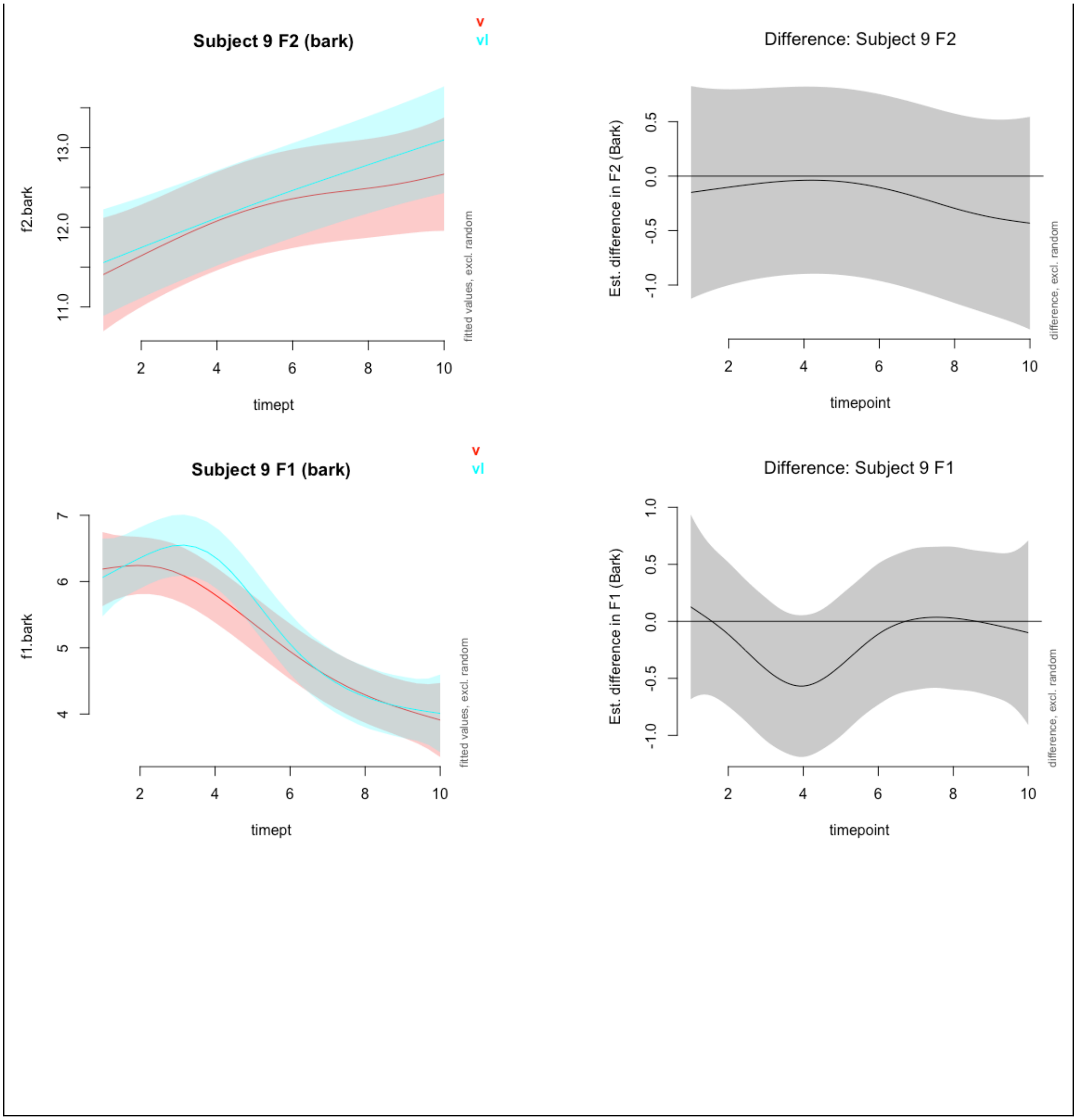

Subject 21 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 21 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 21 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 21 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

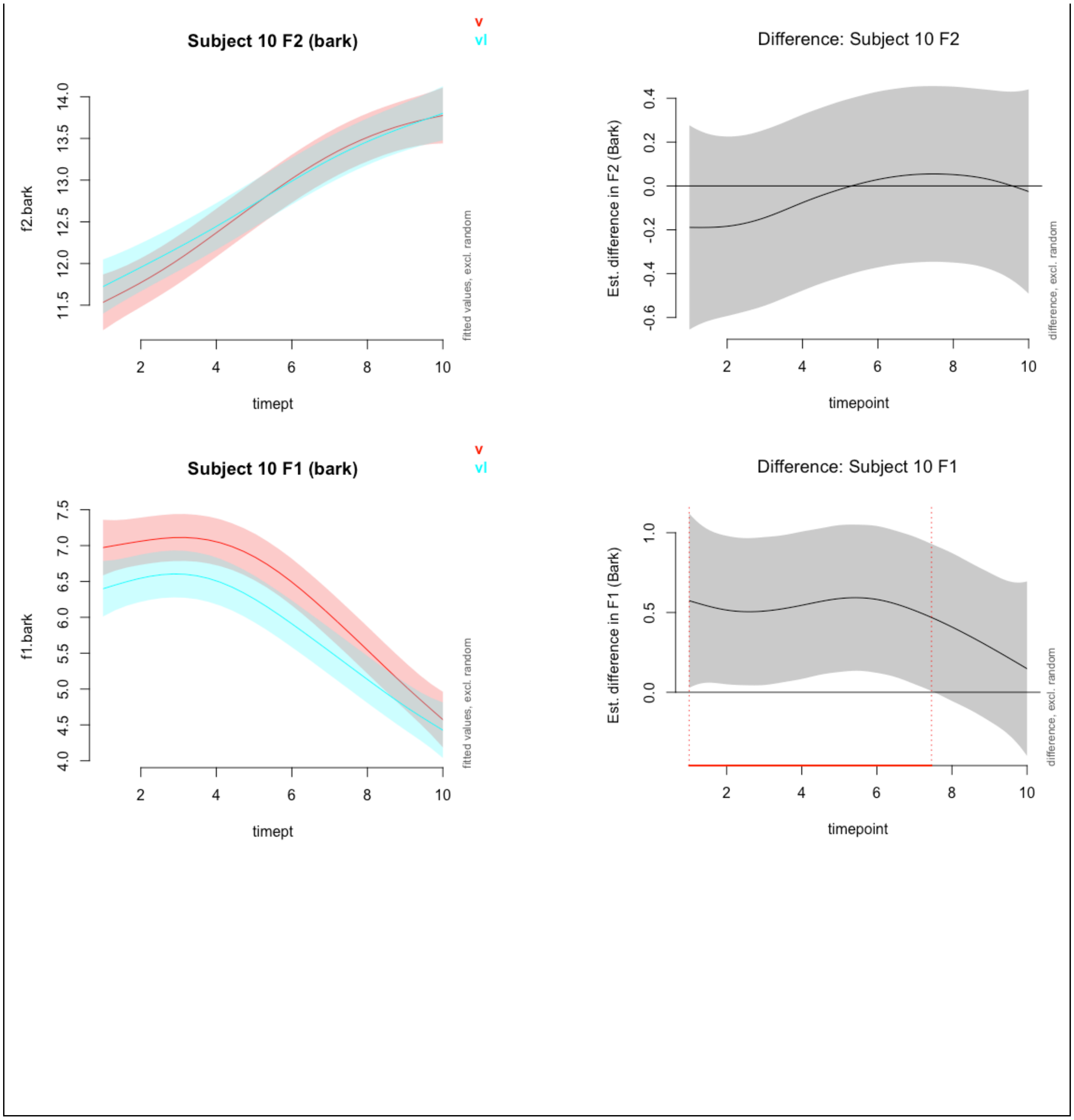
Appendix 2: GAMM and difference plots for pre-flap words

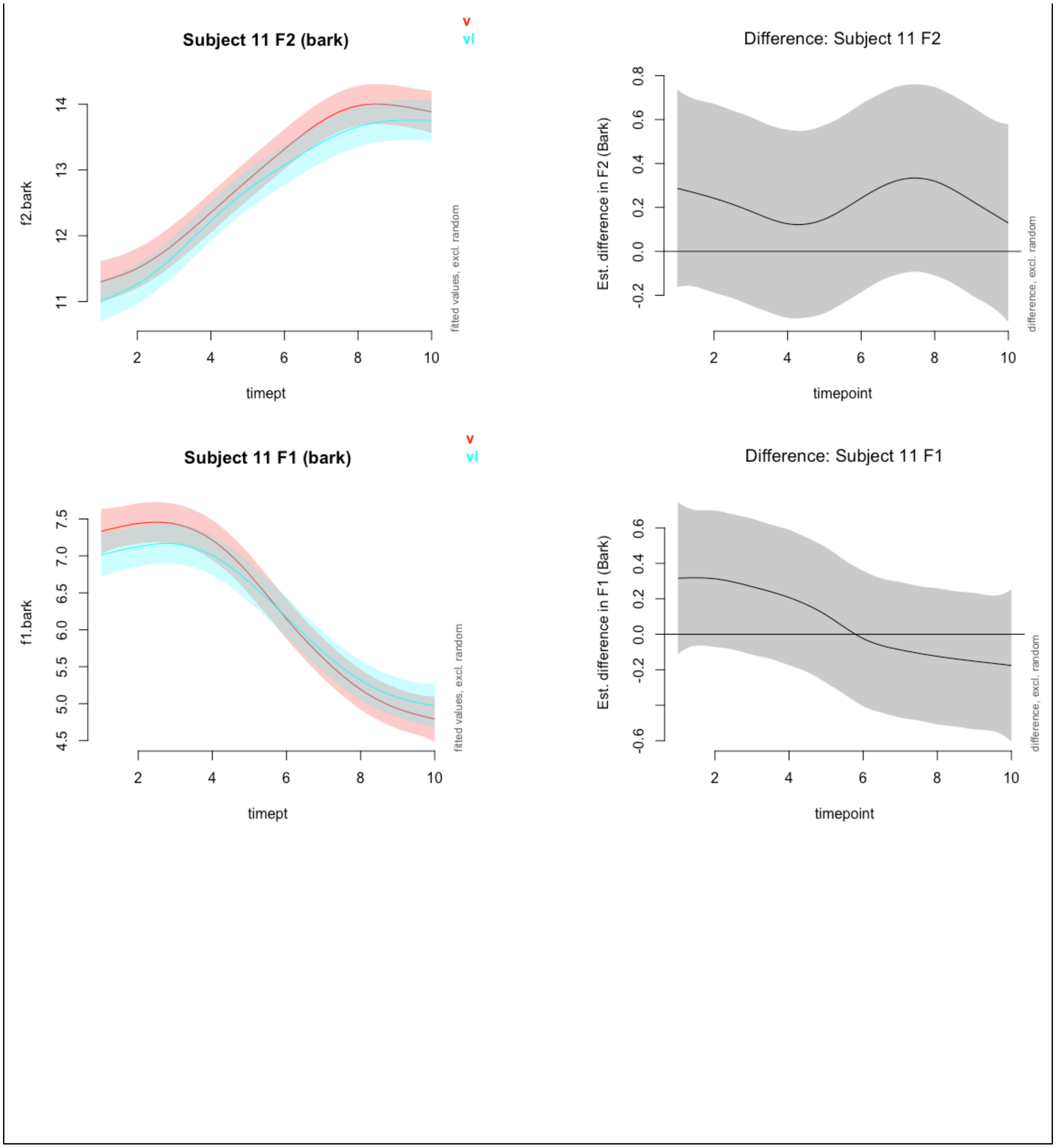


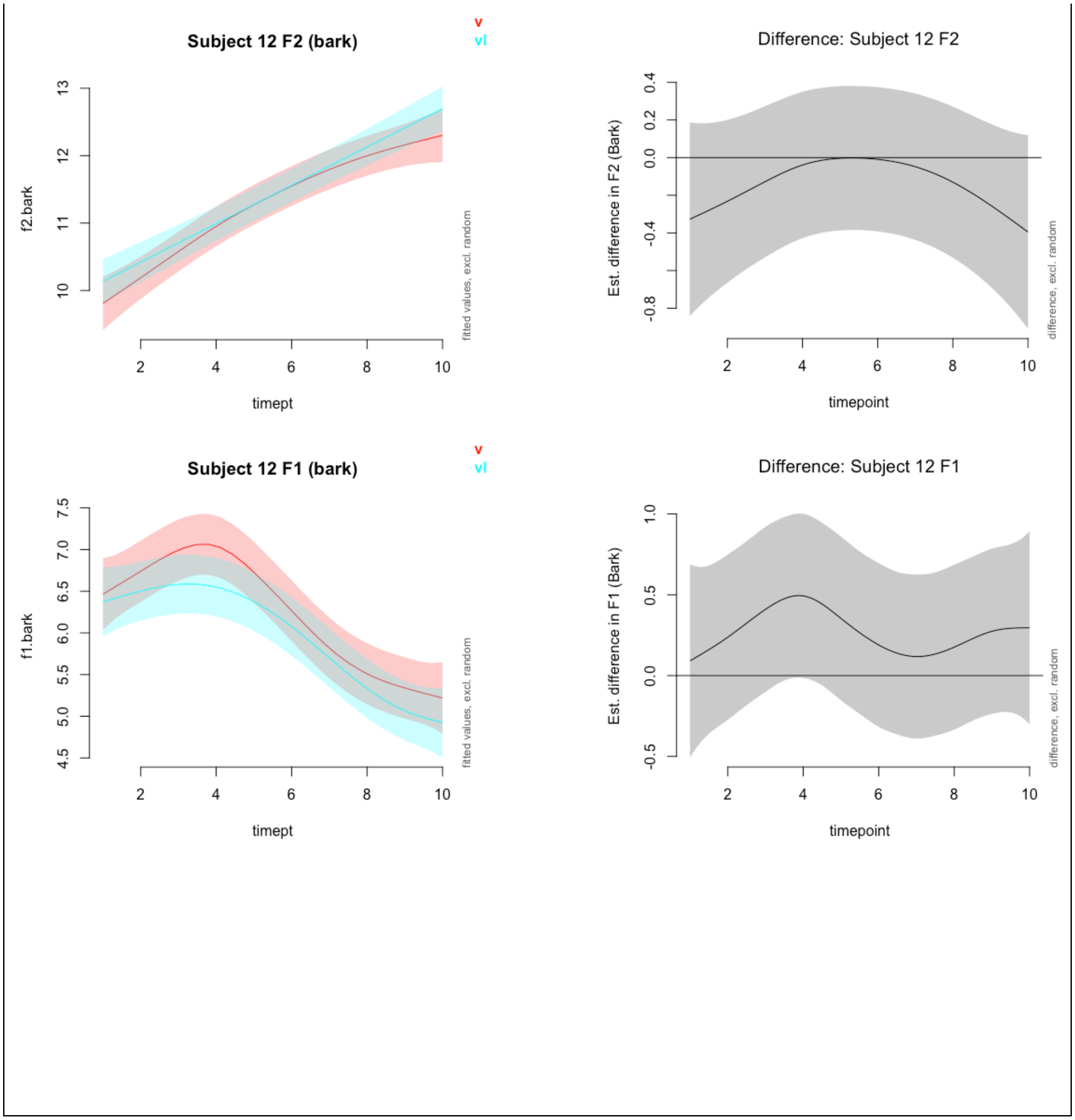


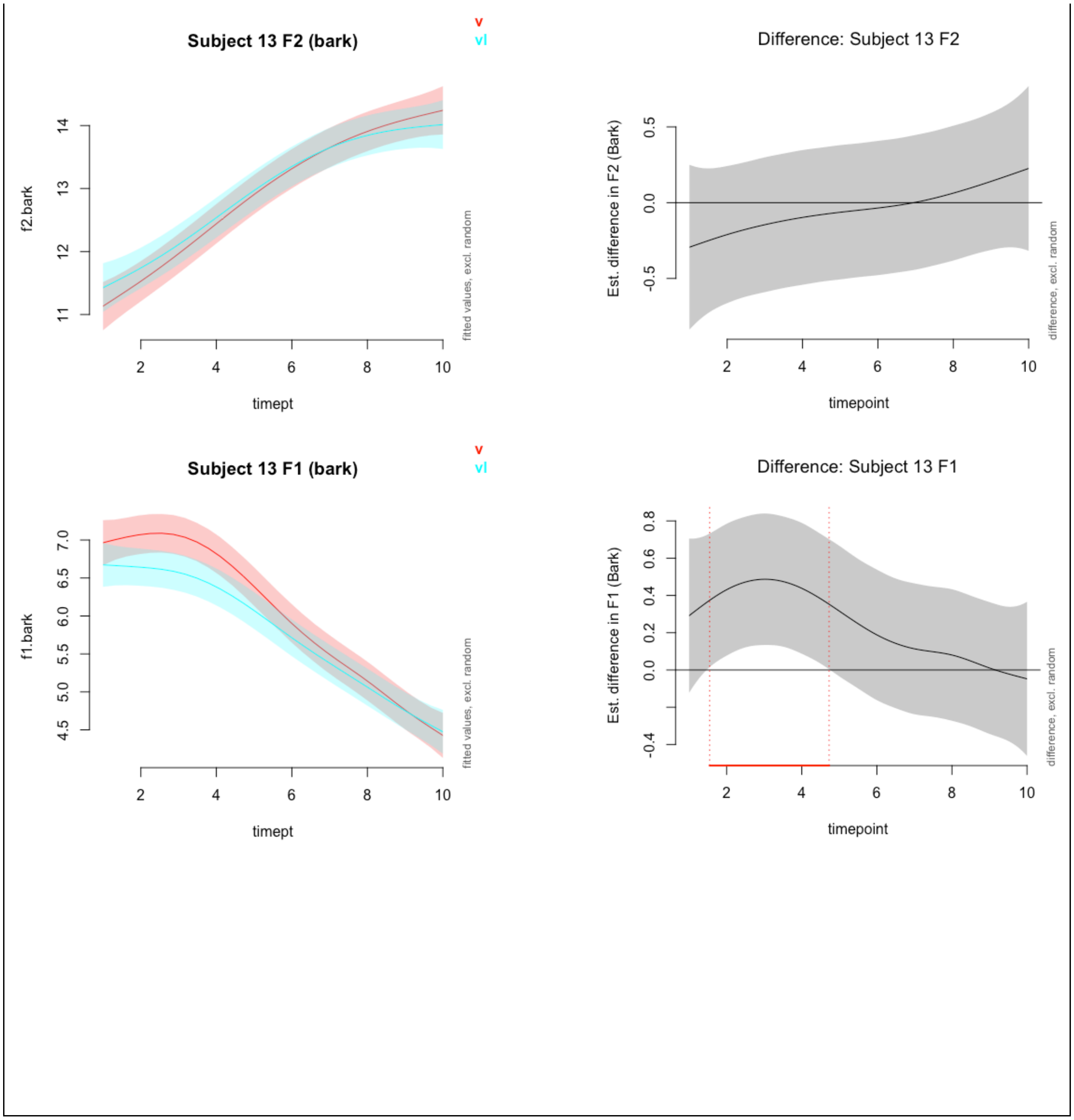
Subject 8 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 8 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 8 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 8 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random


Subject 9 F2 (bark)
v
vl
f2.bark
11.0
12.0
13.0
2
4
6
8
10
timept
fitted values, excl. random
Difference: Subject 9 F2
Est. difference in F2 (Bark)
-1.0
-0.5
0.0
0.5
timepoint
difference, excl. random
Subject 9 F1 (bark)
f1.bark
4
5
6
7
Difference: Subject 9 F1
Est. difference in F1 (Bark)
1.0

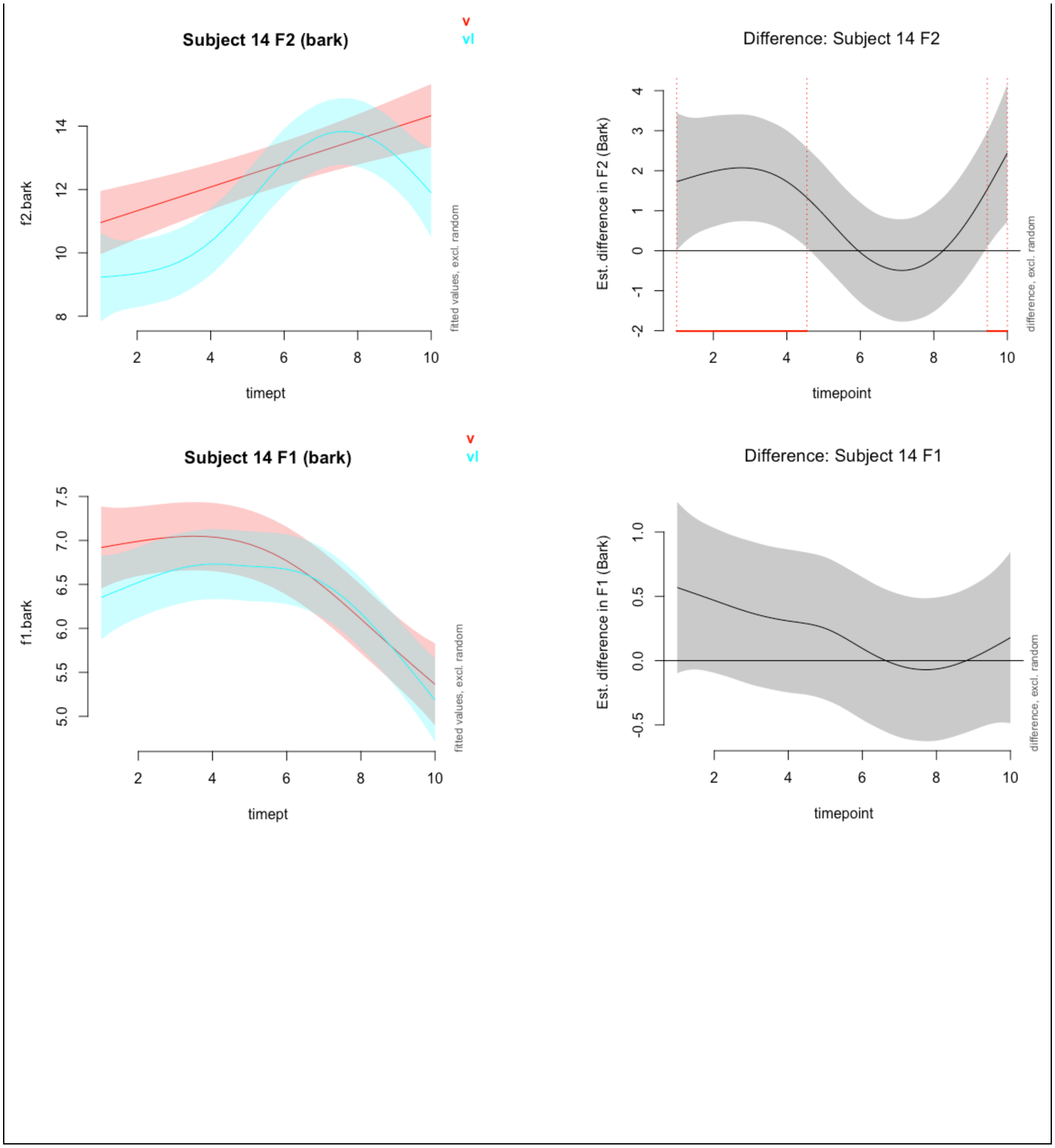

Subject 10 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 10 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 10 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 10 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

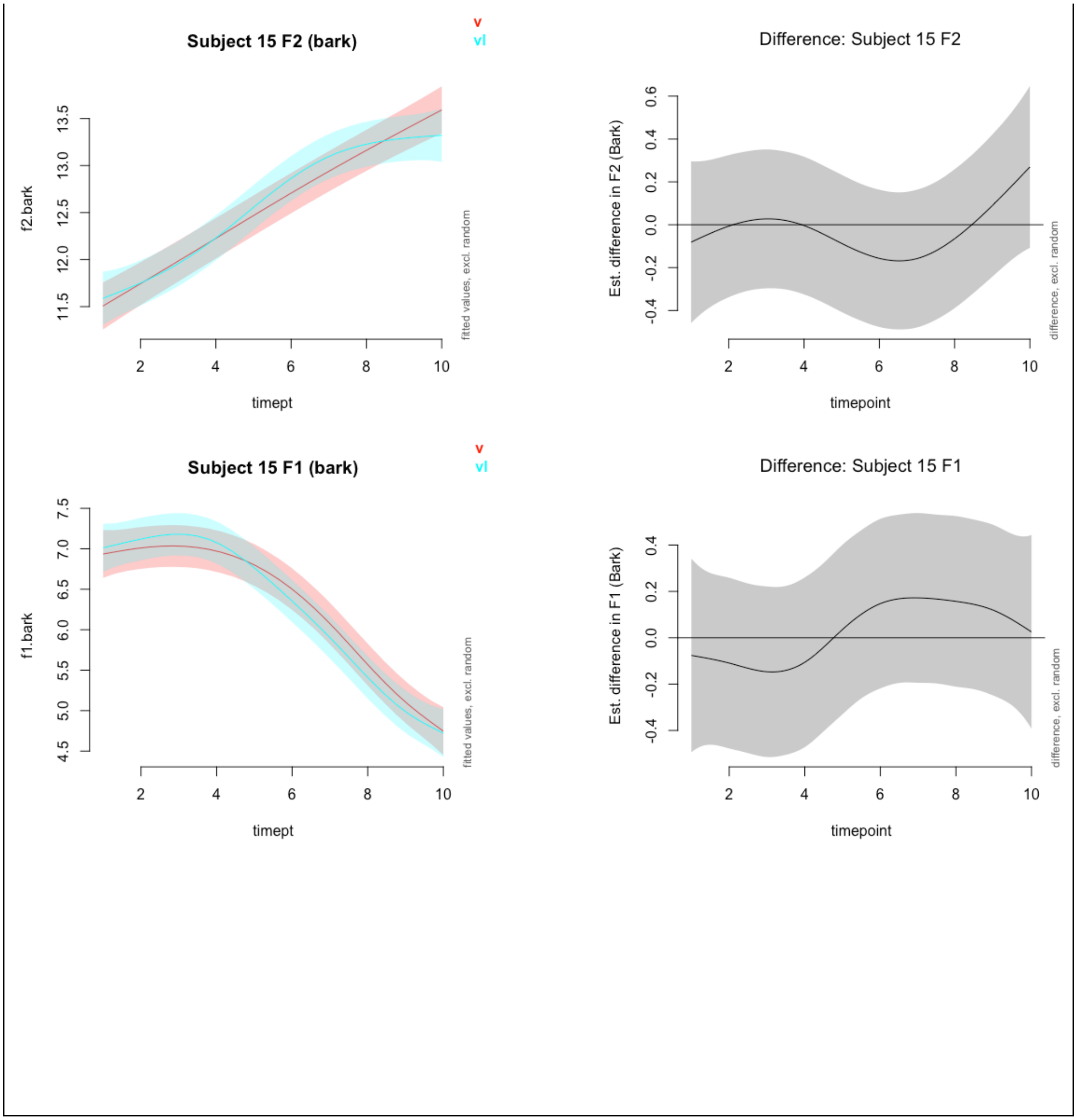

Subject 11 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 11 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 11 F1 (bark)
f1.bark
Difference: Subject 11 F1
Est. difference in F1 (Bark)

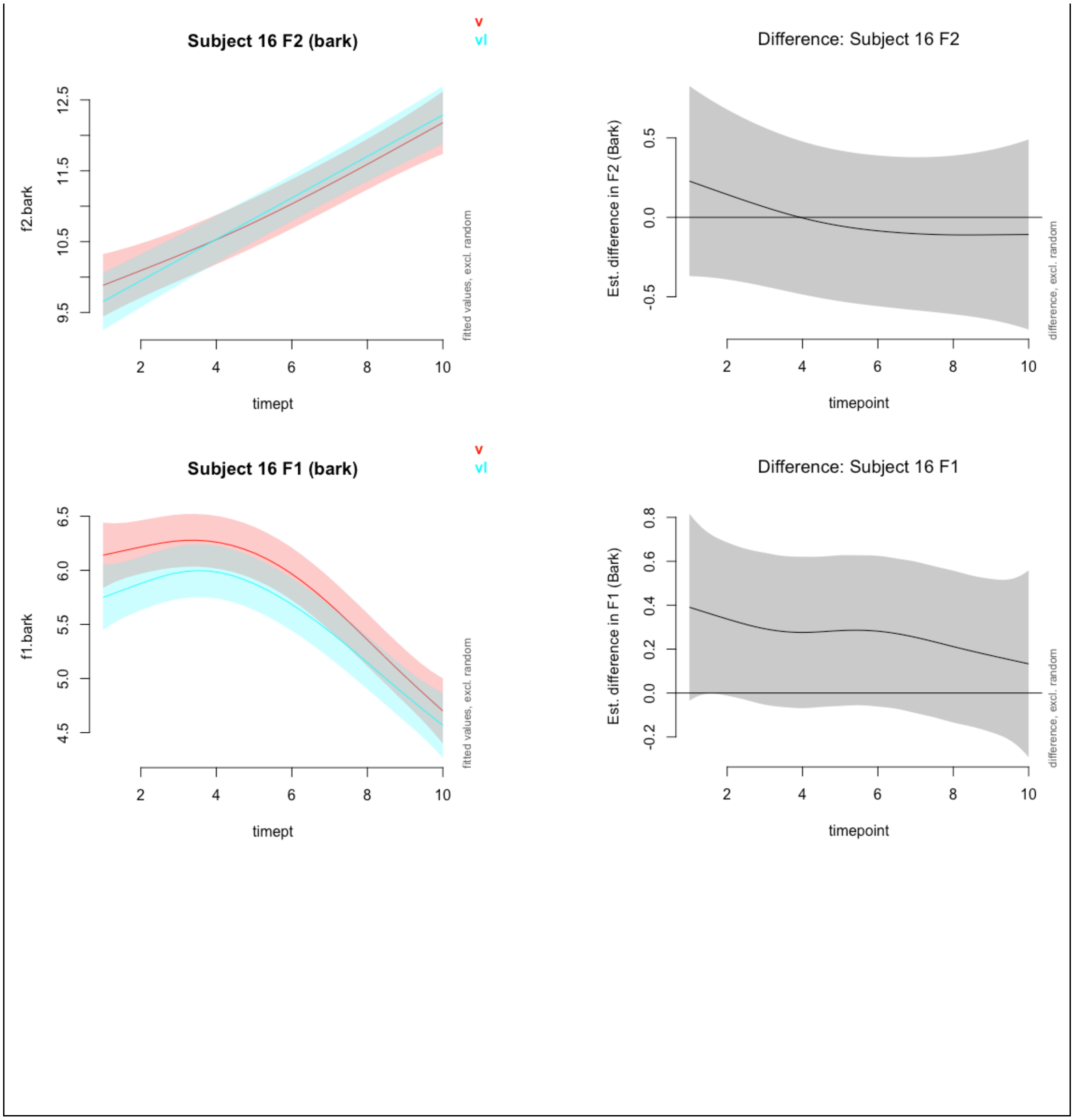

Subject 12 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 12 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 12 F1 (bark)
f1.bark
Difference: Subject 12 F1
Est. difference in F1 (Bark)



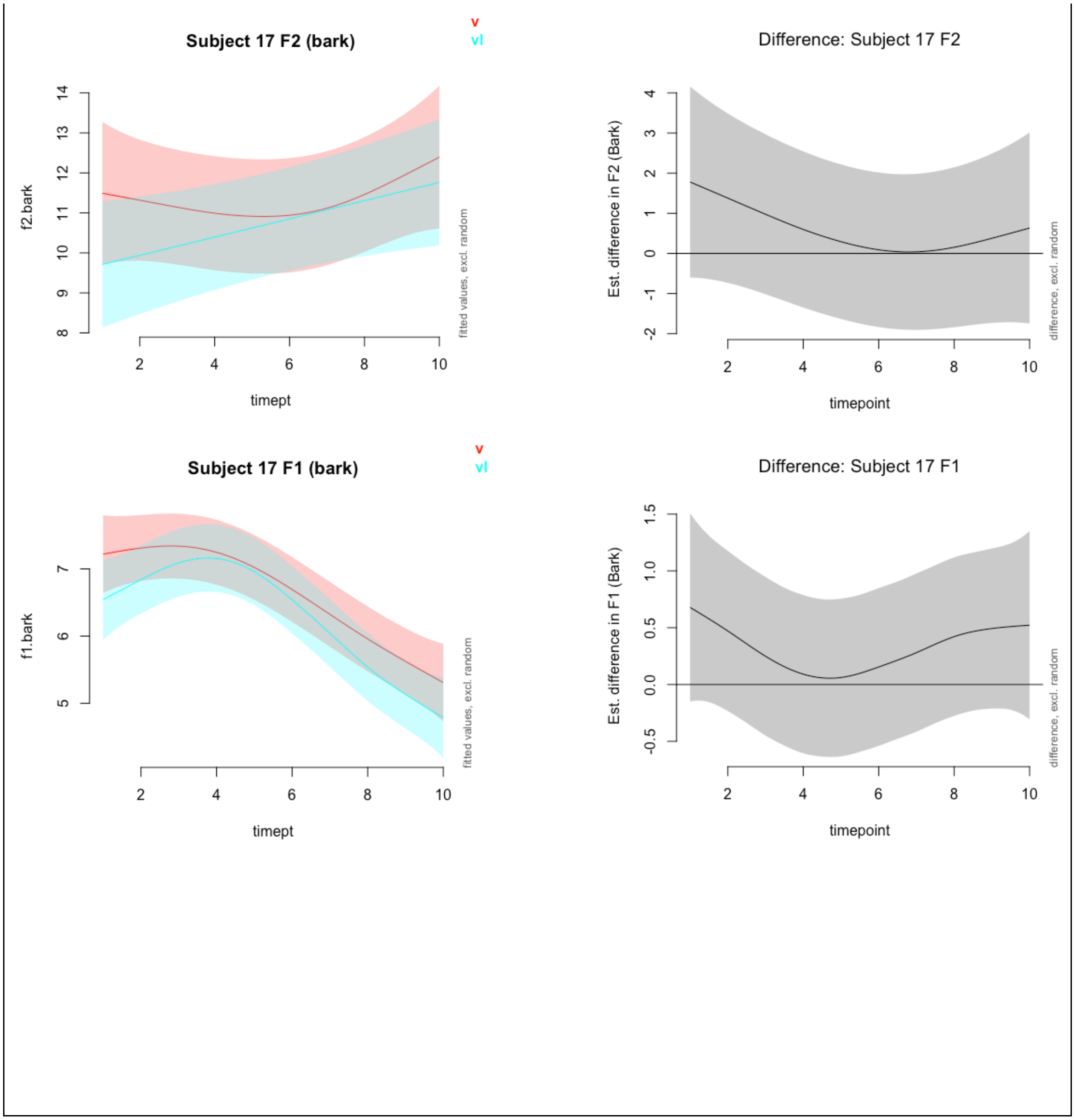
Subject 13 F2 (bark)
v
vl
f2.bark
11
12
13
14
2
4
6
8
10
timept
fitted values, excl. random
Difference: Subject 13 F2
Est. difference in F2 (Bark)
-0.5
0.0
0.5
timepoint
difference, excl. random
Subject 13 F1 (bark)
f1.bark
4.5
5.0
5.5
6.0
6.5
7.0
Difference: Subject 13 F1
Est. difference in F1 (Bark)
-0.4
0.0
0.2
0.4
0.6
0.8

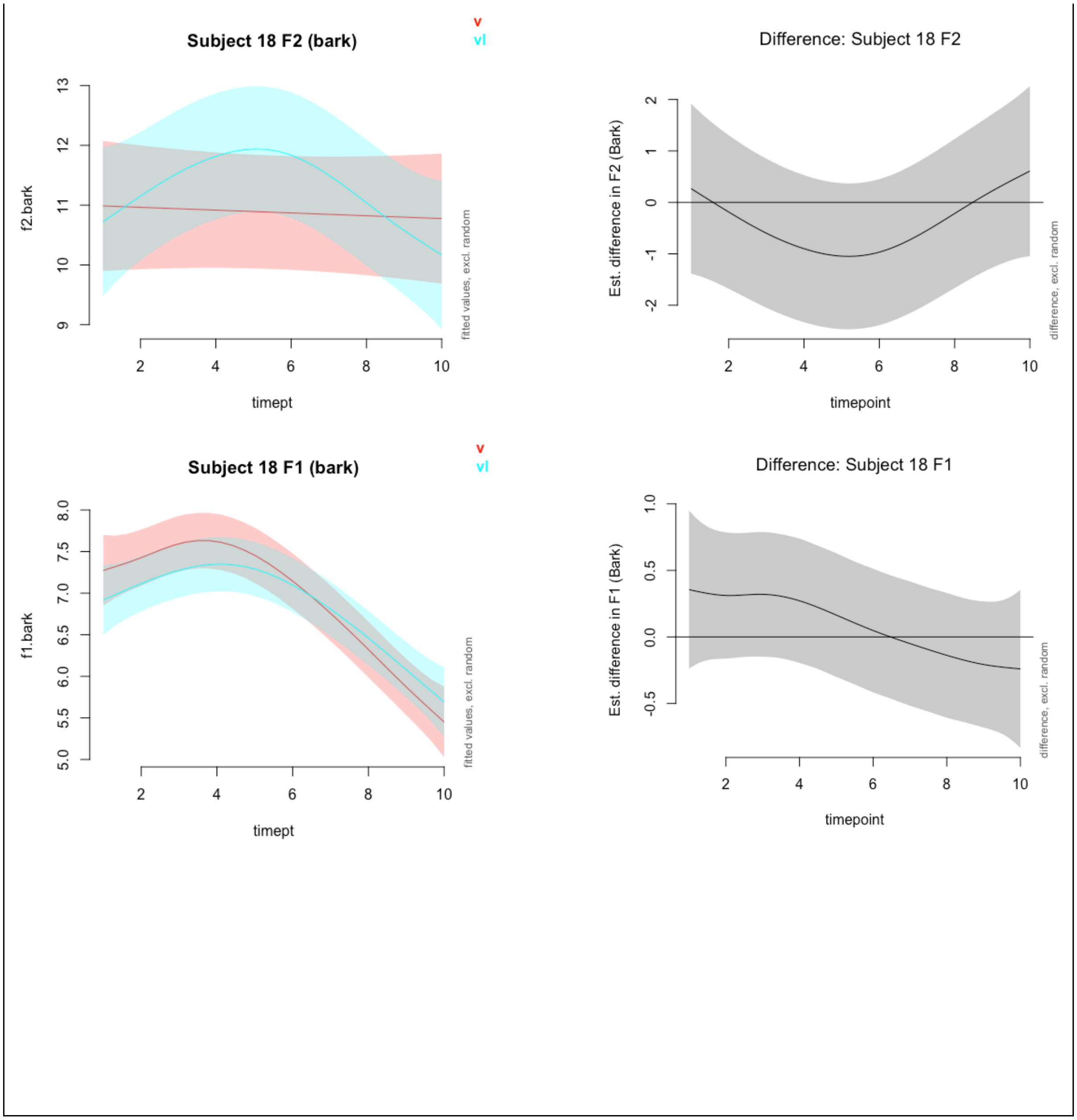

Subject 14 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 14 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 14 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 14 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random


Subject 15 F2 (bark)
v
vl
f2.bark
11.5
12.0
12.5
13.0
13.5
2
4
6
8
10
timept
fitted values, excl. random
Difference: Subject 15 F2
Est. difference in F2 (Bark)
-0.4
-0.2
0.0
0.2
0.4
0.6
2
4
6
8
10
timepoint
difference, excl. random
Subject 15 F1 (bark)
v
vl
f1.bark
4.5
5.0
5.5
6.0
6.5
7.0
7.5
2
4
6
8
10
timept
fitted values, excl. random
Difference: Subject 15 F1
Est. difference in F1 (Bark)
-0.4
-0.2
0.0
0.2
0.4
2
4
6
8
10
timepoint
difference, excl. random

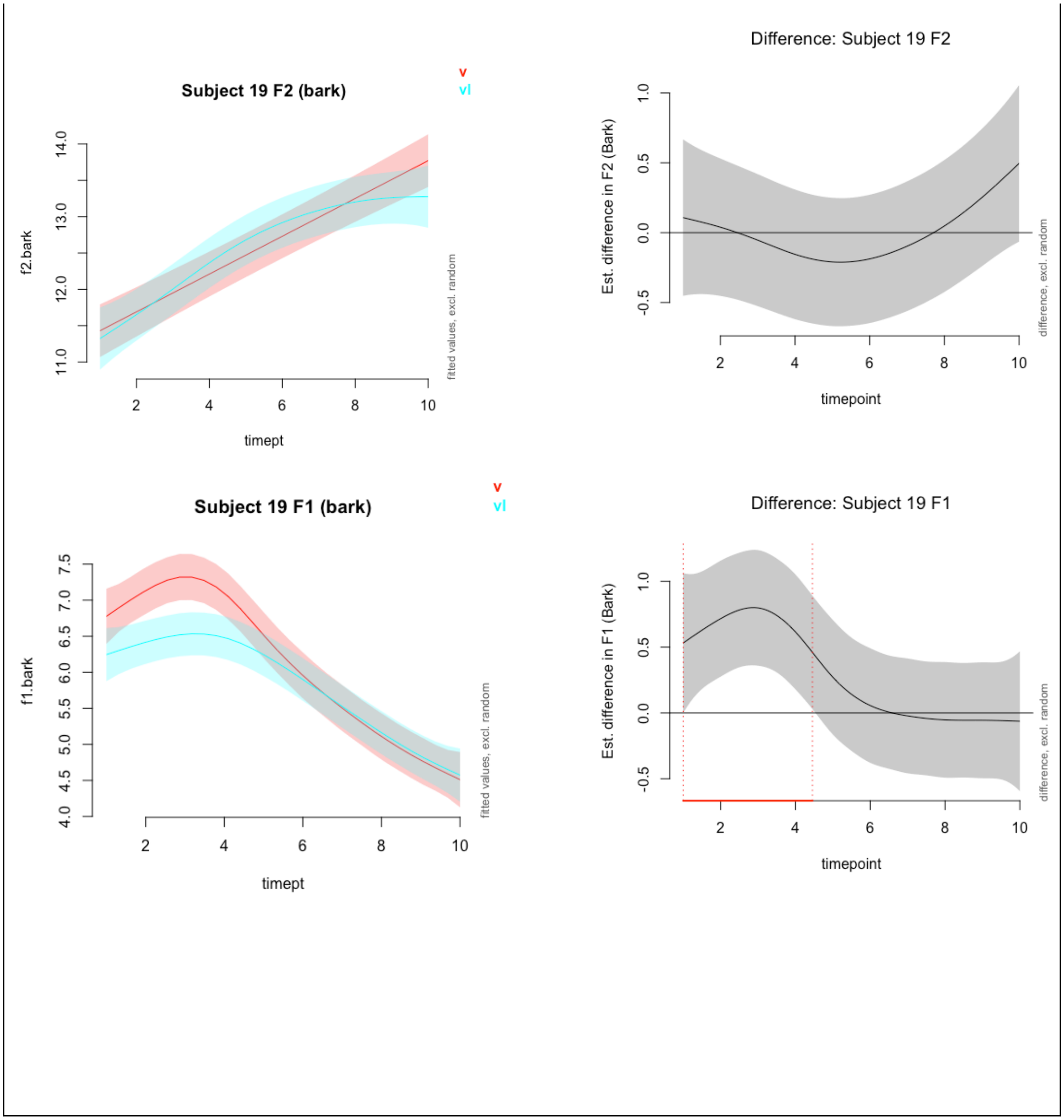


Subject 16 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 16 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 16 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 16 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

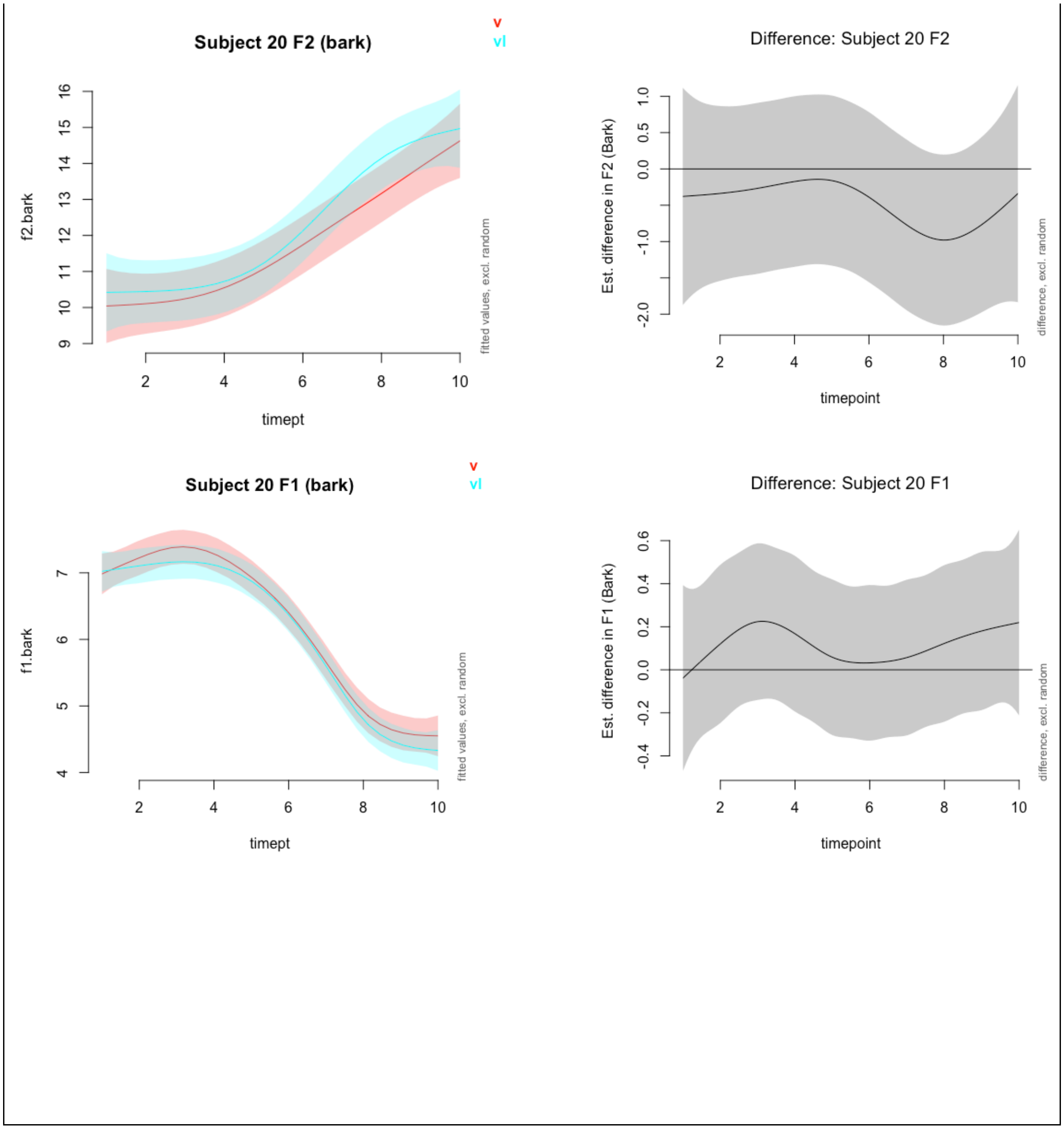

Subject 17 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 17 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 17 F1 (bark)
v
vl
f1.bark
timept
fitted values, excl. random
Difference: Subject 17 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

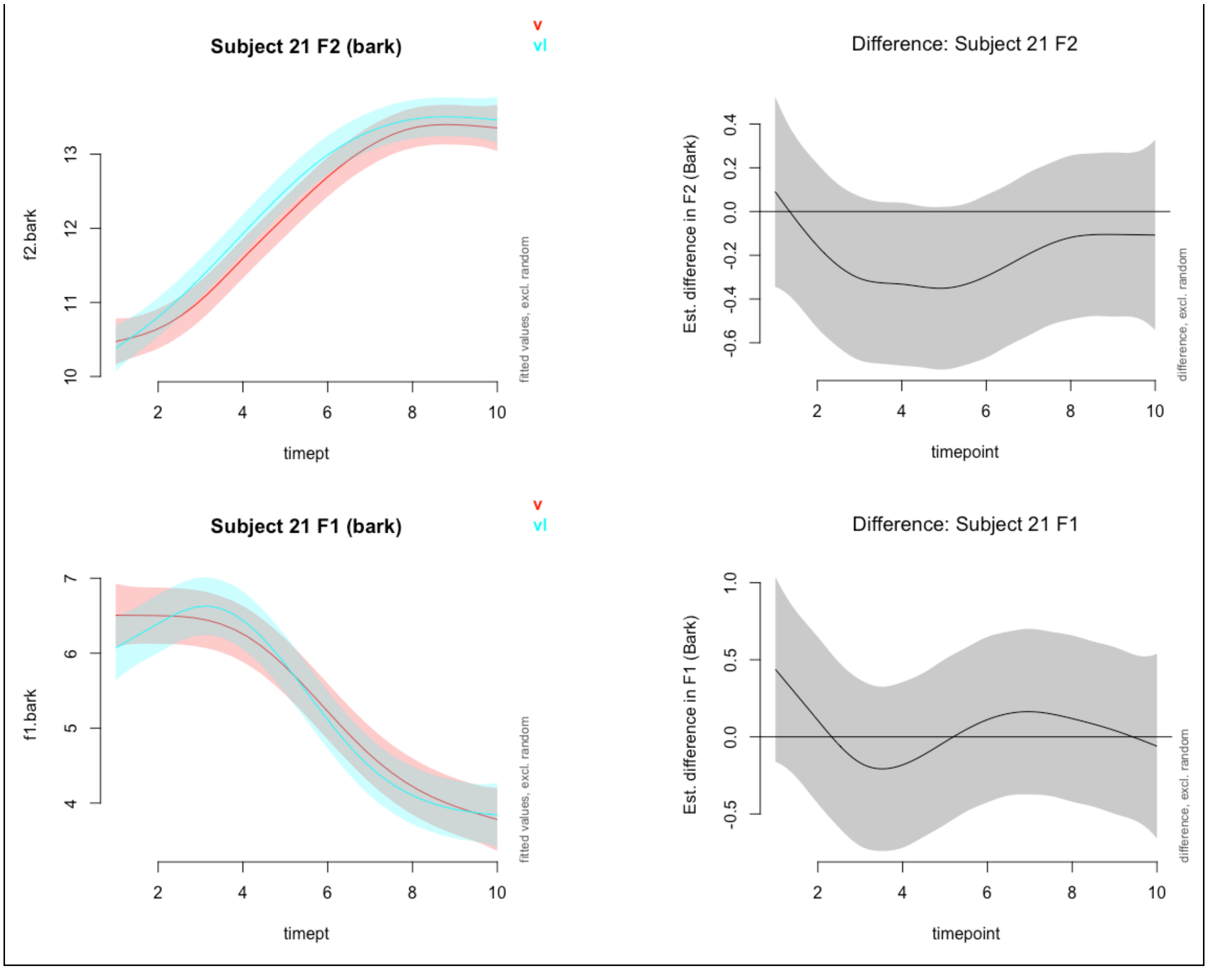


Subject 18 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 18 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 18 F1 (bark)
f1.bark
Difference: Subject 18 F1
Est. difference in F1 (Bark)

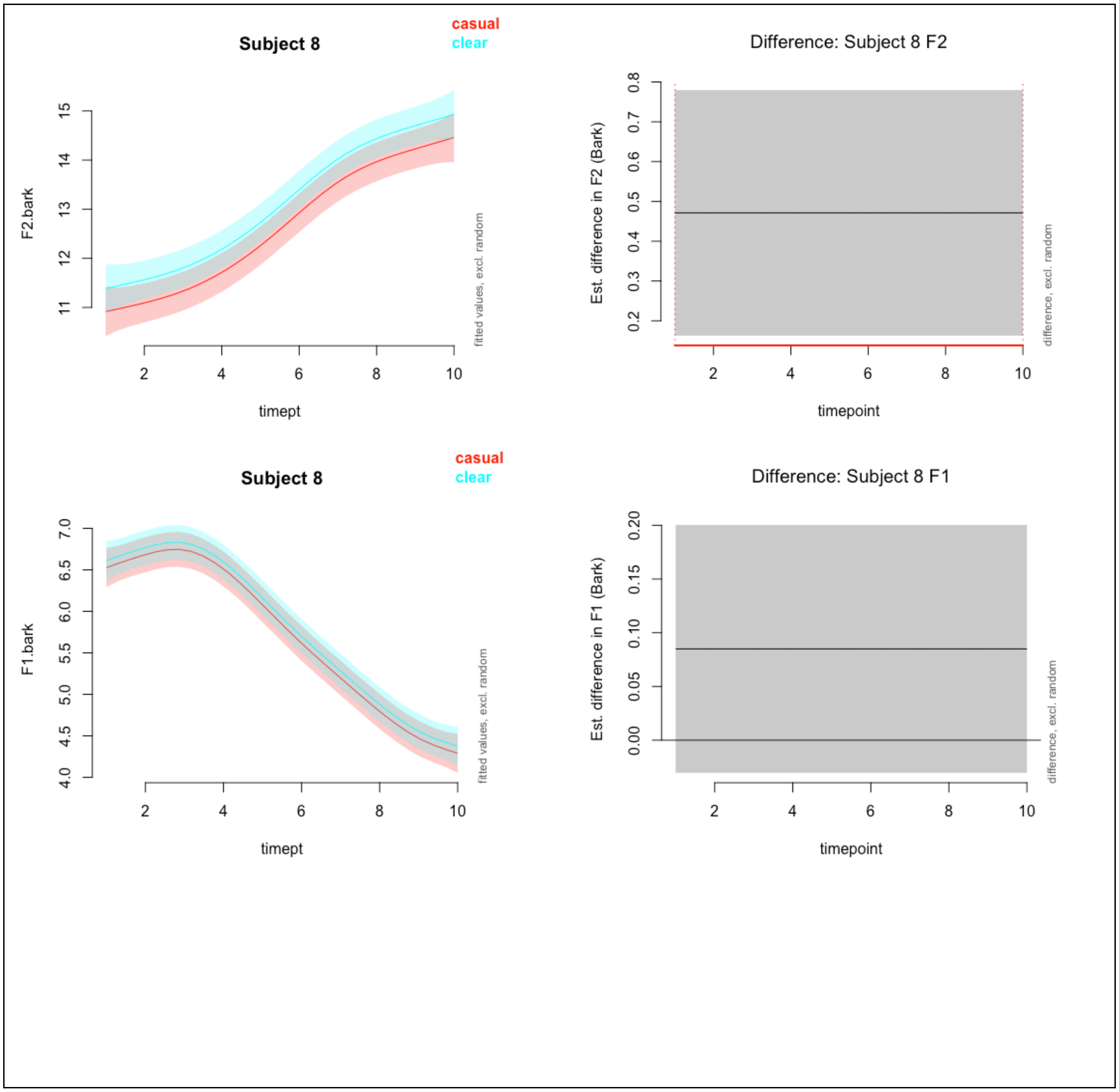


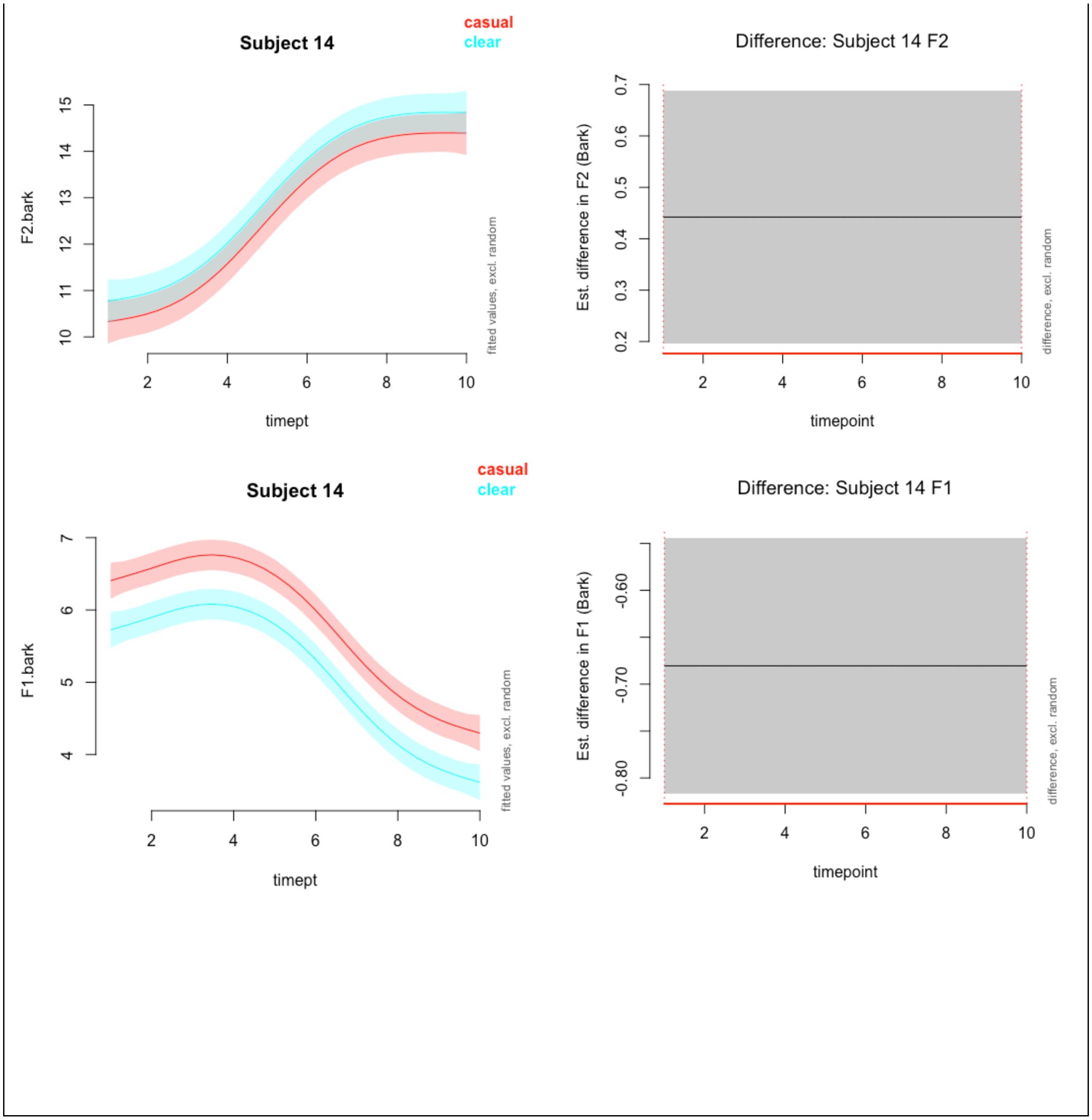
Subject 19 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 19 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 19 F1 (bark)
f1.bark
Difference: Subject 19 F1
Est. difference in F1 (Bark)

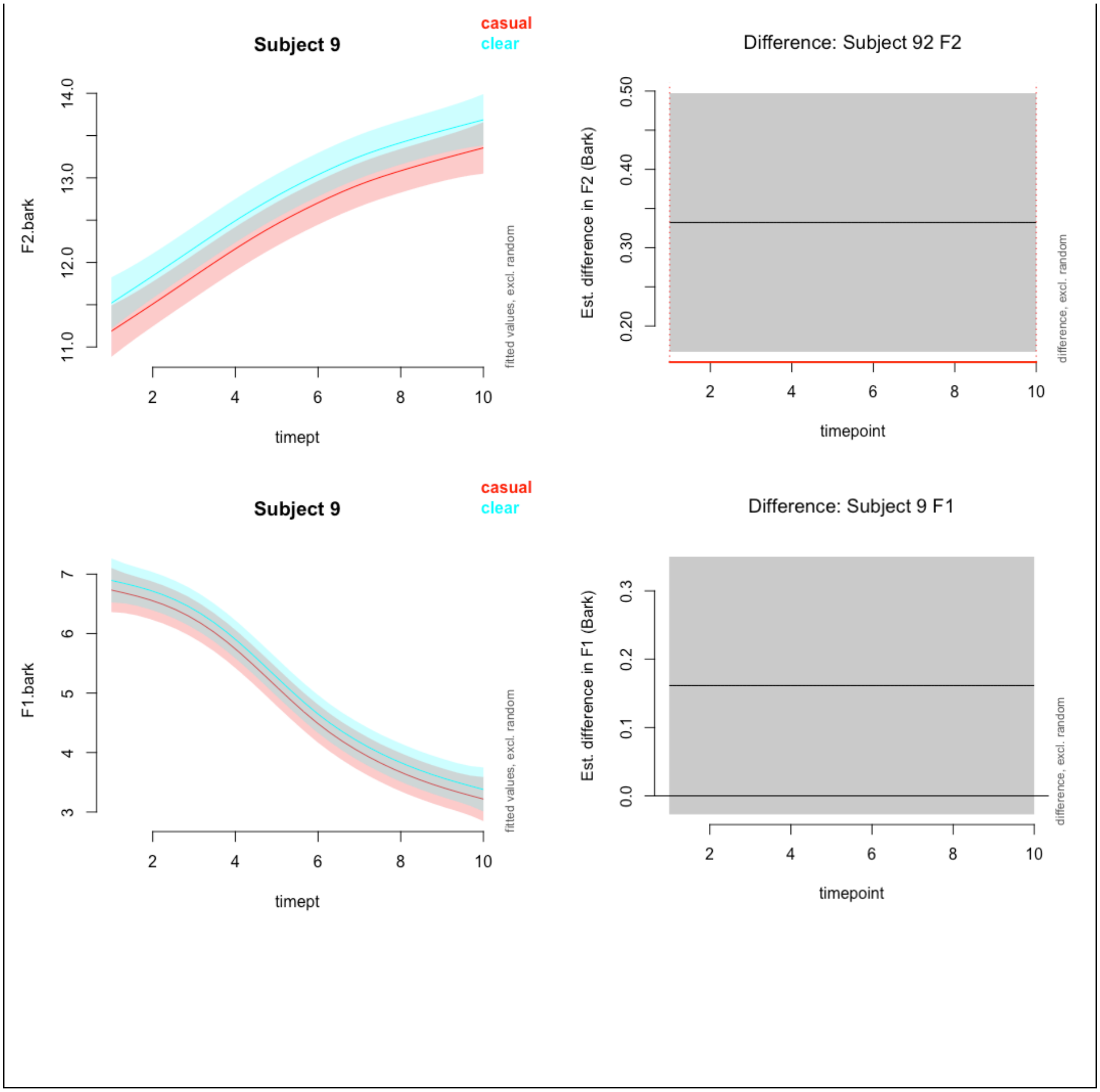

Subject 20 F2 (bark)
v
vl
f2.bark
timept
fitted values, excl. random
Difference: Subject 20 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 20 F1 (bark)
f1.bark
Difference: Subject 20 F1
Est. difference in F1 (Bark)


Subject 21 F2 (bark)
v
vl
f2.bark
13
12
11
10
2
4
6
8
10
timept
fitted values, excl. random
Difference: Subject 21 F2
Est. difference in F2 (Bark)
0.4
0.2
0.0
-0.2
-0.4
-0.6
timepoint
difference, excl. random
Subject 21 F1 (bark)
f1.bark
7
6
5
4
Difference: Subject 21 F1
Est. difference in F1 (Bark)
1.0
0.5
0.0
-0.5

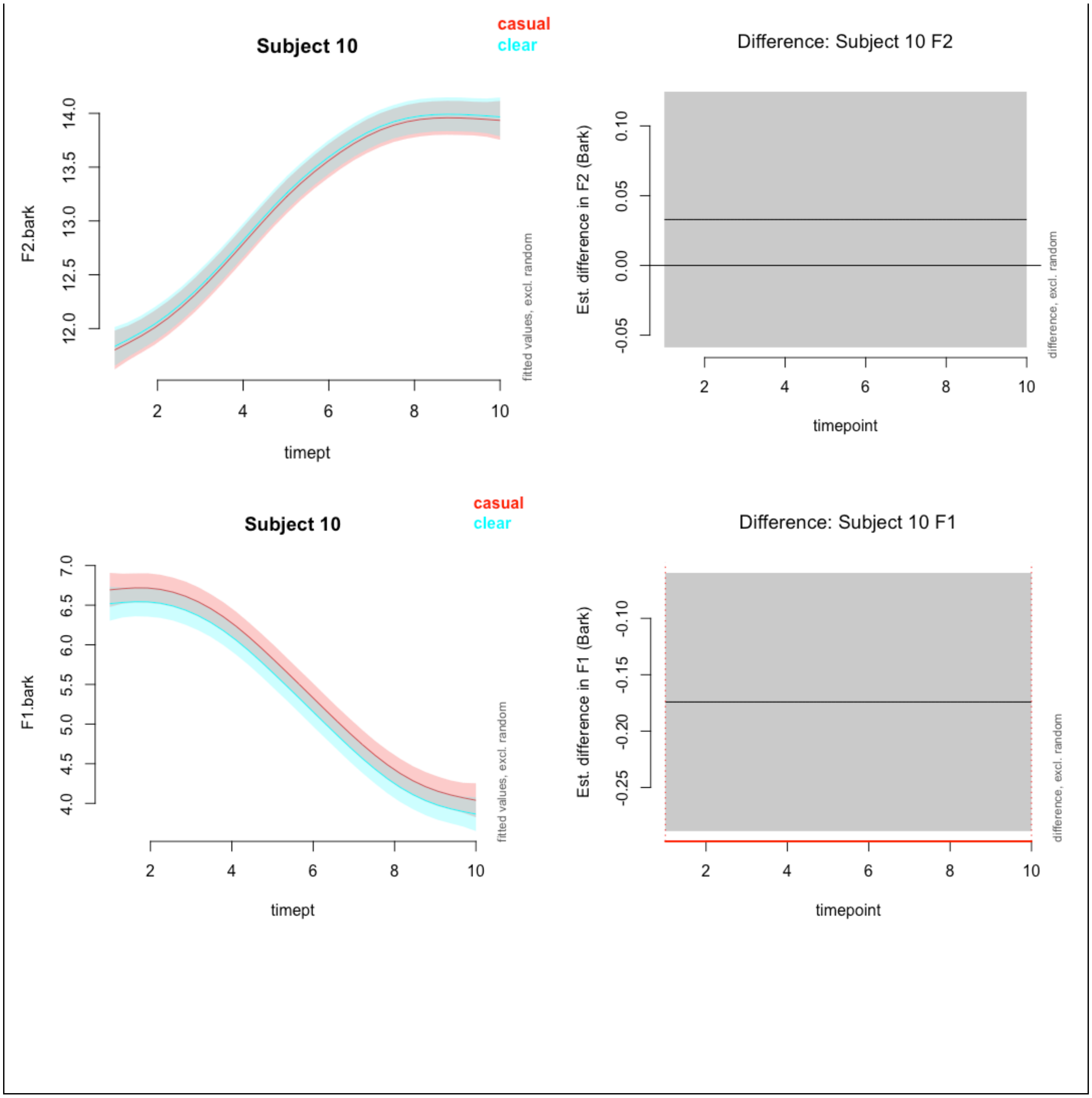
Appendix 3: Clear speech GAMM and difference plots



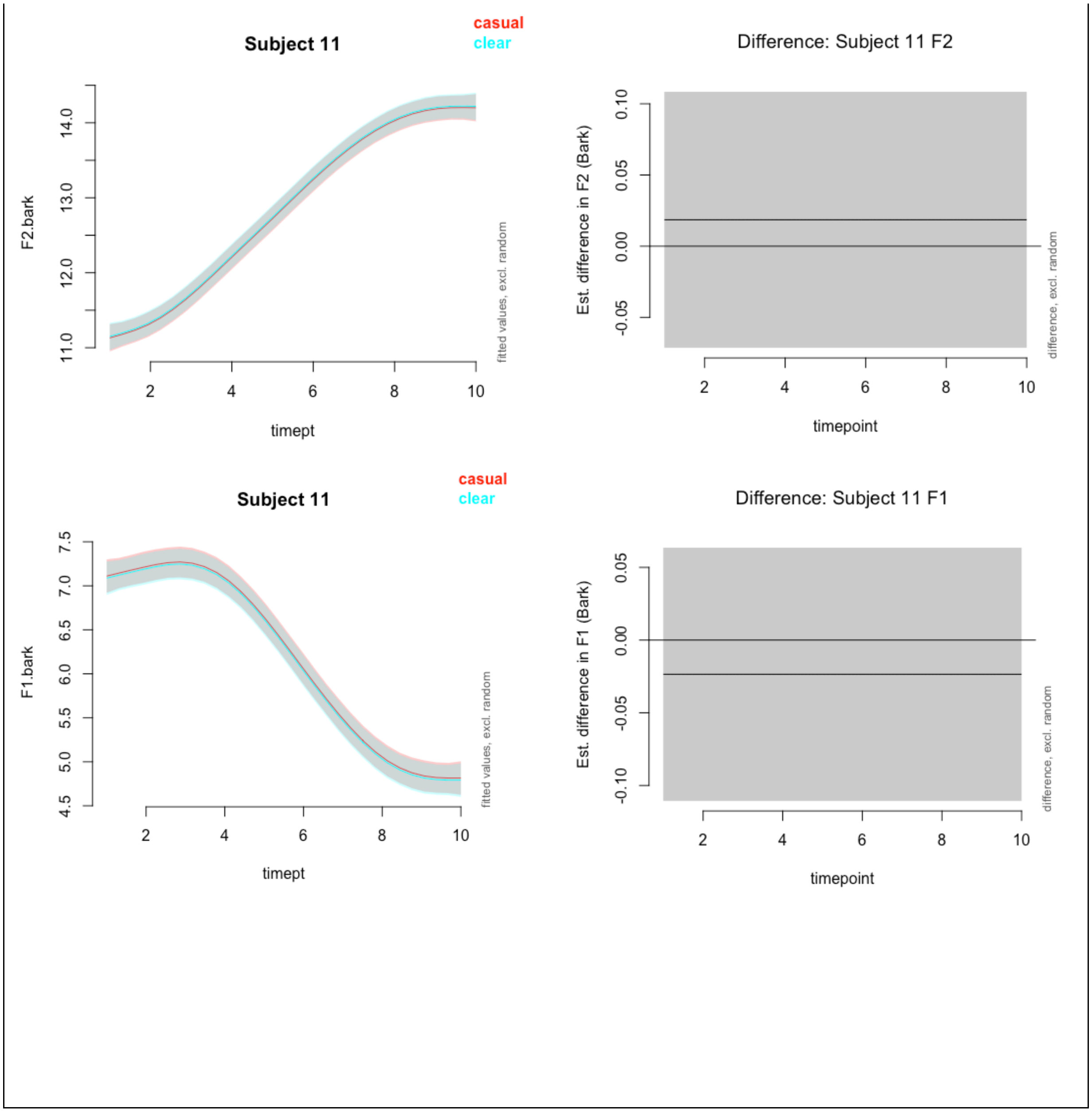

Subject 14
casual
clear
F2.bark
15
14
13
12
11
10
2
4
6
8
10
timept
fitted values, excl. random
Difference: Subject 14 F2
Est. difference in F2 (Bark)
0.7
0.6
0.5
0.4
0.3
0.2
timepoint
difference, excl. random
Subject 14
casual
clear
F1.bark
7
6
5
4
timept
fitted values, excl. random
Difference: Subject 14 F1
Est. difference in F1 (Bark)
-0.60
-0.70
-0.80
timepoint
difference, excl. random

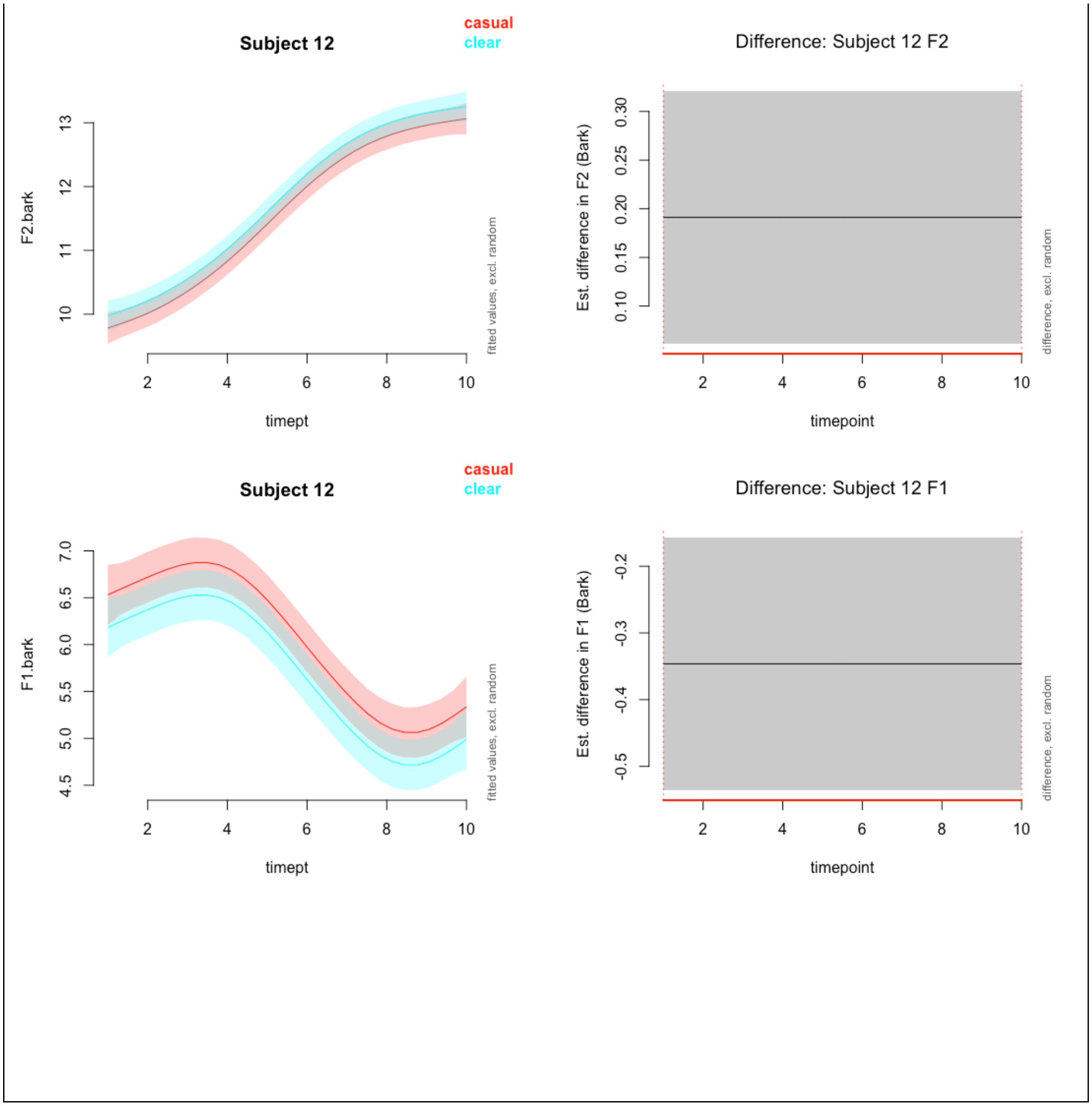


Subject 9
casual
clear
F2.bark
14.0
13.0
12.0
11.0
2
4
6
8
10
timept
fitted values, excl. random
Difference: Subject 92 F2
Est. difference in F2 (Bark)
0.50
0.40
0.30
0.20
timepoint
difference, excl. random
Subject 9
casual
clear
F1.bark
7
6
5
4
3
timept
fitted values, excl. random
Difference: Subject 9 F1
Est. difference in F1 (Bark)
0.3
0.2
0.1
0.0
timepoint
difference, excl. random

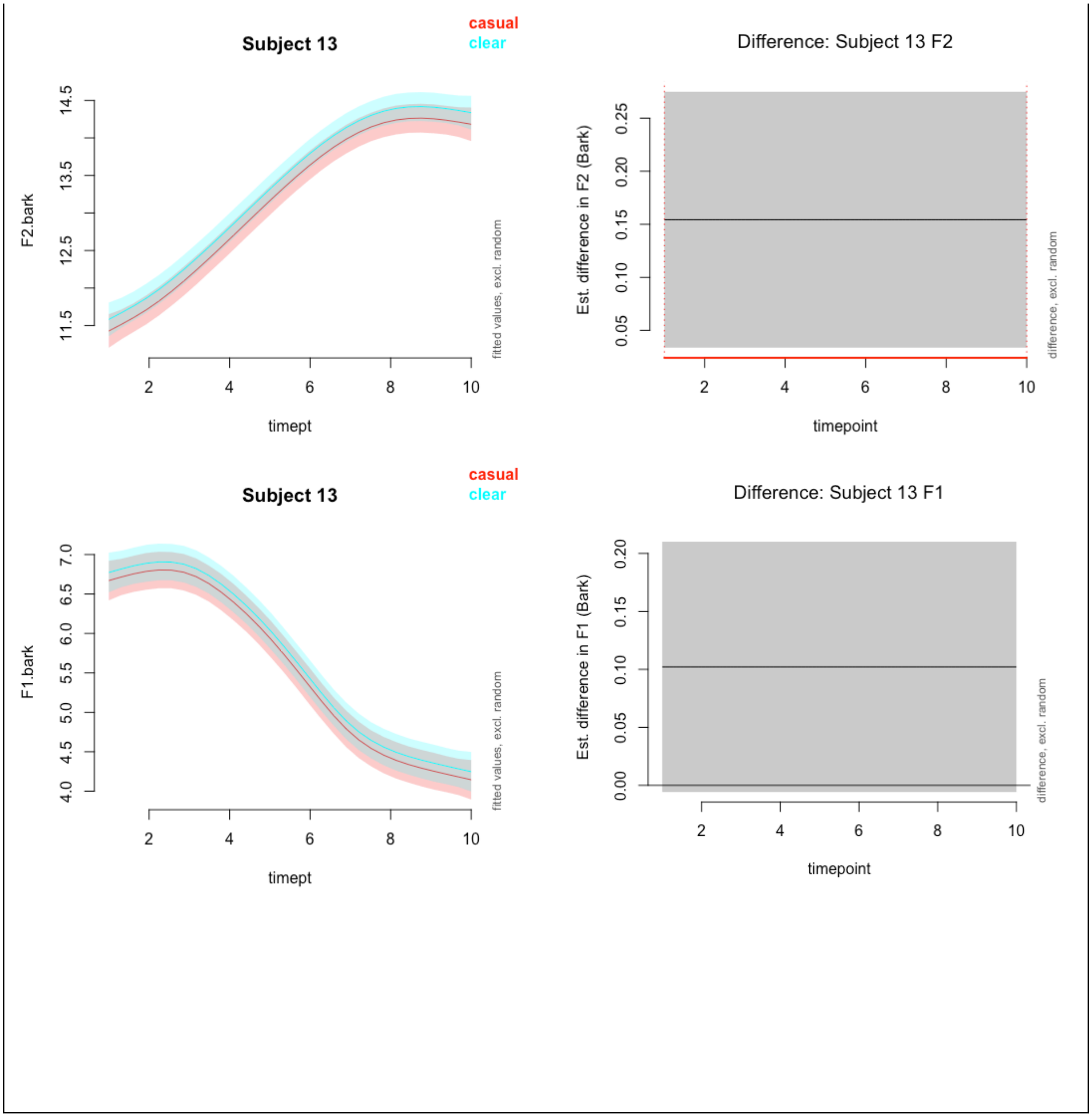


casual
clear
Subject 10
F2.bark
timept
fitted values, excl. random
Difference: Subject 10 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
casual
clear
Subject 10
F1.bark
timept
fitted values, excl. random
Difference: Subject 10 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

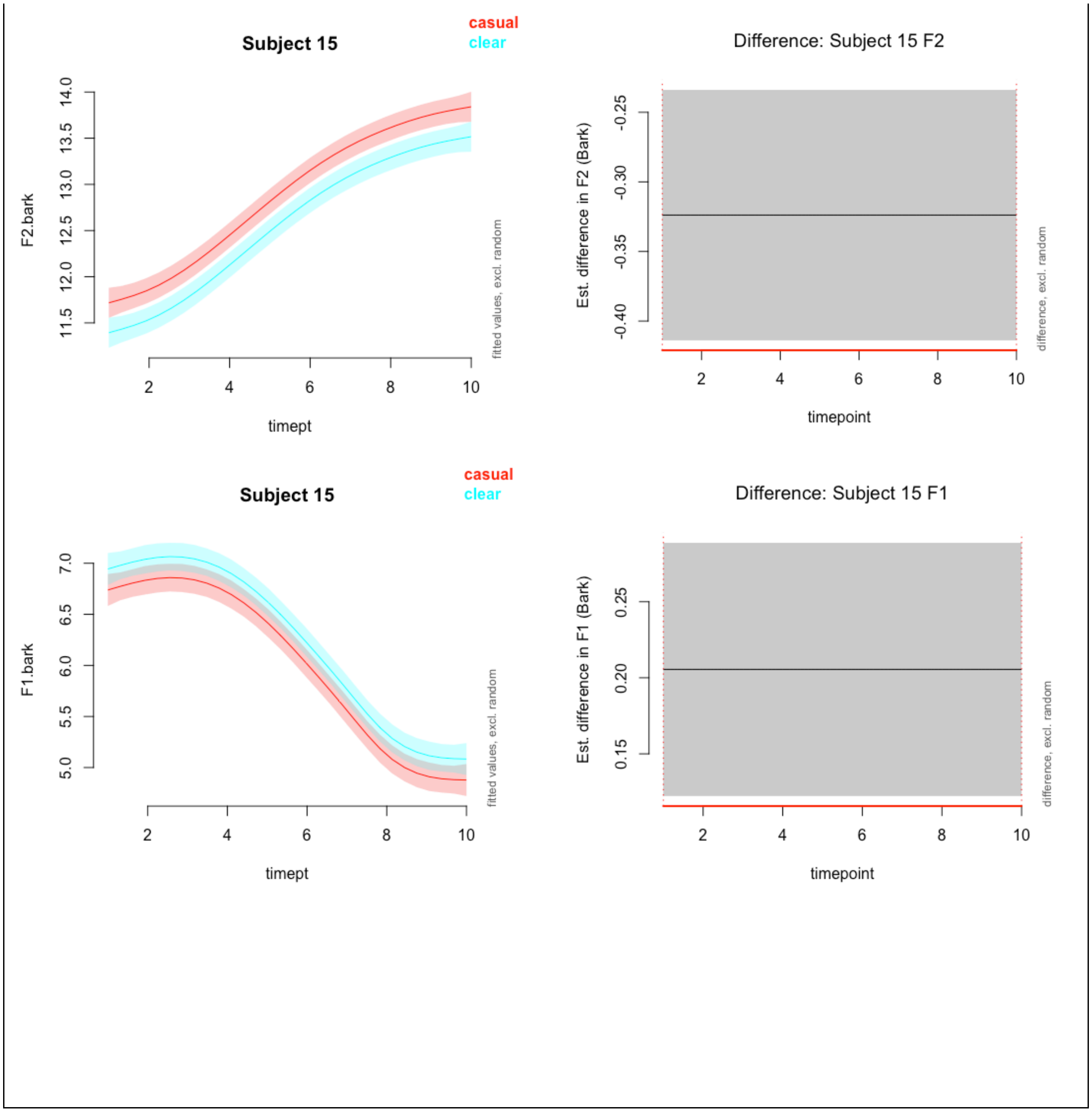

Subject 11
casual
clear
F2.bark
14.0
13.0
12.0
11.0
2
4
6
8
10
timept
fitted values, excl. random
Difference: Subject 11 F2
Est. difference in F2 (Bark)
0.10
0.05
0.00
-0.05
timepoint
difference, excl. random
Subject 11
casual
clear
F1.bark
7.5
7.0
6.5
6.0
5.5
5.0
4.5
timept
fitted values, excl. random
Difference: Subject 11 F1
Est. difference in F1 (Bark)
0.05
0.00
-0.05
-0.10
timepoint
difference, excl. random



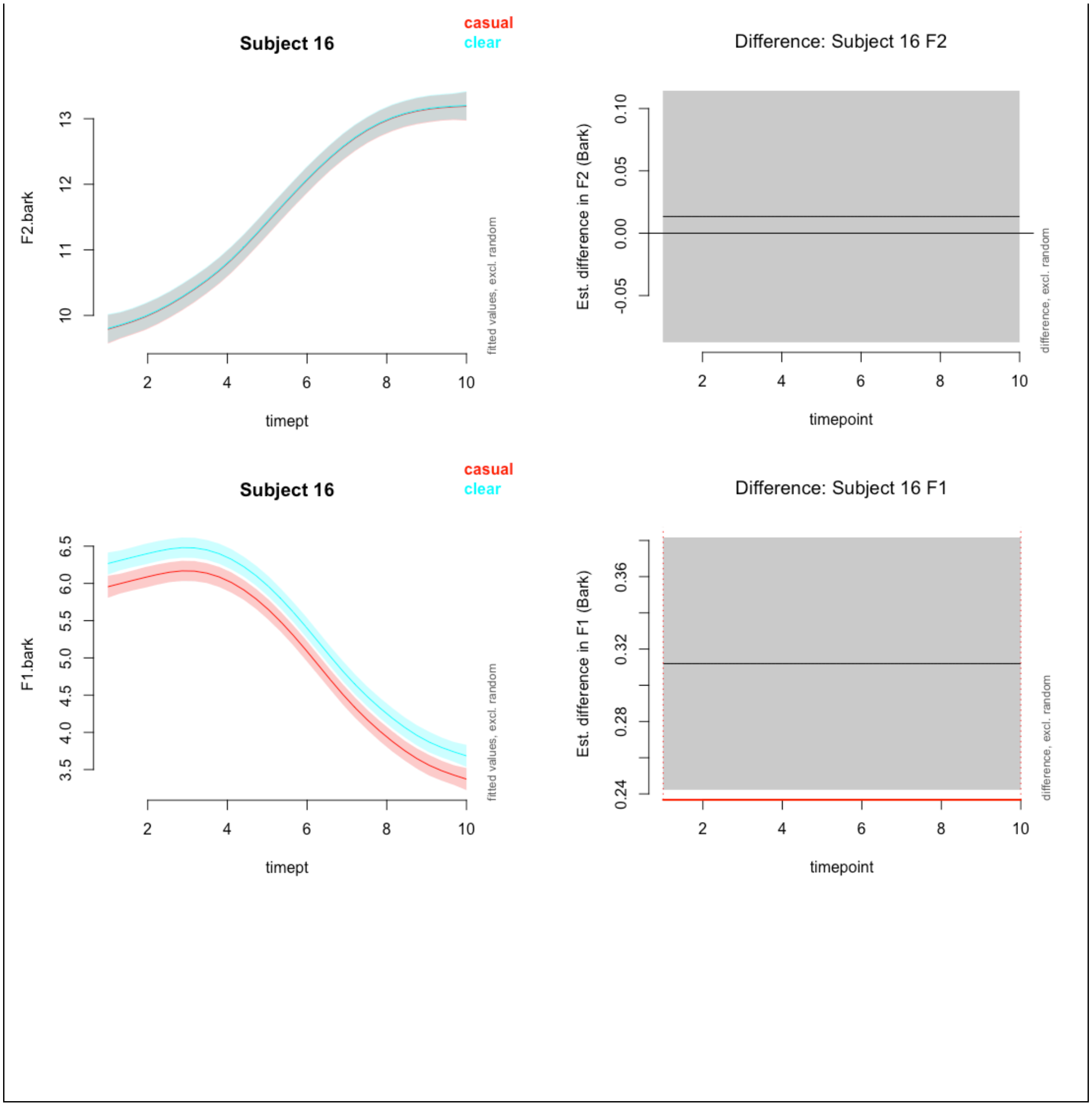
Subject 12
casual
clear
F2.bark
10
11
12
13
2
4
6
8
10
timept
fitted values, excl. random
Difference: Subject 12 F2
Est. difference in F2 (Bark)
0.10
0.15
0.20
0.25
0.30
timepoint
difference, excl. random
Subject 12
casual
clear
F1.bark
4.5
5.0
5.5
6.0
6.5
7.0
timept
fitted values, excl. random
Difference: Subject 12 F1
Est. difference in F1 (Bark)
-0.5
-0.4
-0.3
-0.2
timepoint
difference, excl. random



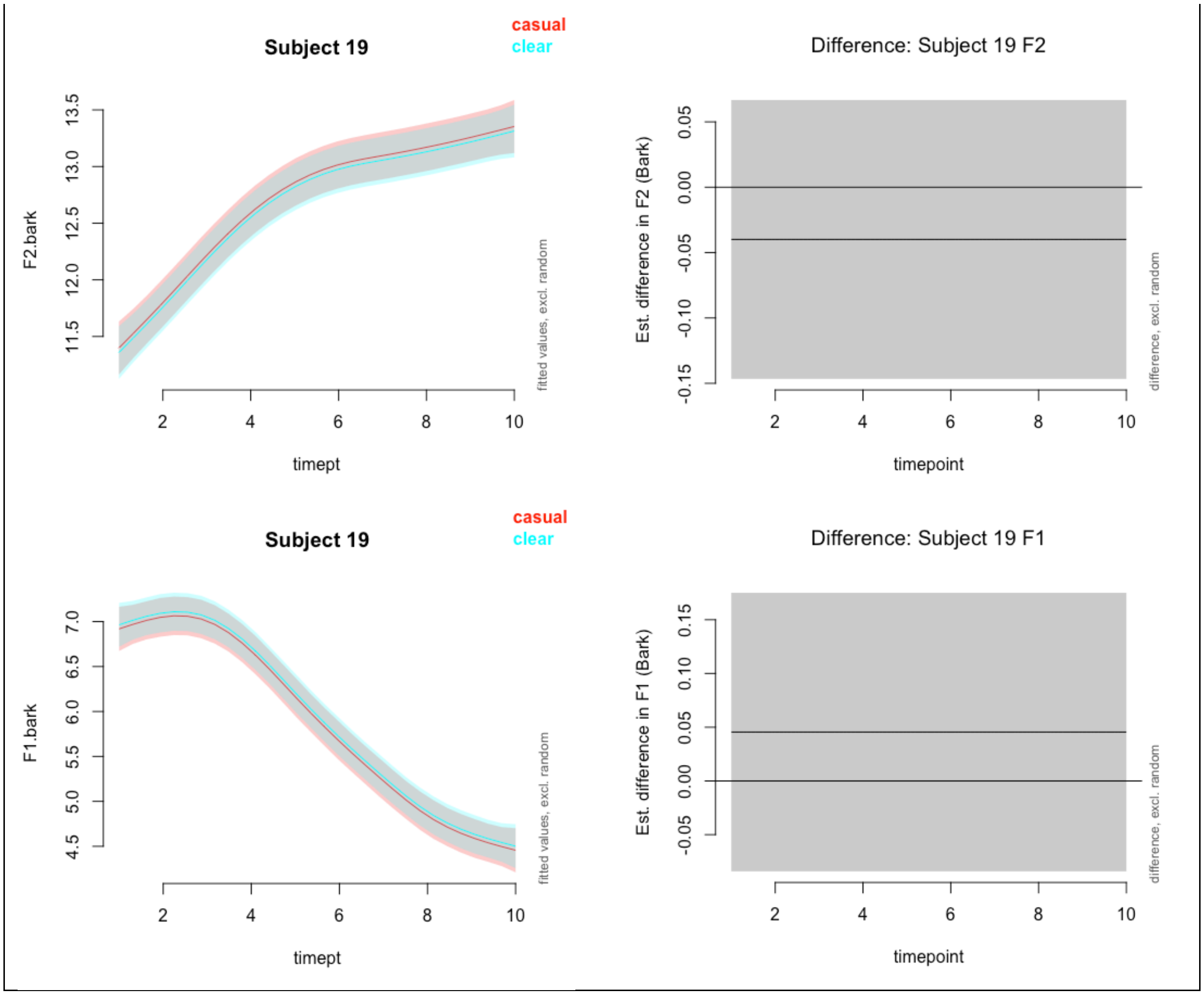
Subject 13
casual
clear
F2.bark
timept
fitted values, excl. random
Difference: Subject 13 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 13
casual
clear
F1.bark
timept
fitted values, excl. random
Difference: Subject 13 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random



Subject 15
casual
clear
F2.bark
timept
fitted values, excl. random
Difference: Subject 15 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 15
casual
clear
F1.bark
timept
fitted values, excl. random
Difference: Subject 15 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random



Subject 16
casual
clear
F2.bark
timept
fitted values, excl. random
Difference: Subject 16 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 16
casual
clear
F1.bark
timept
fitted values, excl. random
Difference: Subject 16 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random


Subject 19
casual
clear
F2.bark
timept
fitted values, excl. random
Difference: Subject 19 F2
Est. difference in F2 (Bark)
timepoint
difference, excl. random
Subject 19
casual
clear
F1.bark
timept
fitted values, excl. random
Difference: Subject 19 F1
Est. difference in F1 (Bark)
timepoint
difference, excl. random

Appendix 4: Word list used in production task

tidal

title

advice

advise

biting

biding

bite

bide

bright

bride

brighter

device

devise

cycle

dice

dies

dive

guide

kite

height

hide

hiding

heighten

ice

eyes

hike

lice

lies

life

light

price

prize

write

ride

writing

riding

quite

rice

rise

sight

sighting

side

siding

spite

strife

strive

tight

tide

gigantic

titanic

psychology

cypress

diameter

dynamic

hibernation

hydraulic

hydrangea

hyena

hyperbole

hypothesis

isolate

itemize

itinerary

librarian

microbial

migration

vibration

spy

lie

pliers

dial

buy

cider

spider

fire

bot

body

rot

rod

got

dot

pod

sod

lot

prod

beat

bead

reed

deed

feet

seat

heat

heed

please

peace

bud

but

rut

hut

suds

luck

lug

cut

cub

sub

food

hoot

shoes

suit