

## Article

# Distributional and Acoustic Characteristics of Filler Particles in German with Consideration of Forensic-Phonetic Aspects

Beeke Muhlack <sup>1,\*</sup> , Jürgen Trouvain <sup>1</sup> and Michael Jessen <sup>2</sup><sup>1</sup> Language Science and Technology, Saarland University, 66123 Saarbrücken, Germany<sup>2</sup> KT 34 Sprache, Audio, Bundeskriminalamt, 65193 Wiesbaden, Germany\* Correspondence: [muhlack@lst.uni-saarland.de](mailto:muhlack@lst.uni-saarland.de)

**Abstract:** In this study, we investigate the use of the filler particles (FPs) *uh*, *um*, *hm*, as well as glottal FPs and tongue clicks of 100 male native German speakers in a corpus of spontaneous speech. For this purpose, the frequency distribution, FP duration, duration of pauses surrounding FPs, voice quality of FPs, and their vowel quality are investigated in two conditions, namely, normal speech and Lombard speech. Speaker-specific patterns are investigated on the basis of twelve sample speakers. Our results show that tongue clicks and glottal FPs are as common as typically described FPs, and should be a part of disfluency research. Moreover, the frequency of *uh*, *um*, and *hm* decreases in the Lombard condition while the opposite is found for tongue clicks. Furthermore, along with the usual F1 increase, a considerable reduction in vowel space is found in the Lombard condition for the vowels in *uh* and *um*. A high degree of within- and between-speaker variation is found on the individual speaker level.

**Keywords:** filler particles; filled pauses; pauses; disfluencies; hesitation; speaker-specificity; forensic phonetics



**Citation:** Muhlack, Beeke, Jürgen Trouvain, and Michael Jessen. 2023. Distributional and Acoustic Characteristics of Filler Particles in German with Consideration of Forensic-Phonetic Aspects. *Languages* 8: 100. <https://doi.org/10.3390/languages8020100>

Academic Editor: Jeanine Treffers-Daller

Received: 16 November 2022

Revised: 2 March 2023

Accepted: 13 March 2023

Published: 31 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The focus of this paper is on filler particles (FPs) such as *uh* and *um* and tongue clicks, phenomena that are typical for, and often observed in, spontaneous speech. Frequently, though not exclusively, FPs occur in the vicinity of speech pauses. The distribution of FPs connected to pauses and stretches of speech is investigated here, as well as the question of whether the different FP types behave differently. A large corpus of German spontaneous speech was analysed with respect to the frequency of occurrence of various FP types, along with their pause context, duration, voice quality, and vowel quality.

Pauses are a vital part of speech (Trouvain and Werner 2022); speech breathing physically requires the interruption of speech, as usually only the egressive pulmonic airstream is used to produce speech (Clark et al. 2007). In addition, pauses function as structuring devices (Oliveira 2002; Swerts 1998), that is, as intentional rhetorical devices to emphasise parts of speech (O'Connell and Kowal 2005, 2008) or as a hesitation strategy allowing the speaker to structure their thoughts and utterances (Clark and Fox Tree 2002; Goldman-Eisler 1972). What is considered as a pause, however, is not as easily described as this term suggests. From a lay point of view, a pause suggests that some kind of action is suspended for a certain amount of time only for this same action to be resumed again later. Therefore, a speech pause is an interruption of speech that lasts for a certain amount of time. The *interruption of speech* can be viewed from at least two different angles: (1) the propositional message can be interrupted and resumed after a certain amount of time (in line with (Clark and Fox Tree 2002)'s idea of a primary signal that is interrupted by collateral signals such as disfluencies), or (2) the speech production activity, and as such the acoustic signal that originates from the speaker, can be interrupted (see (Belz 2021)). While it is true that the articulatory movement is suspended when the propositional message is paused, the

inverse is not necessarily the case. The propositional message can be suspended while the articulatory movement continues, e.g., producing repetitions or (non-)lexical filler particles (*like, you know, uh/um*). While one view is not superior to the other, it should be made very clear how a pause is defined whenever the concept is relevant. For this study, we adopt the second view of pauses, i.e., that a pause is an interruption of the articulatory movement.

FPs are often described as *filled pauses*, which can be considered an oxymoron in the sense that the interruption of speech is filled with speech material (Belz 2021). When considering the first description of a pause, the term *filled pause* makes more sense; the pragmatic message is interrupted, and this interruption is filled with speech material that is not part of the same message. For consistency with our adopted description of a pause, and in line with other scholars (Belz 2021; Fuchs and Rochet-Capellan 2021; Trouvain and Werner 2022), we use the term *filler particle* for such phenomena as *uh* and *um*, and avoid the term *filled pause* altogether. Typical FPs can occur within two stretches of speech production silence,<sup>1</sup> though they are not limited to this position and frequently occur within a speech utterance with no silence on either side. Another phenomenon that serves a similar function as *uh* and *um* and occurs in similar positions as these typical FPs is the purely nasal FP type, *hm*. Less often described are glottal FPs and tongue clicks (however, see (Smith and Clark 1993)). We consider tongue clicks as *potential* FPs that can be used for self-repair (Li 2020) or when having trouble finding a word (Ogden 2020; Trouvain and Malisz 2016) as situations that are typical when using FPs.

Research on pauses, FPs, and disfluencies in general has become more frequent in the last decades, and most work has focused on the frequency distribution, duration, and vowel quality of FPs. However, most of these studies focus on English data, and research on German data mostly focuses on one or two aspects, such as the frequency distribution (Bellinghausen et al. 2019; Braun and Rosin 2015), fundamental frequency and duration (Batliner et al. 1995), vowel quality (Pätzold and Simpson 1995), vowel quality and fundamental frequency (Klug and König 2012), or vowel quality, duration, and frequency (Niebuhr and Fischer 2019). In a few cross-language comparisons, German has been investigated alongside other European languages, such as English, Dutch, and French (de Leeuw 2007; Gerstenberg et al. 2018; Lo 2020; Muhlack 2020). Other studies have looked at FPs in first vs. second languages (L1 vs. L2), e.g., in L2 German by speakers with several different L1 backgrounds (Belz and Klapi 2013; Belz et al. 2017; Muhlack 2020; Reitbrecht 2017). A comprehensive phonetic analysis of FPs in German was recently provided by Belz (2017, 2018, 2021); (Belz and Reichel 2015; Belz and Trouvain 2019) investigated several features of FPs (especially *uh* and *um*), among which were the voice quality of FPs and the occurrence of glottal FPs. The study reported here follows Belz (2021) in describing the phonetic characteristics of FPs in German by examining different features such as their frequency distribution and duration, along with the pause context, voice quality, and vowel quality uttered by 100 male native German speakers.<sup>2</sup> The features of pause context and voice quality are under-researched aspects of FPs which is why they are included here. Benefits from this study include the number of speakers examined, the inclusion of a Lombard condition, and the number of features investigated, both for their general trends and on an individual speaker level. Because data from 100 speakers poses problems for qualitative analyses, speaker specificity is examined by means of comparing the individual patterns of twelve sample speakers. In the following, we provide a brief non-exhaustive overview of the literature on FP research in different languages.

### 1.1. Frequency Distribution

Several previous studies have reported disfluency rates per minute, however, these are mostly not directly comparable, as different phenomena were considered. Braun and Rosin (2015) reported a disfluency rate of 4.5–12.3 per minute (for 10 speakers) when looking at typical FPs (*uh, um*), the nasal FP *hm*, and initial and final vowel and consonant lengthenings in German. Belz (2021) reported an FP rate of 2.9 FPs per minute (range: 1.4–4 disfl./min) for the GECO-FP corpus (Belz 2019) and a rate of 4.3 FPs per minute

(range: 1.9–11.3 disfl./min) for the BeDiaCo (Belz and Mooshammer 2020), both of which are German dialogue corpora. For English, Clark and Fox Tree (2002) reported an FP rate (*uh* and *um* only) of 17.3 per 1000 words ranging from 1.2 to 88.5 FPs per 1000 words for 65 speakers. This rate translates to 2.6 FPs per minute when assuming an average of 150 words<sup>3</sup> produced per minute. Considering FPs, pauses, repetitions, and false starts, Maclay and Osgood (1959) found a mean disfluency rate of 10.97 per 100 words (16.5 disfl./min) ranging from 5 to 15 disfluencies per 100 words for thirteen different speakers of English. Shriberg (1994) reported a disfluency rate of 0.01–0.08 disfluencies per word (1.5–12 disfl./min) for three English corpora, including FPs as well as repetitions, false starts, and repairs. McDougall and Duckworth (2017) reported an FP rate for English (including only *uh* and *um*) ranging from approximately 2–8 FPs per 100 syllables (appr. 5–20 disfl./min).<sup>4</sup> An overview of these studies shows the large variation in the or consideration of disfluencies for each study, and that there is no standard unit in which the rate is reported. Time units can be stated in minutes, as the rate per word, rate per 100 words (or syllables), or even rate per 1000 words. It should be noted that a disfluency rate per word/s may not be the most useful unit, as word lengths may differ considerably (especially in German, where compounding is very frequent), as well as because differences in word length between languages makes cross-language comparison difficult (Trouvain 2004). A rate per, e.g., 100 syllables (or even phones) may be more useful, although the syllable structure and complexity may pose problems for cross-language comparisons in this case as well.

Here, we consider tongue clicks as potential FPs (Belz 2023), as they frequently occur in pauses and their function as a hesitation device has been reported previously in Trouvain and Malisz (2016). We see a large variation between studies and individuals in the use of tongue clicks. A rate of 1.3 clicks per minute was reported for English dialogues in Ogden (2013); however, a high variation between speakers was observed. Trouvain and Malisz (2016) found a click rate of 6–12 per minute for one native English speaker who was regarded as a heavy clicker. However, Zellers (2022) found a rate ranging between 1–5.4 clicks per minute for twelve Swedish speakers. It seems that speakers vary in their clicking behaviour; however, Gold et al. (2013) argued that speakers of English do not vary sufficiently in terms of their click distribution, and that audio material available in forensic cases may be too short for the click frequency to be a useful feature.

### 1.2. Duration of Filler Particles

When looking at the duration of FPs, usually only the FPs *uh*, *um*, and *hm* are considered. Belz (2021) reported the following values for these FPs in German data: a mean of 262 ms (sd = 121 ms) for *uh*, 396 ms (sd = 140 ms) for *um*, and 450 ms (sd = 183 ms) for *hm*. The same duration pattern (*uh* shortest, *hm* longest) was reported by de Leeuw (2007) for German (though not for English and Dutch): 317 ms (sd = 113 ms) for *uh*, 457 ms (sd = 161 ms) for *um*, and 470 ms (sd = 234 ms) for *hm*. It is often reported that the vowel in vocalic-nasal FP types is shorter than in purely vocalic FPs (Belz 2021; Hughes et al. 2016), and that the former type is more often surrounded by silences (Clark and Fox Tree 2002; Hughes et al. 2016). These silences tend to be longer for *um* than for *uh* (Clark and Fox Tree 2002).

### 1.3. Vowel Quality

The vowel quality of FPs is generally considered to be a central vowel in most languages, though with language-specific tendencies. If not reported as a central schwa [ə], the following vowels are used to describe the vowel quality in a number of languages: The front vowel [ɛ] or central vowel [ɐ] for German, the rounded front vowels [ø] or [œ] for French, and the mid-closed front vowel [e] for Spanish (Belz 2021; Candea et al. 2008; Künzel 1987; Simpson 2007). English, especially American English, is typically considered to use a vowel similar to the open-mid vowel [ʌ] in FPs (Shriberg 1994), whereas it is important to note that in phonetic transcription of English the symbol ʌ is usually used to describe an open-mid central vowel (Roach 2009). German FP vowels reportedly have a

lower F1 than English and French FP vowels (Lo 2020; Muhlack 2020). French FP vowels are usually produced with more lip rounding than the German FPs (Lo 2020).

#### 1.4. Voice Quality

A phenomenon that has been less researched in the disfluency area is the voice quality with which FPs are produced. Belz (2017, 2021) introduced glottal FPs into the field, as well as the proportion of creaky voice and glottal plosives within each FP. It has been reported that in Italian tourist guide speech about two-thirds of all FPs are realised with creaky phonation (Cataldo et al. 2019). Shriberg (2001) remarks that FPs may be subject to a decrease in amplitude and a drop in pitch, which supports the production of creaky voice in the final FP position. However, Belz (2021) found that FP initial creaky voice is more frequent than FP final creaky voice, and is more frequent with the vocalic FP *uh* than the vocalic-nasal FP *um*.

#### 1.5. Hypotheses

Through this study, we hope to shed light on the frequency distribution, duration of FPs, and their vowel quality and voice quality in connection with Lombard speech, pause context, and speech tempo. We apply an exploratory analysis, as many aspects of our analysis are under-researched, e.g., disfluencies in the Lombard condition, the production of creakiness during FPs, and the influence of pause context on the acoustic measures of FPs. The literature guides us in a number of aspects, as we expect to find more (and longer) vocalic-nasal FPs in the data than vocalic FPs (Hughes et al. 2016; Wieling et al. 2016). As Gósy and Silber-Varod (2021) found an effect of pause context on the duration of vocalic FPs, we expect to see similar trends here, in that FPs that are surrounded by pauses should be longer than FPs that occur within speech. We do not expect to find a difference in the vowel quality between *uh* and *um*, or any influence of FP duration on vowel quality (Hughes et al. 2016). In line with Belz (2021) and Cataldo et al. (2019), we anticipate observing a high rate of creaky voice within FPs, moreso in the FP initial position than in FP final position.

#### 1.6. Importance for Forensic Phonetics

FPs are frequent in spontaneous speech (Bortfeld et al. 2001; Fox Tree 1995), which makes them interesting for forensic phonetic casework. If they were considered words, as suggested by Clark and Fox Tree (2002), they would probably be grouped under high-frequency words. It is usually assumed that speakers do not produce them intentionally (Braun and Rosin 2015; Kjellmer 2003). The lack of control that speakers usually exert over FPs makes them a practical feature for forensic phonetics, as they are considered to be unaffected by voice disguises (Butterworth 1975; Jessen 2012; Künzel 1987). In forensic phonetic casework, a suspect's voice recording is often compared to a recording of a questioned speaker (forensic voice comparison). The task of the phonetic expert is to make an assessment as to whether the observed linguistic/phonetic features are more likely under the hypothesis of speaker identity or non-identity and how much more likely this is in one or the other direction. FPs and other disfluencies may represent a type of speaker characteristics that can be included in voice comparison analysis. Specific features to be considered might include the general frequency distribution of FPs, their type and duration, the pause context, i.e., whether they occur within a pause or within speech, the proportion of creaky voice produced during the FPs, and the vowel quality of the vowel produced in FP types such as *uh* and *um*. In order for these features to be applied in forensic phonetic casework, it is crucial that the distribution of the features in the relevant population is known to the expert in order to make an assessment about typicality and similarity between the compared recordings (Rose 2002). For example, if a particular feature instantiation (e.g., a certain value of *uh* per minute) is rare in the population, and the similarity in the recordings is high, the strength of evidence in the direction of speaker identity is higher than if the feature instantiation is quite common, even given the same similarity. For all

features that may prove to be useful in forensic casework, the general distribution and variation within the population (between-speaker variation) must to be known, as must the within-speaker consistency.

In this paper, we do not provide a full documentation of between-speaker variation (e.g., histograms about how many speakers show which feature instantiation) for each of the many disfluency features investigated here, nor do we document within-speaker variation to the fullest level of detail. We are primarily focused on mean patterns in two conditions (normal and Lombard), which allow the reader to see the main patterns as concerns what is typical across speakers and what is variable speaker-internally across the studied conditions. In addition, we provide a more in-depth analysis of within- and between-speaker variation in a subset of the speakers investigated here (Section 6). The choice of a Lombard condition is forensically relevant as well, because variations in vocal effort are common in forensic casework.

### 1.7. Outline of the Paper

The outline of the rest of this paper is as follows: Section 2 provides an overview of the phenomena under investigation; Section 3 introduces the corpus (Section 3.1), the annotation scheme used for the data (Section 3.2), the measure of speaking tempo (Section 3.3), and statistical methods (Section 3.4). The results are divided into three parts. We first show general results from the entire dataset in Section 4, then provide separate analyses of the normal and Lombard conditions and compare the results (Section 5), then follow with speaker-specific patterns (Section 6). Each results section includes subsections on the investigated features, namely, the frequency distribution, duration of FPs, pause context, voice quality, and vowel quality in *uh* and *um*. Each section concludes with a discussion, and we finish with a general discussion in Section 7.

## 2. Filler Particle Phenomena

The phenomena under investigation here are different types of FPs, as well as tongue clicks. Both phenomena occur frequently in spontaneous speech (Bortfeld et al. 2001; Fox Tree 1995; Ogden 2013). FPs are often grouped under the umbrella terms of *disfluencies*, *hesitations*, or *discourse markers*. All these terms suggest a specific function for FPs, namely, the indication of production problems for the former two terms and a structuring function for the latter. Determining the function of FPs can be difficult, as they may serve one function or several functions that seem to overlap (Belz 2021). FPs are often said to serve functions such as signalling the search for a word, serving as an editing phase when repairing a speech error (“show flights from Boston-uh-Denver” (Shriberg 1994)), holding or ceding the floor, expressing uncertainty, or securing attention (Clark and Fox Tree 2002; Goodwin 1981; Maclay and Osgood 1959; Shriberg 1994). As the function of an FP is difficult to determine and depends heavily on the conversational situation or speech task, we do not take the different possible functions into account here. However, due to the task of the corpus, it is likely that the FPs under consideration are mainly due to repairing speech errors, searching for the right word, or conceptualising the next sentence, and rather less due to turn-taking issues.

### 2.1. Typical Filler Particles (*uh* and *um*)

By the term “filler particle”, we mean non-lexical speech material that does not add to the propositional message. Clark and Fox Tree (2002) have attributed this additional speech material to a signal of the collateral track. While the message itself belongs to the primary track, the collateral track “refer[s] to the performance itself—to timing, delays, rephrasings, mistakes, repairs, intentions to speak, and the like” (Clark and Fox Tree 2002). Typical FPs that occur in many languages consist of a vowel or a combination of a vowel and a nasal consonant. Their orthographic representation in English is often provided as *uh* or *um*. The German counterparts are more often transcribed as *äh* and *ähm*. For better readability,

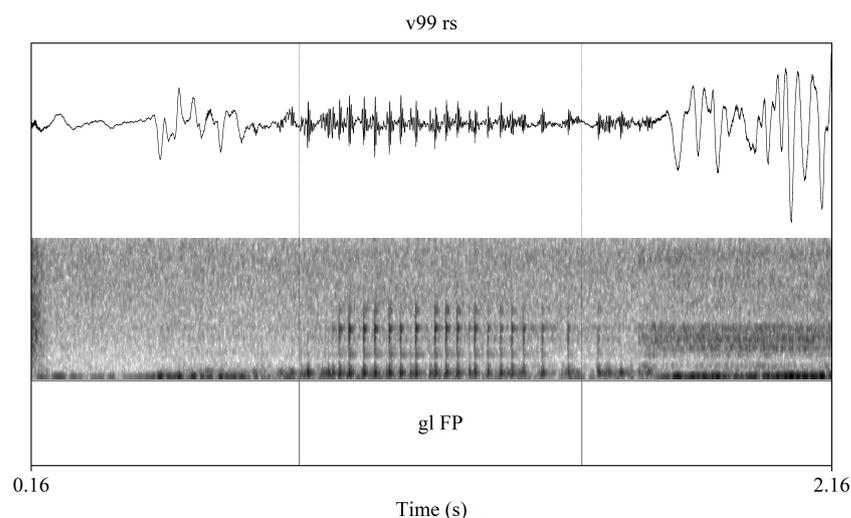
we use the English orthographic forms in this paper, even though we are reporting on German data.

## 2.2. Nasal FP *hm*

Another form of FP is the purely nasal variant *hm*, which occurs in both English and German, though to a lesser extent than the previously described FPs (de Leeuw 2007). The phonetic form of this FP does not necessarily match the orthographic form, as two components are not usually observed in this FP; instead it includes only one, namely, a nasal consonant. The orthographic form *hm* (along with *mh*) is often used for other phenomena as well, such as feedback utterances or discourse particles; for instance, functions such as a reaction signal (“What did you say?”), turn holding (“Let me think.”), completion signal (“Done!”), and expressing appraisal (“This tastes/smells good!”) have been described (Pistor 2017). These discourse particles vary mainly in their intonational contour and duration. Note that the turn holding signal may not be identical to a hesitation, as the former is used intentionally while FPs are usually produced more subconsciously (Kjellmer 2003). In line with Schmidt (2001), we consider *hm* as a consonant with a closed mouth, and with only the intonation as the carrier of phonetic information.

## 2.3. Glottal Filler Particle

A special and often overlooked form of FP is the glottal FP, previously described in Belz (2021). This type stands out due to its specific voice quality, that is, creaky voice instead of modal voice, to the extent that it only consists of a series of glottal pulses with no modal voice parts at all (see Figure 1). Belz (2017) describes this phenomenon as “glottal pulses and creak phonation without coarticulated vowels that seem to be used in a similar way to other FPs.” As observed in the Pool2010 corpus used for this study, glottal FPs seem to be produced with both an open and a closed mouth. They may be understood as variants of *uh* and *hm* produced entirely with a creaky voice quality. In a study with seven female participants, Belz (2017) found that approximately 20% of FPs were glottal FPs and that these did not differ in duration from the vocalic FP *uh*. In a different study with twelve female and twelve male participants, Belz (2021) found that about 5% of FPs were glottal FPs, and that females produce them more often (~7%) compared to males (~3%).



**Figure 1.** Section showing a 2 s selection (spectrogram: 0–8 kHz) from the Pool2010 corpus (speaker: v99 in Lombard condition) showing a glottal filler particle (gl FP).

## 2.4. Tongue Clicks

Tongue clicks are produced by creating a small pocket of air with the tongue against the alveolar ridge. By moving the tongue downwards, the air pocket is enlarged and “the pressure drop in the trapped air generates a short but quite strong inflow of air as

the closure is released" (Clark et al. 2007), which results in the production of a click. In certain languages, clicks are included as phonemes in the sound inventory; however, clicks can occur in various languages as part of the non-linguistic message as well. As Belz (2023) points out, these discourse clicks "are not prototypically used interchangeably with filler particles, but presumably serve different functions in dialogue." Though Belz (2023) regarded clicks as 'candidates of filler particles', they were ignored in the rest of his study. In contrast to Belz (2023), we take clicks into account here in order to attain an idea of the relative frequency of this phenomenon. However, we consider clicks only as *potential* FPs, and do not assign any specific function to single instances of clicks. They can serve functions such as turn-claiming, displaying a stance such as disapproval (e.g., tutting), or signalling difficulty during word search (Ogden 2013). These types of clicks, especially when displaying stance, are used as intentional messages from the sender (Ogden 2013). Closely related to clicks are percussives. In contrast to clicks, these are not produced deliberately, and are a "byproduct" of articulation; they often occur during the preparation phase of speech when the articulators are being separated (Ogden 2013). As function is excluded from the following analyses, we do not make a distinction between clicks and percussives here, and group all percussives under the class of tongue clicks.

### 3. Materials

#### 3.1. Corpus

The data used here are part of the Pool2010 Corpus<sup>5</sup> that was compiled in 2001 by the German Federal Criminal Police Office (Bundeskriminalamt, Wiesbaden) during a research project (Jessen et al. 2005). A total of 107 male native German speakers, most of them employees from this office, were recorded during several tasks. Seven speakers were excluded due to technical issues or problems with a voice disorder on the part of the speaker. While several speech tasks were recorded for this corpus, the present analysis only uses the semi-spontaneous speech task. This speech task was similar to the "taboo" game, in which the speaker has to describe a number of terms in their own words without using two or three "taboo" words.<sup>6</sup> Each speaker described seven words in the mean per condition (range: 2–11 words). Because the number of taboo words may have an influence on the difficulty of the task, each speaker must describe several words with a differing number of taboo words in order to balance the influence of difficulty between speakers. A female interviewer served as the interlocutor for the guessing game, providing the answers when the speaker's information was sufficient. The speakers did not know that she was a confederate and knew the words previously. She extended the speaker's description to a certain degree by not providing the correct word immediately.

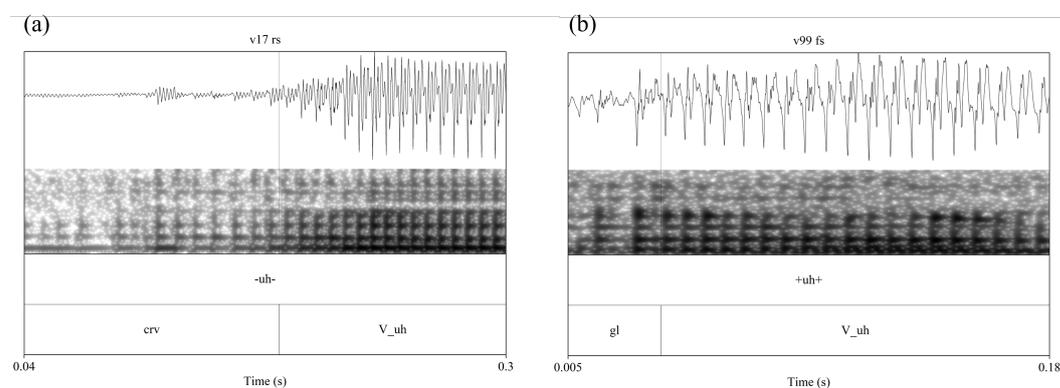
The speakers completed the task in two conditions, one in the normal speech condition and one in the Lombard speech condition; during the latter condition, the speakers heard white noise (80 dB<sub>SPL</sub>) over headphones, leading them to produce to louder speech. The order of the two conditions was changed from subject to subject in order to prevent potential serial order effects. The Lombard effect describes the phenomenon of a noisy environment that causes speakers to increase their level of vocal effort, which results in louder speech with a higher fundamental frequency (Lombard 1911). In addition, the presentation of white noise leads to an impaired feedback loop during speech production. Usually, a speaker hears their own speech and monitors it for possible errors, e.g., slips of the tongue. It is possible that this shortcoming has an impact on the production of disfluencies as well as on FPs. It has been shown that fluency increases in people who stutter when the subjects are presented with noise over headphones (Adams and Hutchinson 1974). Furthermore, the level of noise is negatively correlated with the number of disfluencies.

A microphone was attached to a helmet that included the headphones for the Lombard condition in order to maintain a constant distance to the microphone allow for a quick and easy transition between conditions. The recordings were made with a sampling rate of 16 kHz and a sampling depth of 16 bit. The resulting audio files had a mean duration of

3:52 min (sd = 1:03 min), with a range of 1:43 to 7:53 min. The part of the corpus used for this study consisted of 12h and 56 min when including both conditions together in total.

### 3.2. Annotations

Annotations were produced following Praat (Boersma and Weenink 2022). The files have been re-annotated for this project, including annotation of FPs and their context, i.e., whether they were preceded or followed by a silent section, which we call a pause. A detailed segmentation of FPs into vowels, nasal consonants, and creaky voice parts was included, along with annotation of glottalised FPs and tongue clicks. The entire FP was annotated on one interval tier, while segmentation of the FPs into vowel, nasal, and creaky voice or glottal pulse portions was annotated on another interval tier. To investigate the production of creaky voice and glottal pulse portions during the FPs, these portions were marked during annotation. Creaky voice (crv) was annotated when it was perceived auditorily during the FP; the spectrogram and the pitch contour served as visual aids in determining the beginning and end of these phases. The distinction between creaky voice and glottal pulse portions was based on Belz (2021). Creaky voice (crv) was deemed to occur when more than three glottal pulses occurred within an FP, while when the number of glottal plosives was three or less it was annotated as a glottal pulse (gl) (see Figure 2). Glottal pulses are only audible when listening to a short selection of the FP, not when listening to the FP in its full context. When glottal activity was detected in isolation, i.e., not in other lexical material, it was marked as a glottal FP, denoted gl FP (see Figure 1). Tongue clicks were annotated, though they were not detailed on the time dimension as the tokens are very short (Trouvain and Malisz 2016). Therefore, we did not take duration measurements of tongue clicks into account.

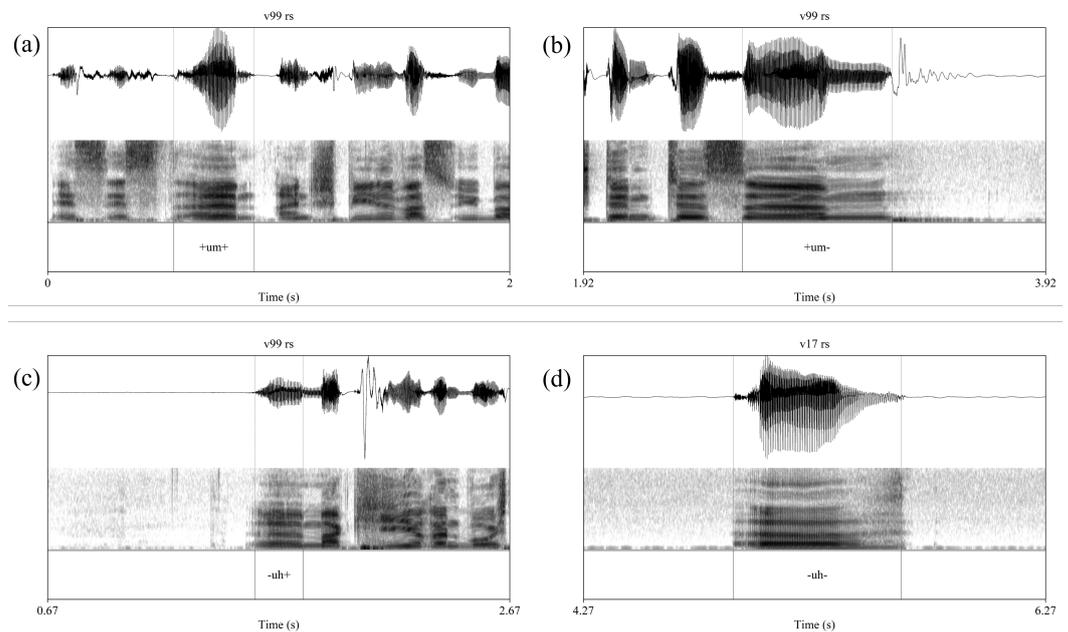


**Figure 2.** Sections (spectrogram: 0–8 kHz) from the Pool2010 corpus (speakers: v99, v17) showing filler particles (a) with initial creaky voice (crv) and (b) with two initial glottal pulses (gl). Note that in 2a only the first 260 ms of the FP are shown.

The FPs *uh*, *um*, and *hm* were labelled according to their pause context, with the original label receiving a plus or a minus before and after the label (e.g., *+uh-*). A minus sign denotes that a pause, i.e., a silent phase, occurs on the one side of the FP, while a plus sign denotes that the FP is directly connected to speech and no perceived pause was detected (see Figure 3). We did not apply a minimum (or maximum) pause threshold when annotating pauses (Campiono and Véronis 2002). When the first segment after a pause was a plosive, the initial boundary of the segment was moved 50 ms to the left to account for the closure phase of the stop (Belz and Trouvain 2019).

We labelled three different types of pauses, namely, before and after *hm* and the typical FPs *uh* and *um*. Due to the task, both simple perceived pauses (without a minimum threshold) and pauses occurring when the participant was clearly waiting for a response from the interviewer (guessing the correct word) were specifically labelled (“waiting pause” = *p\_w*). Speech pauses that occurred between tasks, i.e., when one word was successfully

explained and the participant was moving on to the next word, were marked as task change (tc).



**Figure 3.** Sections of 2 s duration (spectrogram: 0–8 kHz) from the Pool2010 corpus (speakers: v99, v17) showing FPs with (a) speech as left and right context (+FP+), (b) speech as left and silence as right context (+FP−), (c) silence as left and speech as right context (−FP+), and (d) silence as left and right context (−FP−).

The speech interval between pauses is referred to as inter-pausal unit (IPU). As syntactic information was not annotated, we cannot be certain that complete phrases were produced and that every pause acts as a phrase boundary. However, it is assumed that the waiting pauses and the task changes mark a phrase boundary, either because a response from the interviewer is expected or because a new task begins.

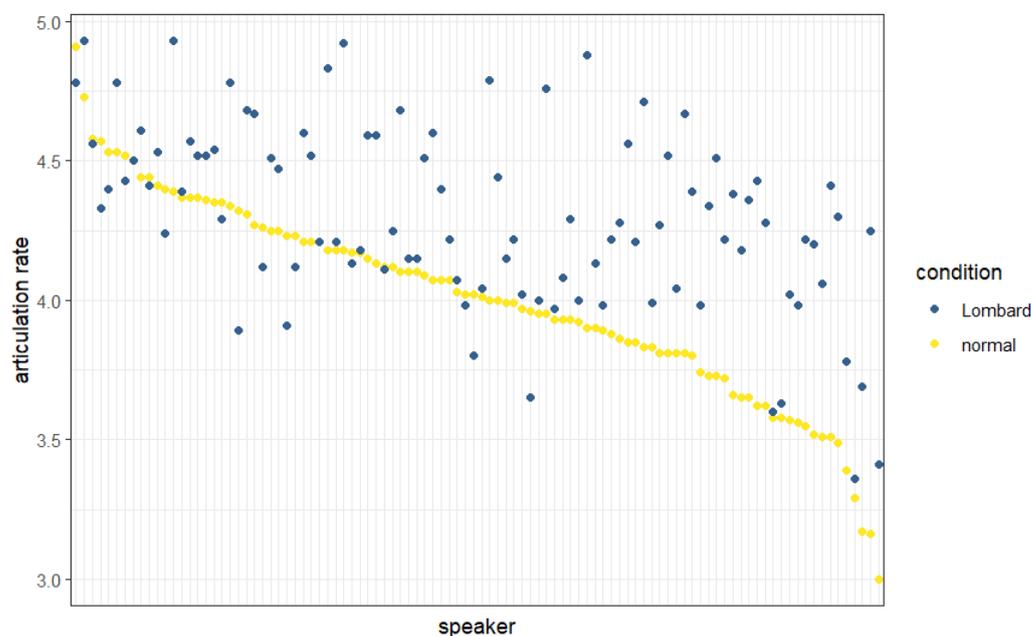
The corner vowels [a:], [i:], and [u:] of each speaker were annotated in selected syllables carrying the lexical stress where they would appear in Standard German. The vowels were measured in the same way as reported for *uh* and *um*, and observations within three standard deviations of the mean of each vowel and formant were kept in the dataset. The aim was to collect ten tokens for each vowel and condition per speaker; this aim, however, could not be reached for several speakers, as the recordings did not provide enough tokens. This was especially the case concerning the close rounded back vowel, which is under-represented in the dataset of lexical vowels (token numbers after reduction: *i* = 1214, *a* = 1460, *u* = 514).

### 3.3. Speaking Tempo

We measured the speech tempo of each file with the help of the Praat script “Praat Script Syllable Nuclei v2” created by [de Jong and Wempe \(2009\)](#). As the data include long pauses (p\_w, tc) that influence the speech rate, we excluded these pause types using a Praat script and used the resulting new (shorter) files for the speaking tempo measurements. Simple pauses (p) were kept intact. The script detects intensity peaks that are surrounded by intensity dips, which are then considered as syllable nuclei. Three measures of the script are adjustable: the silence threshold, the minimum dip between peaks, and the minimum pause duration. To find the optimal settings for the script, we manually annotated ten files of the corpus, marking the phonetic syllables and the intervals where speech occurs in line with the output TextGrids that the script produces. This allowed us to determine the values for syllable number per file and the total speaking time, which we then used to test the settings of the script. The settings that resulted in the minimum deviance from the

manual values on the ten files were as follows: a silence threshold of  $-20$  dB, a minimum dip of 2, and a minimum pause duration of 500 ms. The mean deviance from the manual values took the form of over-counting of syllables by 13.8 and misdetection of speaking time by 7.2 s per file. The articulation rate and speech rate were determined for every file using the optimal settings of the script. Two files resulted in particularly low values for these measures. On further inspection, these files contained and loud bursts and laughs; after these were deleted, the performance of the script improved with the above-mentioned settings. While it is possible that the script performed better on certain files than on others, we assume that these values are a good approximation of the true speaking tempo and a time-efficient alternative to manual annotation.

Figure 4 shows that the articulation rate for most speakers in the Lombard condition is actually faster than in the normal condition, contrary to previous findings (Tuomainen et al. 2021), although Tuomainen et al. (2021) raise the question of whether the differences in articulation rate are perceptually salient (cf. Quené 2007). Deviations from the articulation rate reported in Jessen (2007) on the same corpus can be explained by our use of the script. The authors of the script report that automatic values are generally lower than manually obtained values, and need to be multiplied by 1.28 to predict the manual values (de Jong and Wempe 2009). This is due to the failure of the script to detect certain unstressed syllables; thus, fewer syllables are counted, leading to a lower articulation rate. Jessen (2007) reported a mean articulation rate of 5.21 syll/s in the normal speech condition, while the automatically obtained mean value for the same data is 4.0 syll/s. When using the factor reported above to predict the manual mean values from our automatically obtained mean, we found that this conversion approximated the manual value quite well ( $4.0 \times 1.28 = 5.12$ ).



**Figure 4.** Articulation rate (syll/s) per speaker as a function of condition.

### 3.4. Statistical Methods

Analyses and plots were created using R (R Core Team 2022) (version 4.1.3) and the tidyverse package (Wickham et al. 2019). Linear mixed models were created using the lme4 package (Bates et al. 2015) with FP duration for F1 and F2, and frequency as dependent variables and FP type, condition, pause context, FP duration for F1 and F2, and speaking tempo measures as independent variables. The speaker was included as a random effect, allowing the intercept (though not the slope) to vary between subjects; The lmerTest package (Kuznetsova et al. 2017) was used to obtain  $p$ -values. Models were built by including the condition and FP type with the interaction term and adding the other

variables as control factors without the interaction term. We are aware that when building linear mixed models one should aim for maximal models; however, when including all the interaction terms of numerous factors the results would become difficult to interpret and the model would not serve the research question. As the speech rate and articulation rate are co-dependent variables, only one of these measures was included. The articulation rate was chosen, as it is independent from any pause rate and is considered to be more independent from the disfluency rate than the speech rate measure. In addition, the difference between the speech conditions is larger for the articulation rate than for speech rate.

Furthermore, a Pillai score was calculated as a measure of overlap between vowels (Kelley and Tucker 2020). This was determined by conducting a multivariate analysis of variance (MANOVA) using the first and the second formants as response variables and the FP type as the predictor variable (using the manova-function from the stats package (R Core Team 2022)).

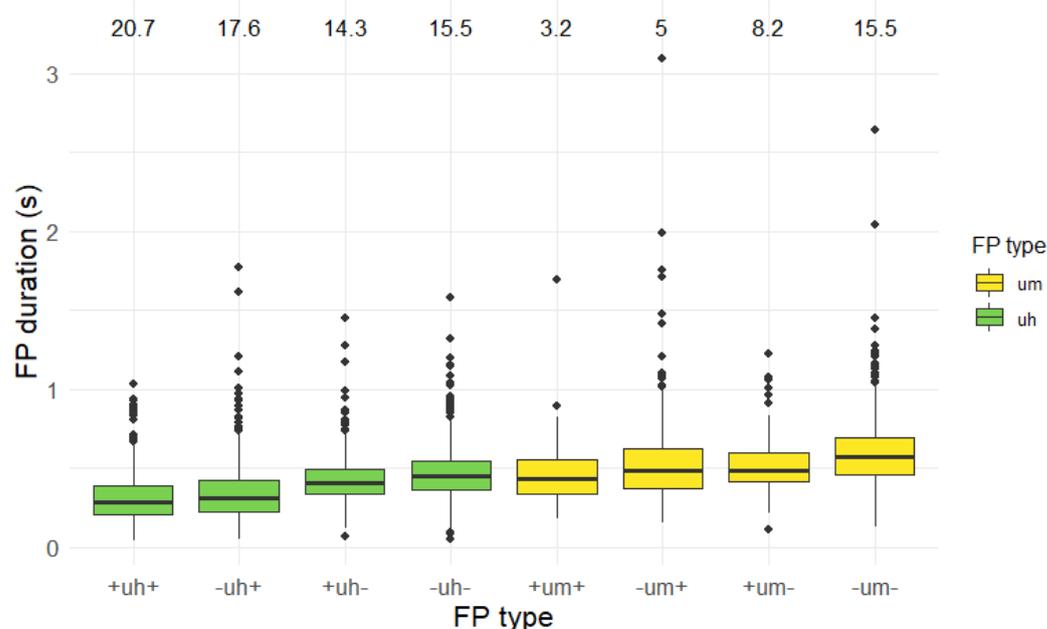
#### 4. General Results

In the following, a general overview of the FPs in the Pool2010 corpus is provided for the normal condition and the Lombard condition. A comparison of normal speech to Lombard speech follows in Section 5. Section 6 focuses on between-speaker variation and within-speaker consistency. The Lombard speech condition is regarded as a second speech mode in order to compare two files of the same speaker.

##### 4.1. Frequency Distribution

We detected 6734 FPs have in the entire corpus of material (12 h 56 min), resulting in a ratio of 8.67 FPs per minute. When looking at the rate of FPs per type (Table 1), it becomes clear that tongue clicks are the most frequent, closely followed by the vocalic FPs. The vocalic-nasal FP type *um* is only half as frequent as the vocalic type *uh*. The number of *um* instances is as high as the number of glottal FPs and the nasal type *hm* taken together.

Considering only the FPs *uh* and *um*, as in Figure 5, along with their pause context, it is apparent that the *um* FPs amount to only around one-third of all typical FPs. The most frequent FP type in pause contexts is *uh* surrounded by speech, i.e., not within silence.



**Figure 5.** Duration of the FPs *uh* and *um* and their pause contexts in seconds (s). Context types are described using a (+) to denote speech and a (−) to denote a silent phase surrounding the FP. The values at the top refer to the percentage of each displayed FP type for all *uh* and *um* FPs.

**Table 1.** Absolute number of the FPs (*uh*, *um*, *hm*, glottal FPs, and tongue clicks), with the mean (sd) durations (in ms) of the phenomena and the vowel duration of *uh* and *um* FPs. NA (not applicable) means that the duration was not measured (e.g., for clicks) or that the phenomenon did not include a vowel. Creaky voice portions are included in the total duration and vowel duration.

FP Type	Absolute	Rate: FPs/min	Duration Mean (sd)	Vowel Duration Mean (sd)
<i>uh</i>	2250	2.9	382 (180)	382 (180)
<i>um</i>	1054	1.4	559 (234)	281 (125)
<i>hm</i>	314	0.4	442 (224)	NA
glottal FP	757	1.0	244 (332)	NA
clicks	2359	3.0	NA	NA

#### 4.2. Duration

The duration of FPs is highly variable, as can be seen by the high standard deviation values in Table 1 for each phenomenon. Glottal FPs have the shortest duration (and the highest standard deviation), followed by the vocalic type (*uh*), the nasal type (*hm*), and the vocalic-nasal type (*um*), with the last being the longest. The vowels in *uh* are longer than in *um*.

Considering only the FPs *uh* and *um* in their pause context (see Figure 5), it becomes apparent that vocalic-nasal types are longer than vocalic types and that there seems to be a duration hierarchy depending on the pause context. FPs in speech (+FP+) are shorter than FPs surrounded by pauses (−FP−). Moreover, IPU-final FPs (+FP−) are longer than IPU-initial FPs (−FP+). This pattern applies to the vocalic type *uh* as well as to the vocalic-nasal type *um*.

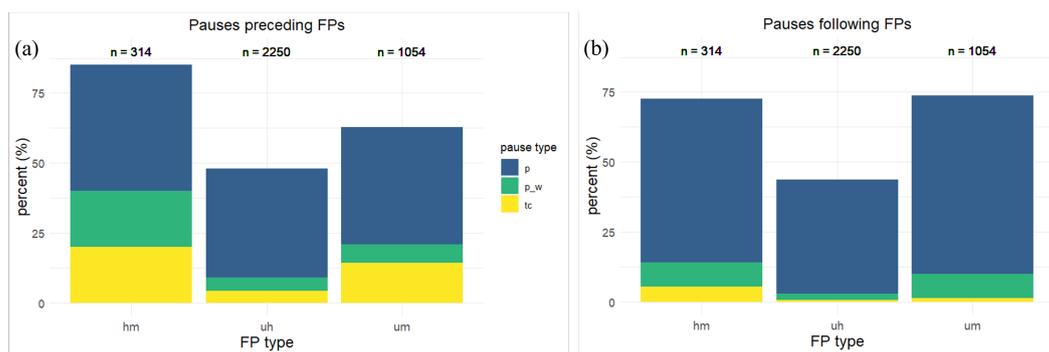
#### 4.3. Pause Context

Figure 6 shows how many FPs are preceded and followed by a pause, as well as the type of pause. Simple pauses are clearly more frequent than the other pause types, and pauses in general are more frequent surrounding the nasal and the vocalic-nasal FP types than surrounding the vocalic type. Additionally, the pause type “tc” is more frequent preceding an FP than following an FP, which means that a task is started with an FP more often than it is closed with an FP. This observation is not surprising; it has been reported that turn-initial FPs are more frequent than turn-final ones in other corpora (O’Connell and Kowal 2005; Swerts 1998).

The duration of the pauses surrounding an FP varies to a high degree, as indicated again by the high standard deviation values (Table 2). A linear model including pause type (simple, waiting, task change) and pause position (pre-FP, post-FP) as predictors for pause duration indicates that the three different pause types (simple, waiting, task change) are significantly different from one another (Table A1). Furthermore, while simple pauses (p) and task changes (tc) do not differ in their duration depending on pause position, waiting pauses (p\_w) do differ significantly in their duration. These pauses are longer before an FP than after an FP.

**Table 2.** Duration (in ms) of different pause types pooled over both speech conditions (normal, Lombard).

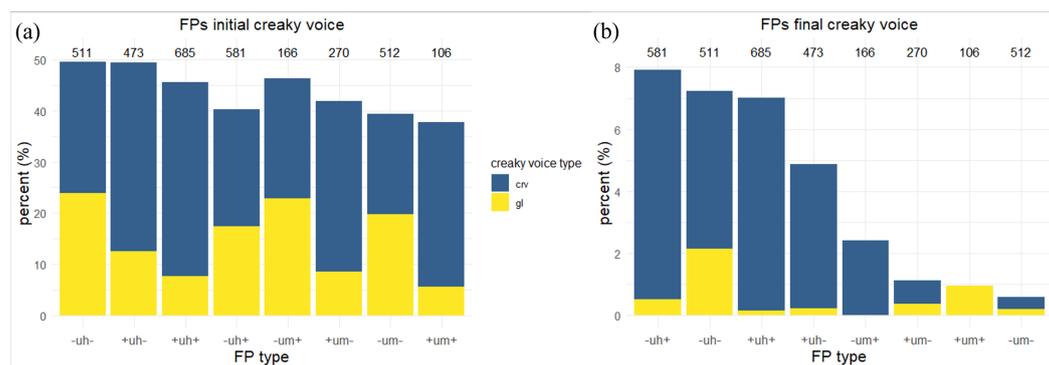
Pause Type	Pre FP	Post FP
simple pause (p)	1177 (1222)	1083 (1227)
waiting pause (p_w)	3182 (2302)	2285 (1563)
task change (tc)	3962 (2706)	3560 (2283)



**Figure 6.** Percentage of different FP types (a) preceded and (b) followed by a pause. The colours show different types of pauses: simple pause (p), waiting pause (p\_w), and task change (tc). The values at the top show the values representing 100% for each FP type.

#### 4.4. Voice Quality

A large percentage of FPs are produced with initial creaky voice. Nearly 46% of *uh* FPs include initial creaky-voiced portions or glottal pulses, while 41% percent of *um* FPs include these sections. Less than 7% of *uh* FPs and only 1.14% of *um* FPs include final creaky-voiced portions or glottal pulses. Of all the *uh* and *um* (5.7%) tokens, 189 are produced with 100% creaky voice; these are included under the initial creaky voice category.<sup>7</sup> Figure 7 shows the ratio of FPs that are produced with creaky voice portions or glottal pulses in the beginning (a) or end (b) of the FP as a function of pause context. It is apparent that creaky voice is frequent in the beginning of FPs and is relatively rare at the end of FPs. There are especially low numbers of creaky voice portions for the vocalic-nasal type FP *um* in the final position. It seems that the nasal consonant leads to an FP that is less likely to be produced with creaky voice. Duration measurements of creaky voice portions and glottal pulses seem to be stable across the *uh* and *um* FP types. The difference between the mean duration values of initial and final creaky voice portions/glottal pulses is not significant, though only just so, as determined by a *t*-test (initial 122 ms vs. final 145 ms;  $t = -1.95, p = 0.05$ ). Standard deviation values are 144 ms against 141 ms, which is quite long considering the mean values. This again shows the high variation within the feature of non-modal voice quality.



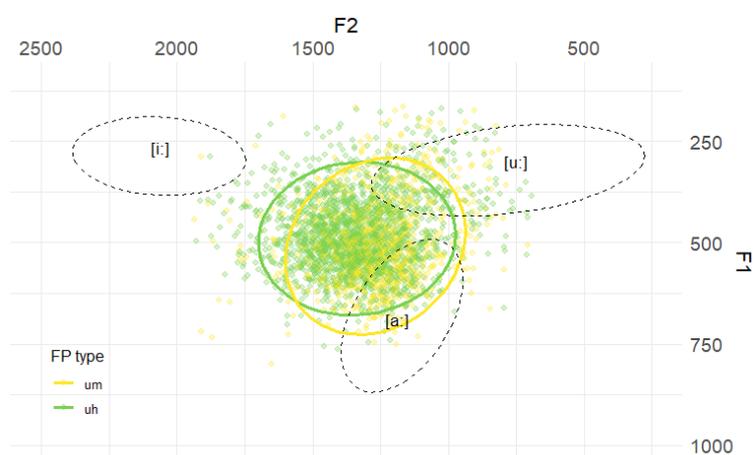
**Figure 7.** Creaky voice portion at (a) the beginning and (b) the end of FPs. The values at the top show the values representing 100% for each FP type. Note that the scales of the graphs are different, as there are considerably fewer FPs including creaky voice or glottal pulses at the end of the FP.

#### 4.5. Vowel Quality

Vowel quality was measured at the temporal midpoint of the vowel for the FPs *uh* and *um* using the Burg method provided by Praat<sup>8</sup> (Boersma and Weenink 2022). FPs are considered to have rather stable formants, *uh* more so than *um*, which is why a midpoint measurement was chosen; see Hughes et al. (2016). For the analysis of vowel quality, and to reduce any measurement errors, the dataset ( $n = 3304$ ) was reduced to only those

observations within three standard deviations from the mean of each the first and the second formants ( $n = 2996$ ; 308 observations excluded).

As seen in Figure 8, the vowels of the FPs *uh* and *um* show a high degree of overlap, which is further supported by the Pillai-score of 0.03. The Pillai score is a measure of overlap, ranging from 1 (no overlap) to 0 (complete overlap) (Kelley and Tucker 2020); it is determined by conducting a multivariate analysis of variance (MANOVA) using the first and the second formants as response variables and the FP type as the predictor variable. A Pillai score was calculated for each corner vowel as well, in order to measure the overlap of each lexical vowel with the FP vowel. The values support the relationship between vowels that can be seen in Figure 8. The high front vowel has little overlap with the FPs (Pillai = 0.8); the German central [a:] has more overlap (Pillai = 0.5), and the high rounded back vowel has even more overlap (Pillai = 0.3). It is important to note that the number of tokens for [u:] is considerably lower than for the other two corner vowels.



**Figure 8.** Vowel quality of the FPs *uh* and *um* in an F1–F2 chart in comparison to the corner vowels [a:], [i:], and [u:] by the same speakers. Ellipses include 95% of all data points.

#### 4.6. Discussion

To conclude the general results, it can be seen that tongue clicks are frequent in this corpus, as is the *uh* FP type. The overall disfluency rate of 8.67 FPs per minute is driven considerably by the high rate of clicks in the corpus. When excluding clicks (as is often done when looking at FPs), the rate of 5.7 FPs per minute is closer to the rates previously reported in the literature (Belz 2021). Contrary to our assumption (see Section 1.5), the vocalic-nasal type *um* is less frequent than the vocalic type. This is surprising, as Wieling et al. (2016) and de Leeuw (2007) report a change towards *um* being more frequent than *uh* in Germanic languages. This may be due to the fact that our corpus represents the FP use in the year of recording, which was 2001. Furthermore, our participants are only male speakers, while language change is more often led by females (Labov 1990; Wieling et al. 2016). The least frequent FP type is the nasal type, with even glottal FPs being more frequent. Glottal FPs should be considered in analyses of disfluencies, as they are nearly as frequent as the FP type *um* in this corpus.

As expected, vocalic-nasal FPs are produced with the longest duration (Hughes et al. 2016), followed by nasal FPs and vocalic FPs. Glottal FPs have the shortest duration. Furthermore, for the typical FPs *uh* and *um*, a duration hierarchy can be seen where FPs surrounded by pauses are the longest (–FP–), followed by IPU-final FPs (+FP–) and IPU-initial FPs (–FP+). Within-speech FPs (+FP+) show the shortest duration. Similar trends were found by Gósy and Silber-Varod (2021) for Hungarian vocalic FPs. For *uh*, the within-speech context is most frequent, while for *um* the within-pause context is most frequent.

In general, simple pauses are more frequent than waiting pauses or task changes. After task change pauses, the FP types *um* and *hm* are much more frequent than the vocalic *uh*.

This pause type (task change) only occurs infrequently after FPs, which suggests that FPs are used when the speaker is starting their speech. Pauses occur more often with the nasal or the vocalic-nasal type than with the vocalic FP type. In terms of pause duration, we find a significant difference in the duration of the waiting pause type, meaning that these pauses before an FP are longer than after an FP. This suggests that when an IPU ends with an FP before a waiting pause, the speaker is quicker in picking up their thoughts than after a waiting pause. In this case, the speaker may use the FP to buy time for formulating their next thought.

Creaky voice and glottal pulses are considerably more frequent in particle-initial position than in particle-final position, as reported by Belz (2021). In particle-final position, only a small portion of *um* FPs show creaky voice, while the percentage for *uh* is higher. A possible explanation for this striking difference is that because the FPs *uh* and *um* begin with a vowel, this vowel-initial position corresponds to the context in which a glottal stop can occur in German (i.e., words beginning with vowels). According to Kohler (1994), a glottal stop in German is frequently realised as creaky voice.

In terms of vowel quality, the vowels of *uh* and *um* show a high degree of overlap with each other. Furthermore, they are spread over a large portion of the central vowel space. The same similarity between the vowels of the two FP types was found by Hughes et al. (2016) for British English.

## 5. Normal vs. Lombard Speech Condition

Lombard speech is produced with a higher vocal effort, which typically results in a rise in the fundamental frequency (Jessen et al. 2005). Hyperarticulation, especially of the jaw and the tongue, has been reported by Šimko et al. (2016). This increased jaw opening accounts for the increase in first formant values that are typically reported in Lombard speech (Van Summers et al. 1988). Other effects of Lombard speech include a slower speaking tempo, which includes a lower articulation rate (Tuomainen et al. 2021) and possibly a higher pause rate. Our data reveals a lower speech rate in Lombard speech, as determined by a *t*-test ( $t = -2.13, p < 0.03$ ; 2.84 syll/s vs. 2.68 syll/s), as well as a higher articulation rate in Lombard speech ( $t = 6.07, p < 0.001$ ; 4.01 syll/s vs. 4.3 syll/s) (see Table 4). Effect sizes reveal a large effect for the difference in articulation rate ( $d = 0.86$ ) and a small effect size for speech rate ( $d = -0.3$ ), which is why articulation rate is the preferred factor for the following linear mixed models.

### 5.1. Frequency Distribution

The rate of FPs in normal speech (8.67 FPs/min) and Lombard speech (8.69 FPs/min) appear stable. However, when looking at the different FP types it becomes apparent that they are not as stable as the overall rate suggests. This is determined using a linear mixed model, with the rate of each FP type per speaker as the dependent variable, the FP type, condition, and articulation rate as independent variables, and the speaker as a random intercept:  $lmer(freq\_rate \sim fp\_type * condition + articulationrate + (1 | speaker), data = data)$ . The effect plot in Figure 9 shows that the rate of clicks and glottal FPs increases in the Lombard condition; the model (see Table A2 in the Appendix A) shows that the difference in conditions is not significant for the glottal FPs. The rate of the FPs *uh*, *um*, and *hm* decreases in the Lombard condition, and this difference is statistically significant. While articulation rate is included in the model, the factor does not reach statistical significance, i.e., it has no influence on the frequency of FPs. Possible reasons for these results are outlined in the discussion in Section 5.6.

### 5.2. Duration

To determine the influence of condition on FP duration, the following model was fitted:  $lmer(freq\_dur \sim condition * fp\_type + articulationrate + prepause + postpause + (1 | speaker), data = data)$ . The model output can be seen in Table A3 in the Appendix A. It shows that FPs in the Lombard condition are on average 48 ms longer than in the normal

condition, with the vocalic-nasal FP *um* being 162 ms longer than the other FPs. The pause context has an effect on the duration, in that FPs are shorter when no pause precedes or follows the FP. The pause after an FP has a larger effect on duration (−109 ms) than the pause before an FP (−37 ms). The articulation rate is included as the control variable; however, it has no influence on FP duration. Furthermore, the aforementioned duration hierarchy is applicable, with the pause context affecting the duration of *uh* and *um* to a similar degree.

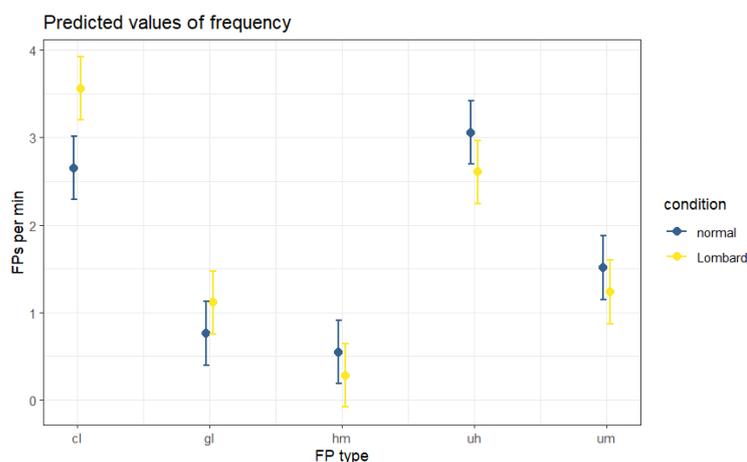


Figure 9. Effect plot of frequency (FPs/min) as a function of FP type and condition.

### 5.3. Pause Context

A similar distribution of pause types preceding and following FPs is observed in Lombard speech compared to normal speech. The FP types *hm* and *um* are more often followed and preceded by a pause than the vocalic type. The Lombard condition does not seem to affect either the rate of FPs that are surrounded by pauses or the type of pause.

Pause duration, however, is affected by the Lombard condition, as determined by a linear mixed model with pause duration as the dependent variable and pause position (before/after FP), condition, and pause type as predictor variables (Table A4). A significant effect of pause type shows that waiting pauses are longer than simple pauses by 1.2 s on average, while task change pauses are longer than simple pauses by 2.2 s on average. Furthermore, the Lombard condition increases the pause duration by 200 ms; moreover, there is an interaction between pause type, condition, and pause position that makes waiting pauses before FPs longer in the Lombard condition than in the normal condition (see Table 3).

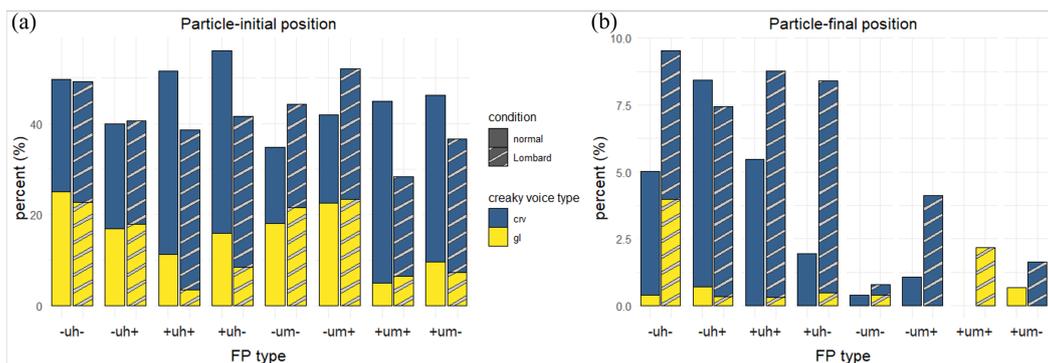
Table 3. Comparison of pause duration (mean and standard deviation) in normal vs. Lombard speech, divided by position (−FP = preceding an FP; FP− = following an FP) and type of pause (simple pause = p; waiting pause = p\_w; task change pause = tc).

Pause Position	Pause Type	Normal Mean (sd) in ms	Lombard Mean (sd) in ms	Difference in ms
−FP	p	1030 (919)	1330 (1457)	300
	p_w	2642 (1473)	3978 (2984)	1336
	tc	3389 (1903)	4706 (3346)	1317
FP−	p	985 (1123)	1196 (1329)	211
	p_w	2189 (1313)	2388 (1795)	199
	tc	3192 (1807)	4167 (2863)	975

### 5.4. Voice Quality

The percentage of FPs that include creaky voice portions or glottal pulses in normal compared to Lombard speech seems to only differ for the FP type *uh*. In normal speech,

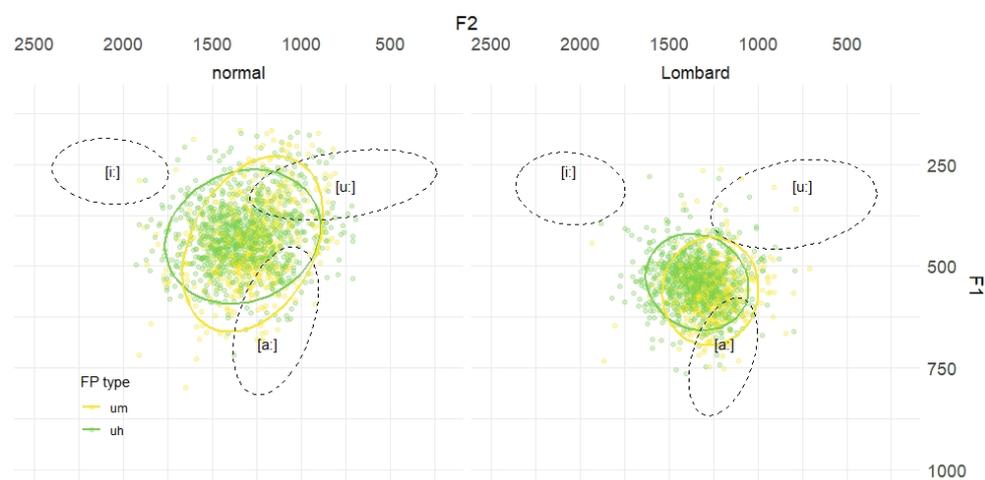
49% of all vocalic FPs include creaky voice or glottal pulses, while only 42% are affected in Lombard speech. The difference for the *um* FP type is not as large, with a slightly higher portion affected in Lombard speech compared to normal speech (normal = 40%, Lombard = 42%). There seems to be no pattern for pause context (see Figure 10). The 189 FPs that include 100% creaky voice are included in Figure 10a as initial creaky voice portions (132 tokens in normal speech, 57 in Lombard). Creaky voiced portions (or glottal pulses) in FP-final position increase, while in FP-initial position this is only true for IPU-initial *um* (*-um+*) and *um* in isolation (*-um-*).



**Figure 10.** Creaky voice portions and glottal pulses (colours) in the FPs *uh* and *um* divided by their pause context (+/–) and the speech condition (patterns).

### 5.5. Vowel Quality

In Figure 11, it can be seen that the Lombard condition has relatively little influence on the vowel space taken up by the three corner vowels. The close front vowel is distributed similarly in both conditions, the vowel space for the open central vowel decreases slightly, and the ellipse for the close back vowel is somewhat more round in the Lombard condition than in the normal speech condition. However, the vowel space taken up by the vowels in the FPs *uh* and *um* is drastically decreased. Fitting a linear model suggests that the Lombard condition has a significant effect on the vowel height of FPs:  $lmer(f1 \sim condition * fp\_type + articulationrate + prepause + postpause + (1 | speaker), data = data)$ . The F1 is in the mean 97 Hz higher in the Lombard condition than in the normal condition, i.e., the vowels are produced with a lower tongue position or with a more open jaw. The articulation rate has a significant effect on the F1, in that with every one-unit increase (e.g., from 3 syll/s to 4 syll/s), the F1 increases by 30 Hz. The effect of the occurrence of a pause before the FP reaches significance as well, though we consider the difference of an increase by 10 Hz to be negligible (Whalen et al. 2022). A increase in F1 values due to a greater jaw aperture can be observed in the literature (Šimko et al. 2016). A correlation between signal amplitude (resulting from speech production alone, not environmental or technical factors) and the F1 value has previously been observed by Ibrahim et al. (2022), suggesting that the increased F1 is not independent of the increased intensity. There is no significant difference in F2 values between the conditions determined by the following model:  $lmer(f2 \sim condition * fp\_type + articulationrate + prepause + postpause + (1 | speaker), data = data)$ . A main effect of duration reveals that a longer FP duration results in a lower F2, i.e., when the FP duration increases by one unit (=1 s) the F2 decreases by 107 Hz, or using more realistic numbers for this field, when the FP duration increases by 100 ms, the F2 decreases by approximately 11 Hz. The significant interaction of Lombard condition and FP type suggests that the condition has an effect on the vocalic-nasal FP and has no effect on the vocalic FP. The difference of 34 Hz is rather small, and may be a negligible effect (Whalen et al. 2022).



**Figure 11.** Vowel quality of the FPs *uh* and *um* in normal vs. Lombard speech in comparison to the corner vowels [a:], [i:], and [u:] by the same speakers. Ellipses include 95% of all data points.

### 5.6. Discussion

To conclude this section, we summarise the results briefly and try to explain the more surprising results. The rate of the typical FPs (*uh*, *um*), and *hm* decreases in Lombard speech, while the rate of the glottal FPs and tongue clicks increases. The duration of FPs increases in the Lombard condition by 48 ms, while the articulation rate is not a significant factor for this difference. Pause context affects the duration of the FPs in that they are longer when surrounded by pauses and shorter with no pauses. FP-following pauses have a greater effect on the duration of the FP than FP-preceding pauses, which may be explained by the common phenomenon of final lengthening (Lindblom 1968). Pause durations of both types, both FP-preceding and following, increase, which is more the case for FP-preceding pauses. Furthermore, pause types are 200 ms longer on average in the Lombard condition than in the normal condition. Waiting pauses are affected to a different degree: in FP-preceding position, the Lombard condition increases its duration by over a second, while in FP-following position this difference is considerably smaller (200 ms).

When looking at voice quality, the most noticeable effect is that creaky voice portions and glottal pulses increase in particle final position. In terms of vowel quality in Lombard speech, an increase in F1 mean values is detected for the vowels in *uh* and *um* along with a reduction in the vowel space used for these vowels. While the increase in F1 values can be explained by a greater jaw aperture (Šimko et al. 2016), the effect on vowel space is more surprising. It seems that the vowel target for the FP is reached in the Lombard condition better than in the normal condition, which is surprising, as it is unclear whether FPs have vowel targets at all (Gick et al. 2004). It could be argued that this reduction of vowel space is an act of articulatory precision consistent with Lombard speech as a form of clear speech. However, Lombard speech does not show all characteristics of clear speech (for the concept of clear speech, see (Smiljanić and Bradlow 2009)). In particular, there is no general effect of increased formant dispersion in Lombard speech, as shown here with the corner vowels as well as by other studies in which no robust Lombard effect on F2 was found (Garnier et al. 2006; Gully et al. 2019; Hay et al. 2017; Šimko et al. 2016; Van Summers et al. 1988). It is possible that the reduction of vowel space in FPs is a secondary effect of the Lombard condition. For example, the increased muscle tension that is needed for the increased vocal effort may play a role (Wohlert and Hammen 2000).

One of our results that is in need of explanation is that the FP rate for *uh*, *um*, and *hm* is lower in Lombard speech than in normal speech. A possible explanation could be as follows. Although it is undisputed that FPs can be used by the listener, it is not yet clear whether the speaker actively produces FPs as a signal to the listener or whether FPs are the result of planning processes, etc., on the part of the speaker which, while they

can be detected by the listener, are not actively intended by the speaker (see [Corley and Stewart \(2008\)](#) for a further discussion). If FPs from the set of *uh*, *um*, and *hm* are caused by active signalling from the speaker, this could mean that communicative interaction in the Lombard situation of the corpus was somewhat inhibited, resulting in a reduction in the FPs involving active signalling.

This explanation would not cover the behaviour of glottal FPs and clicks, which are actually more frequent in Lombard speech than in normal speech. It is possible that these two types of FPs are hard to produce deliberately, and generally more difficult to perceive (unless very long or loud) than the other FP types. In terms of production, creaky voice and clicks are partially the byproducts of aerodynamic principles, and it is probably difficult to learn the gestural coordination patterns that are necessary to actively and deliberately produce them (as evidenced by the fact that they are infrequently used phonemically in languages). If glottal FPs and clicks are more difficult to control in production, the active signalling function might not work well enough, and if they are more difficult to perceive, speakers would realise that their signalling intention might not reach the listener. Therefore, active signalling of glottal FPs and tongue clicks are unlikely to occur, meaning that the above explanation would no longer apply. This does not explain why glottal FPs and clicks are actually more frequent in Lombard speech than in normal speech, instead of merely being equal in number. For glottal FPs, the effect turns out to be non-significant, while clicks are significantly more frequent in Lombard speech than in normal speech. One possible explanation for this could be that increased jaw lowering ([Schulman 1989](#)) and associated tongue lowering in loud speech increase negative air pressure, thereby enhancing the production of clicks.

A similarly difficult situation occurs when trying to explain the patterns of creaky voice and glottal pulses in *uh* and *um*, shown in Figure 10a. [Keating et al. \(2015\)](#), in their typology of different kinds of creaky voice, make (among other criteria) a distinction between types that are associated with glottal constriction and those without (cf. [Jessen \(2012\): 52ff.](#), for the forensic relevance of the distinction between constricted and non-constricted types of creaky voice). For predictions about the effect of Lombard speech, it matters whether creaky voice in FPs is of constricted or non-constricted type. For the non-constricted type, the prediction is that creaky voice is reduced in the Lombard condition; Lombard speech is associated with increased subglottal air pressure ( $P_{\text{sub}}$ ), which would have an inhibiting effect on non-constricted creak, which is associated with lowered  $P_{\text{sub}}$ . For the constricted type, the prediction is that creaky voice is increased in the Lombard condition. A clear glottal adduction gesture is useful in Lombard speech in order to withstand the increase in  $P_{\text{sub}}$  due to Lombard speech (i.e., with increased vocal effort). Particle-initial position is a context in which the constricted type of creaky voice is to be expected as a correlate of the glottal stop in German ([Kohler 1994](#)). This would predict more creaky voice in Lombard speech than in normal speech, which as a general trend cannot be seen. However, when looking separately at FPs preceded by a pause (–FP) or by speech (+FP), creaky voice gains ground in Lombard speech in the post-pausal (–FP) position. This is a position in which a glottal stop is more likely than without a preceding pause ([Kohler 1994](#); [Krech 1968](#)). This means that the second explanation seems to have an effect, although it appears to interact with other factors as well. The patterns in Figure 10b (particle-final position) are particularly difficult to understand, as it is not clear what kind of creaky voice occurs; however, particle-final position could involve non-constricted creak, especially when occurring before a pause. Nonetheless, there is generally more creak rather than less in Lombard, and there is no trend across *uh* and *um* of this effect being smaller before a pause than without a following pause.

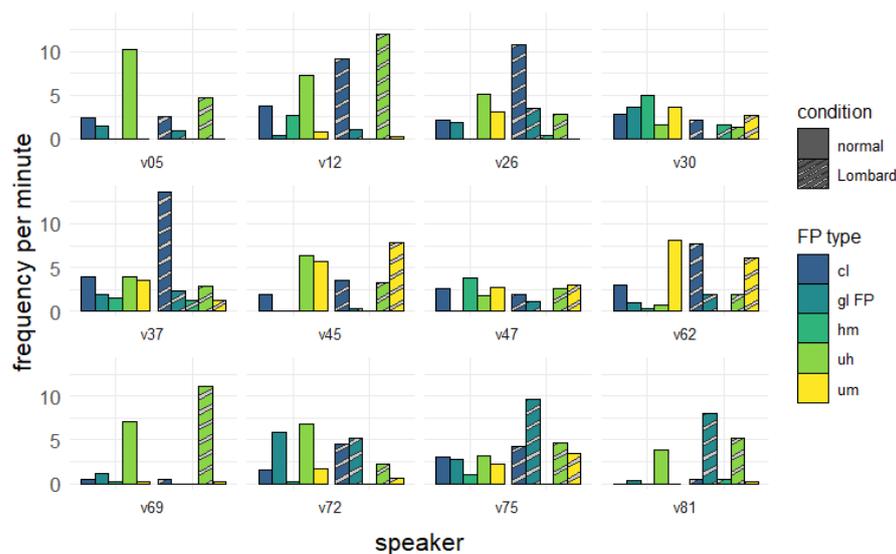
## 6. Speaker Specificity

As outlined before, in phonetic casework it is important to be aware of the distribution of a certain feature within a relevant population as well as the within-speaker differences. In this section, we show differences between speakers for twelve sample speakers from the

corpus, each of whom were selected due to their heavy use of one of the five phenomena under investigation here (*uh*, *um*, *hm*, glottal FPs, and clicks). For each phenomenon, we chose two speakers with the highest rate per minute, and we additionally selected two more speakers occurring among heavy users in more than one category. The question was whether the frequent use of one phenomenon had an effect on the use of the other FPs (Section 6.1). Within-speaker differences were addressed by comparing the two conditions, i.e., normal speech vs. Lombard speech (Sections 6.1 and 6.4), by examining per-speaker standard deviations across FP tokens (Section 6.3) for both (Sections 6.2 and 6.5).

### 6.1. Frequency Distribution

Figure 12 shows the frequency distribution of the five phenomena under investigation in the normal and Lombard speech conditions for twelve sample speakers. Recall that the general tendency discussed in Section 5 is that the frequency of the typical FPs (*uh*, *um*) and *hm* decreases, while the frequency of glottal FPs and tongue clicks increases. This general trend, however, cannot be seen in all individual speakers. For certain speakers, the rate of tongue clicks increases visibly (e.g., v12, v26, v37); for speaker v75, both clicks and glottal FPs increase, and the rate of *uh* and *um* increases in Lombard speech. The rate of the FPs *uh*, *um*, and *hm* increases for certain speakers (v12, v45, v69) in Lombard speech as well, which is contrary to the general trend. Individual patterns are consistent for certain speakers (e.g., v05, v45, v62, v69), for whom the most frequently used FPs in normal speech occur in Lombard speech as well, though the frequencies change somewhat. For other speakers, one FP type simply does not occur in Lombard speech at all, such as *um* for speaker v26, *hm* for speaker v47, or glottal FPs for speaker v30.



**Figure 12.** Frequency distribution of FPs for twelve sample speakers comparing their production of FPs in normal speech (left-hand side for each speaker) vs. Lombard speech (diagonal stripes; right-hand side for each speaker).

### 6.2. Duration

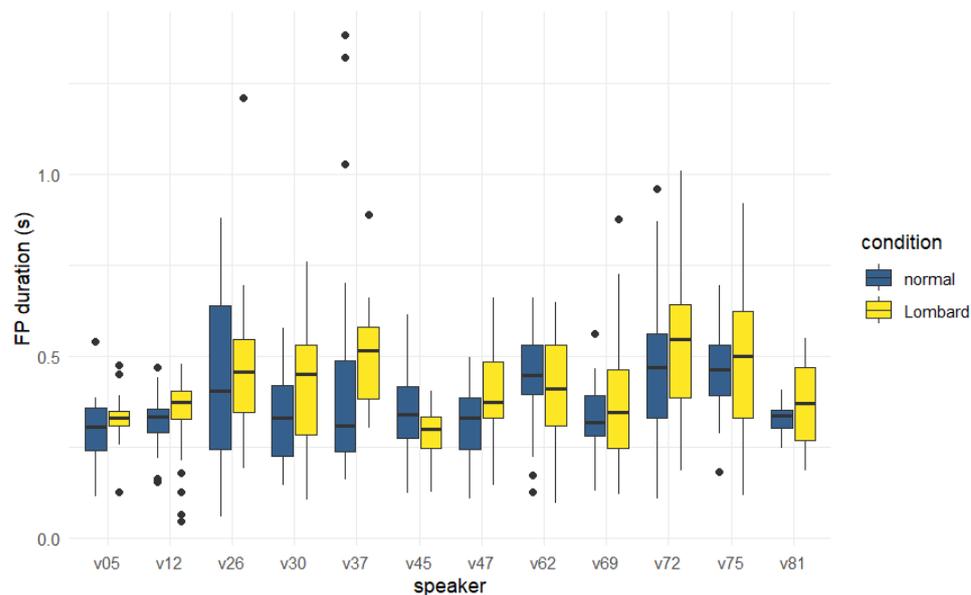
FP durations for all tokens in normal and Lombard speech conditions combined vary considerably between speakers (Table 4): mean glottal FP durations range from 93 ms to 629 ms, while standard deviations per speaker range from as little as 24 ms to over 600 ms. A small standard deviation for an individual speaker means that the speaker is quite consistent in producing FPs with similar durations, while a high standard deviation means that the speaker exhibits considerable variance. Mean duration for the nasal FP ranges from 160 ms to 721 ms (sd: 82–350 ms) for the twelve sample speakers. The durations for the vocalic FP are less variable, ranging from 190 ms to 450 ms, and within-speaker variation

is smaller (sd: 70 ms to 280 ms). The general trend that *um* is longer can be seen for the speakers selected here (mean: 330–650 ms, sd: 46–315 ms). Glottal FPs seem to vary the most in duration between speakers, while the vocalic FP *uh* varies the least.

**Table 4.** Minimum and maximum mean duration values and standard deviation (in ms) of the FPs in the sample set of twelve speakers.

FP Type	Minimum	Maximum
uh	188 (69)	448 (278)
um	334 (46)	648 (315)
hm	161 (82)	721 (349)
gl FP	93 (24)	629 (615)

To illustrate the speakers’ individual patterns, Figure 13 shows the pooled durations of the typical FPs *uh* and *um* in the normal and Lombard conditions. It is apparent when looking at the sample set that the durations of FPs in both conditions seem to not depend on each other. It is not the case that FPs are always longer in one condition than in the other, nor that the variance is always higher in one of the two conditions. While speakers v12 and v47 show similar standard deviations in both conditions and a similar increase in duration in the Lombard condition, other speakers (v26, v81) do not show this pattern. Between-speaker differences in FP duration seem to be slightly larger than the within-speaker differences that result from the normal–Lombard difference. For example, speaker v12 has a relatively low mean FP duration in both the normal and Lombard conditions, whereas speaker v72 has relatively high values in both conditions. Overall, however, the difference in the within-speaker and between-speaker variations is not large.

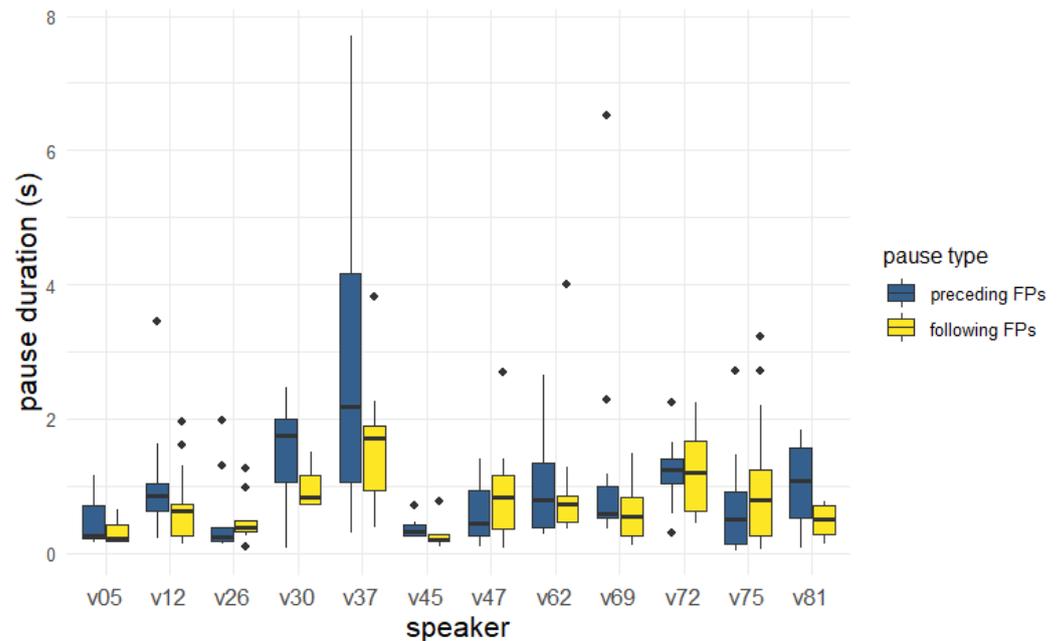


**Figure 13.** FP duration of the typical FPs *uh* and *um* per speaker (pooled) as a function of speech condition.

### 6.3. Pause Context

The mean durations of pauses per speaker for all tokens in the normal and Lombard conditions combined seem to be even more variable than the mean durations of the FPs themselves (see Figure 14 for simple pauses). FP-preceding pauses, regardless of type (simple pauses, waiting pauses, task changes), range from almost 800 ms to 2831 ms, with the standard deviation being similarly variable (sd: 562–2977 ms). The high standard deviation value seems to be caused by speaker v37 (see Figure 14), as the other participants

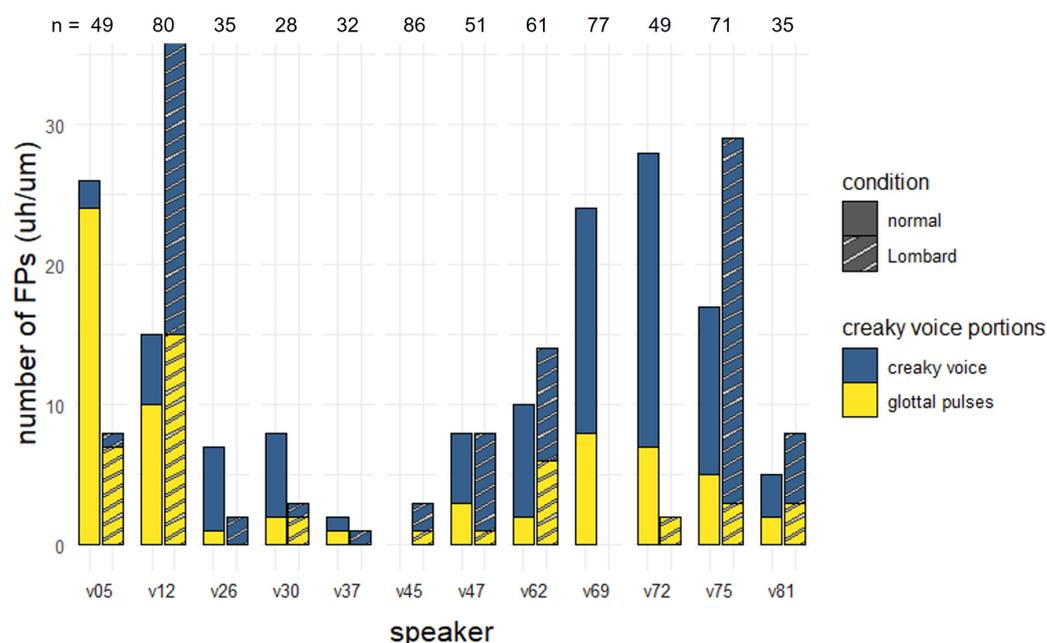
show smaller variations from the mean. FP-following pauses show a general trend of being shorter, ranging from 474 ms to 1561 ms (sd: 309–1895 ms). The addition of speech condition to this variable does not add any insightful information other than more variance, as was the case above for FP duration; for this reason, we refrain from adding the factor to Figure 14.



**Figure 14.** Duration of pauses surrounding FPs per speaker. Only simple pauses (p) are considered; waiting pauses and task changes are excluded.

#### 6.4. Voice Quality

In this section we only look at particle-initial creaky voice or glottal pulses, as they are very infrequent in final position, as shown above in Section 5.4. The production of creaky voice and glottal pulses in the set of sample speakers is highly variable. Certain speakers (v26, v30, v37, v45, v47, v81) produced fewer than ten instances in each condition with a portion of this voice quality, as measured independently of their total number of FPs produced (see Figure 15). A speaker to highlight here is v45, who produced 86 *uh* and *um* FPs in total, of which only three instances were produced with initial creaky voice or glottal pulses. All of these were produced in the Lombard condition. Equally low is the number for speaker v37; this participant, however, produced only 32 *uh* and *um* FPs in total. The speakers who produced more than ten tokens using creaky voice or glottal pulses in their FPs tend to show more variance between conditions. For example, speaker v05 produced around 70% of all their FPs with a non-modal voice in the beginning of the FPs. The majority of these were produced with glottal pulses instead of creaky voice, which was more the case in the normal condition than in the Lombard condition. Speaker v12, in contrast, produced non-modal voice portions more frequently in the Lombard condition. Two speakers (v69, v72) produced more than twenty FP tokens with creaky voice or glottal pulses in the normal condition and none or very few in the Lombard condition. Only four out of twelve speakers (v37, v47, v62, v81) can be considered to not show differences between the two conditions. Thus, between- and within-speaker differences seem to be quite high for this feature.



**Figure 15.** Number of particle-initial creaky voice portions and glottal pulses in the twelve sample speakers. The values at the top denote the speakers' total number of the FPs *uh* and *um*, which are the FPs for which creaky voice and glottal pauses were annotated.

### 6.5. Vowel Quality

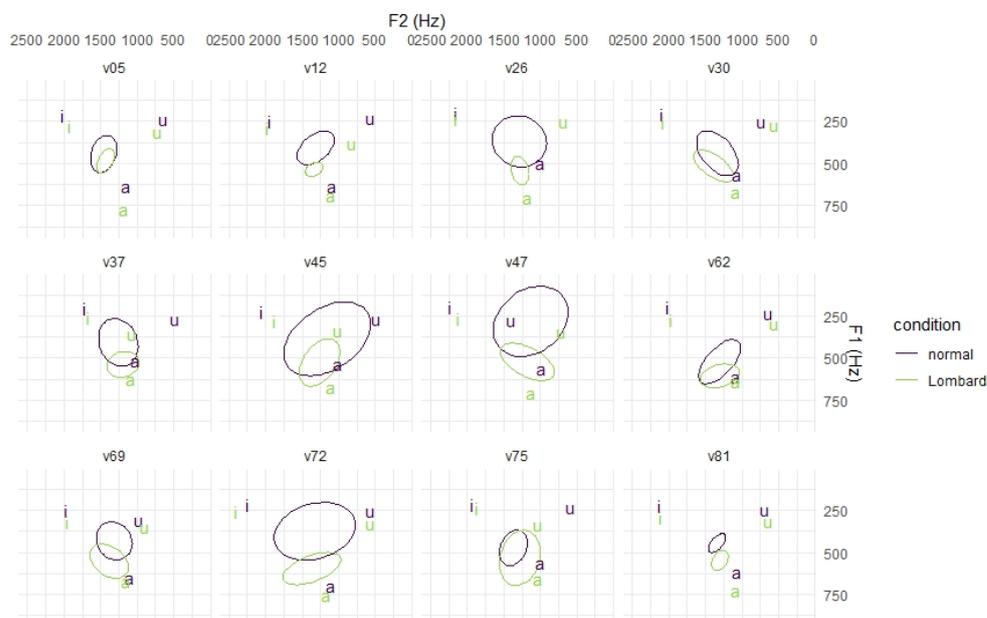
The general trend for the Lombard condition is a decreased vowel space and a lowering of the mean F1 value. For our sample speakers, a decrease of the vowel space can be seen for nine out of the twelve speakers (Figure 16). Of the remaining three, two (v69, v81) show a rather similar vowel space for the vowels in the FPs, while only one (v75) shows a clear increase in vowel space in the Lombard condition. All of the sample speakers show a lowering of the vowel space in the Lombard condition compared to the normal condition, which is equivalent to an F1 increase. A change in F2 seems to be absent. The speakers vary considerably in how much of the vowel space is used for the FP vowels. While certain speakers (v05, v12, v75, v81) produce their FP vowels in a quite limited vowel space, other speakers' vowels (v45, v47, v72) expand over a large portion of the central vowel space.

### 6.6. Discussion

Our glimpse into speaker specificity on the basis of twelve sample speakers shows high variation in FP frequency distribution, FP duration, pause duration, voice quality, and vowel quality. A certain amount of this variation is due to variation between speakers, which has positive implications for forensic voice comparison; however, there is substantial, though slightly less, variation within speakers. Part of this within-speaker variation follows the statistical patterns addressed in the previous section, while part of it does not, and the degree to which the patterns are congruent or incongruent with the general trend differs. There are mixed findings in the literature in terms of speaker specificity when using a disfluency profile. [McDougall and Duckworth \(2018\)](#) found a consistent pattern in two tasks (an interview and a telephone conversation), while [Harrington et al. \(2021\)](#) showed deviating patterns for the same speakers in a third task (a voicemail message).

A specific characteristic of this study lies in the fact that two speech conditions, normal and Lombard, were investigated; the differences between these conditions have relatively strong impacts on several speaker characteristics, and as such they are classified as "mismatched conditions" ([Alexander et al. 2005](#)). This is in contrast to matching conditions (e.g., the same or similar speech task performed two weeks apart). Intra-speaker variation is expected to be stronger under mismatched conditions than under matching conditions. It would be possible to reduce the intra-speaker variation by applying a normalisation

procedure that takes into account the statistically dominant patterns of the normal–Lombard distinction. For example, the duration of FPs could be increased in normal condition or decreased in Lombard condition before the voice comparisons are conducted. This would increase the comparison scores (i.e., reduce the difference) when speakers who follow the dominant trend closely are compared while decreasing them for other speakers; altogether, such a normalisation procedure would probably increase speaker discrimination. Had the same methods we used here been used on a dataset using matching conditions instead, intra-speaker variation for the FP features would be expected to be lower. As both the results of the aforementioned studies by (Harrington et al. 2021; McDougall and Duckworth 2018) and those of the present paper suggest, the difference between matching and mismatched conditions is somewhat of a continuum, at least as far as FP patterns are concerned.



**Figure 16.** Vowel quality of the FPs *uh* and *um* in normal speech vs. Lombard speech in comparison to the corner vowels [a:], [i:], and [u:]. Ellipses include 95% of all data points.

### 7. General Discussion

As outlined above, we investigated the filler particles (FP) used in a spontaneous speech corpus of 100 male native German speakers. In our analyses, we looked into the use of the five phenomena *uh*, *um*, *hm*, glottal FPs, and tongue clicks and the feature frequency distribution of FPs, FP duration, pause duration, voice quality, and vowel quality.

All in all, we showed that the phenomena which are often disregarded in disfluency research, namely, glottal FPs and tongue clicks, are equally as frequent as the FPs *uh*, *um*, and *hm*. In this corpus, the vocalic type *uh* and tongue clicks occur as the most frequent FP phenomena, and glottal FPs are nearly as common as the FP type *um*, though it should be noted that percussives are grouped under tongue clicks here and may contribute to the high rate of clicks in the corpus. Nasal FPs *hm* are quite rare. Considering the duration of the FP types *uh* and *um*, the pause context seems to have an influence on their duration, in that FPs that are produced within pauses are longest, followed by IPU-final FPs and then IPU-initial FPs. FPs occurring within speech stretches are shortest. We call this a duration hierarchy. Furthermore, we can confirm previous findings that the FP *uh* is less often flanked by a pause than the FP *um* (Belz 2021; Clark and Fox Tree 2002); *uh* is used very rarely after waiting pauses and task changes, which means that the introduction of new content entails the FPs *um* and *hm* rather than *uh*. It is apparent that non-modal voice is very common in the beginning of the FPs *uh* and *um*, as over 40% of each FP type includes an initial glottal stop or creaky voice of varying length. The vowel quality of the typical FPs (*uh/um*) is

spread over a large area of the central vowel space, and they show a very high degree of overlap.

The main findings of the comparison between the normal speech condition and the Lombard condition include that Lombard speech promotes the production of tongue clicks and that the frequency of the FPs *uh*, *um*, and *hm* decreases. Durational measures of FPs and surrounding pauses increase, which is moreso the case for the longer *waiting pause* type. An interesting effect on vowel quality can be observed in Lombard speech. The vowel space in the Lombard condition taken up by the vowels of the FPs is drastically reduced compared to the normal condition, along with the typical F1 increase. The reasons for this are yet unknown, though the increased muscle tension that is needed for the increased vocal effort and the higher intensity may play a role (Wohlert and Hammen 2000). Articulation rate was included as a control variable in the linear mixed models, though it is only a significant factor in predicting the first formant values; the difference of 30 Hz per increase in articulation rate may, however, be rather small considering the large increase in speaking tempo. For example, a change in articulation rate of two units (from 3 syll/sec to 5 syll/sec) only increases the F1 by 60 Hz, according to the linear mixed model.

Speaker specificity was investigated on the basis of twelve sample speakers. We found a high degree of between-speaker variation as well as a substantial amount of within-speaker variation in terms of both the features investigated and the different filler particles. A certain amount of the within-speaker variation is internal to the conditions (expressed as the standard deviations per condition), while the rest is due to the normal and Lombard conditions. The details have been presented in Section 6. The two conditions are of the mismatched type, which means that strong intra-speaker variation can be expected. It would be possible to reduce some of this variation by implementing a normalisation procedure in which the statistical differences are taken into consideration, though even then the speaker discrimination performance would probably remain lower compared to the case of matching conditions. It is possible that the patterns of intra- and inter-speaker variation would look somewhat different if the twelve speakers had been selected randomly instead of using the selection process described at the beginning of Section 6. Further research into speaker specific disfluency patterns in German using several recordings from the same speaker in a variety of conditions could provide a more complete picture of the nature and size of the within-speaker effects. Another research goal executable with the available dataset would be to use all 100 speakers instead of only twelve for a full investigation of the speaker discrimination potential of the FP features. Using the likelihood ratio framework for such an investigation would have the advantage that the implications for forensic phonetics could be fully worked out. Though such a study was beyond the scope of the present paper, even with the current results two implications for forensic phonetics can be pointed out. First, the average values shown for the different FPs and their features in the two conditions, along with the standard deviations, provide an indication of the typicality patterns addressed in Section 1.6, i.e., which feature values are typical of the relevant population of speakers (here, adult male German speakers) and which are non-typical. Second, while the articulation rate was measured and included in the statistical models, it had almost no effect on the features. This indicates strong independence between the disfluencies studied here and the articulation rate. Because both disfluency and articulation rate are frequently used speaker characteristics in voice comparison casework, this independence is beneficial when combining results from different characteristics (Gold and French 2011; Jessen 2018).

**Author Contributions:** Conceptualization, B.M. and J.T.; Data curation, M.J.; Formal analysis, B.M.; Funding acquisition, J.T.; Investigation, B.M.; Supervision, J.T. and M.J.; Visualisation, B.M.; Writing—original draft, B.M.; Writing—review and editing, J.T. and M.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by private funds from the Max Mangold estate in the form of a scholarship for the first author, and in part by the Deutsche Forschungsgemeinschaft (Project TR 468/3-1) for the second author.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Due to legal issues, the corpus is not freely available though the data files and our R script are available on OSF: <https://osf.io/yf3et/> (accessed on 20 October 2022).

**Acknowledgments:** The authors would like to thank Ivan Yuen for advice on statistical modelling, three anonymous reviewers for valuable feedback on an earlier version of this paper, as well as student assistants Hanna Zimmermann, Chiara Pletto, Verena Fus, and Sonja Persch for their annotation work.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Linear Models Output

**Table A1.** Model output of the linear model for pause duration as the dependent variable and pause position (pre/post) and pause type (p, p\_w, tc) as independent variables.

	Estimate	Std. Error	t-Value	Pr (< t )
(Intercept)	1.083	0.036	30.32	<0.001 ***
typepre	0.094	0.053	1.77	0.08
pausetypep_w	1.202	0.121	9.96	<0.001 ***
pausetypep_w	2.477	0.227	10.92	<0.001 ***
typepre:pausetypep_w	0.803	0.16	5.02	<0.001 ***
typepre:pausetypep_w	0.308	0.246	1.26	0.21

**Table A2.** Model output of the linear mixed model for frequency, with the rate per minute of FPs as the dependent variable and the FP type, condition, and articulation rate as independent variables.

	Estimate	Std. Error	df	t-Value	Pr (< t )
(Intercept)	3.31	0.87	280.07	3.81	<0.001 ***
fp_typegl	-1.89	0.24	890.83	-7.81	<0.001 ***
fp_typehm	-2.1	0.24	890.83	-8.7	<0.001 ***
fp_typeuh	0.41	0.24	890.83	1.67	0.09
fp_typeum	-1.14	0.24	890.83	-4.71	<0.001 ***
conditionLombard	0.91	0.25	947.46	3.65	<0.001 ***
articulationrate	-0.16	0.21	266.15	-0.75	0.45
fp_typegl:conditionLombard	-0.56	0.34	890.83	-1.63	0.1
fp_typehm:conditionLombard	-1.18	0.34	890.83	-3.45	<0.001 ***
fp_typeuh:conditionLombard	-1.36	0.34	890.83	-3.98	<0.001 ***
fp_typeum:conditionLombard	-1.19	0.34	890.83	-3.47	<0.001 ***

**Table A3.** Model output of the linear mixed model for duration with the duration of FPs (in seconds) as the dependent variable and the FP type, condition, and articulation rate as independent variables.

	Estimate	Std. Error	df	t-Value	Pr (< t )
(Intercept)	0.502	0.065	724.9	7.7	<0.001 ***
conditionLombard	0.048	0.009	3027	5.55	<0.001 ***
fp_typeum	0.162	0.01	3289	16.11	<0.001 ***
articulationrate	-0.017	0.016	772	-1.04	0.3
prepause+	-0.037	0.007	3296	-5.63	<0.001 ***
postpause+	-0.109	0.007	3280	-16.26	<0.001 ***
conditionLombard:fp_typeum	-0.023	0.013	3261	-1.73	0.08

**Table A4.** Model output of the linear model for the pause duration as the dependent variable and the pause position (pre/post), pause type (p, p\_w, tc), and condition (normal/Lombard) as independent variables.

	Estimate	Std. Error	t-Value	Pr (< t )
(Intercept)	0.985	0.048	20.53	<0.001 ***
typepre	0.045	0.072	0.62	0.54
pausetypep_w	1.204	0.165	7.31	<0.001 ***
pausetypepc	2.206	0.284	7.78	<0.001 ***
conditionLombard	0.211	0.07	2.99	<0.01 **
typepre:pausetypep_w	0.408	0.213	1.92	0.06
typepre:pausetypepc	0.152	0.31	0.49	0.62
typepre:conditionLombard	0.09	0.105	0.85	0.39
pausetypep_w:conditionLombard	−0.012	0.238	−0.05	0.96
pausetypepc:conditionLombard	0.765	0.46	1.66	0.1
typepre:pausetypep_w:conditionLombard	1.048	0.317	3.31	<0.001 ***
typepre:pausetypepc:conditionLombard	0.253	0.496	0.51	0.61

**Table A5.** Model output of the linear mixed model for the first formant with the F1 (in Hz) as the dependent variable and the FP type, condition, and articulation rate as independent variables.

	Estimate	Std. Error	df	t-Value	Pr (< t )
(Intercept)	312	34	1115.03	9.27	<0.001 ***
conditionLombard	97	4	2978.47	22.32	<0.001 ***
fp_typeum	0.6	5	3061.22	0.12	0.9
articulationrate	30	8	1181.06	3.61	<0.001 ***
fp_dur	10	9	3054.38	1.18	0.24
prepause+	−10	3	3046.97	−3.08	<0.01 **
postpause+	−0.5	3	3031.99	−0.14	0.89
conditionLombard:fp_typeum	12	7	3010.67	1.85	0.06

**Table A6.** Model output of the linear mixed model for the second formant with the F2 (in Hz) as the dependent variable and the FP type, condition, and articulation rate as independent variables.

	Estimate	Std. Error	df	t-Value	Pr(< t )
(Intercept)	1307	74	1383.6	17.63	<0.001 ***
conditionLombard	15	9	3020.62	1.61	0.11
fp_typeum	10	11	3056.78	0.93	0.35
articulationrate	10	18	1506.81	0.55	0.59
fp_dur	−107	19	3047.17	−5.63	<0.001 ***
prepause+	−4	7	3039.32	−0.54	0.59
postpause+	29	7	3025.15	3.99	<0.001 ***
conditionLombard:fp_typeum	−34	14	3007.09	−2.47	<0.05 *

**Notes**

- 1 As opposed to “true” acoustic silence, where background noise is absent.
- 2 The majority of forensic phonetic casework deals with male voices, which is why most research in this area focuses on this speaker group.
- 3 Maclay and Osgood (1959) found a mean of 152 words/min.
- 4 Converting this unit to a rate per minute is more difficult than for the rate per 100 words, as syllable duration is highly depend on the syllable structure and the stress and pause context (Crystal and House 1990). We reached an approximation by taking the most frequent syllable structure (CVC) reported by Crystal and House (1990) and calculating the mean duration of the CVC type before and after pauses in stressed and unstressed position (mean = 250 ms).
- 5 Due to legal issues, the corpus is not freely available though the data files and our R script are available on OSF: <https://osf.io/yf3et/> (accessed on 20 October 2022).

- <sup>6</sup> In certain cases there may be one or four taboo words instead.
- <sup>7</sup> Note that these are not the same as glottal FPs, as a vowel may still be discernible.
- <sup>8</sup> Maximum formant: 5000 Hz; maximum number of formants: 5; window length: 0.025 s; dynamic range: 50 Hz.

## References

- Adams, Martin R., and John Hutchinson. 1974. The effects of three levels of auditory masking on selected vocal characteristics and the frequency of disfluency of adult stutterers. *Journal of Speech and Hearing* 17: 682–88. [CrossRef] [PubMed]
- Alexander, Anil, Damien Dessimoz, Filippo Botti, and Andrzej Drygajlo. 2005. Aural and automatic forensic speaker recognition in mismatched conditions. *International Journal of Speech, Language and the Law* 12: 214–34. [CrossRef]
- Bates, Douglas, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]
- Batliner, Anton, Andreas Kießling, Susanne Burger, and Elmar Nöth. 1995. Filled Pauses in Spontaneous Speech. Paper presented at International Congress of Phonetic Sciences (ICPhS), Stockholm, Sweden, August 13–19; pp. 472–75.
- Bellinghausen, Charlotte, Simon Betz, Katharina Zahner, Alina Sasdrich, Marin Schröer, and Bernhard Schröder. 2019. Disfluencies in German adult-and infant-directed speech. Paper presented at SEFOS: 1st International Seminar on the Foundations of Speech. Breathing, Pausing and The Voice, Sønderborg, Denmark, December 1–3; pp. 44–46.
- Belz, Malte. 2017. Glottal filled pauses in German. In *Workshop on Disfluency in Spontaneous Speech (DiSS 2017)*. Stockholm: KTH Royal Institute of Technology, pp. 5–8.
- Belz, Malte. 2018. Vowel quality of German äh and ähm in dialogue moves. In *Phonetik und Phonologie im Deutschsprachigen Raum (P&P14)*. Wien: Universität Wien, pp. 13–17.
- Belz, Malte. 2019. *GECO-FP*. Berlin: Humboldt-Universität zu . 20794. [CrossRef]
- Belz, Malte. 2021. *Die Phonetik von äh und ähm: Akustische Variation von Füllpartikeln im Deutschen*. Berlin: Metzler. [CrossRef]
- Belz, Malte. 2023. Defining filler particles: A phonetic account of the terminology, form, and grammatical classification of “filled pauses”. *Languages* 8: 57. [CrossRef]
- Belz, Malte, and Myriam Klapi. 2013. Pauses following fillers in L1 and L2 German Map Task Dialogues. Paper presented at Workshop on Disfluency in Spontaneous Speech (DiSS 2013), Stockholm, Sweden, August 21–23; pp. 9–12.
- Belz, Malte, and Christine Mooshammer. 2020. *Berlin Dialogue Corpus (BeDiaCo). (Version 1)*. Berlin: Humboldt-Universität zu.
- Belz, Malte, and Uwe D. Reichel. 2015. Pitch Characteristics of Filled Pauses. Paper presented at 7th Workshop on Disfluency in Spontaneous Speech (DiSS 2015), Edinburgh, UK, August 8–9. pp. 1–4.
- Belz, Malte, Simon Sauer, Anke Lüdeling, and Christine Mooshammer. 2017. Fluently disfluent?: Pauses and repairs of advanced learners and native speakers of German. *International Journal of Learner Corpus Research* 3: 118–48. [CrossRef]
- Belz, Malte, and Jürgen Trouvain. 2019. Are ‘silent’ pauses always silent? Paper presented at International Congress of Phonetic Sciences (ICPhS), Melbourne, Australia, August 5–9; pp. 2744–48.
- Boersma, Paul, and David Weenink. 2022. Praat: Doing Phonetics by Computer. Available online: <https://www.fon.hum.uva.nl/praat/> (accessed on 20 October 2022).
- Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober, and Susan E Brennan. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech* 44: 123–47. [CrossRef] [PubMed]
- Braun, Angelika, and Annabelle Rosin. 2015. On the speaker-specificity of hesitation markers. In *International Congress of Phonetic Sciences (ICPhS)*. Glasgow: International Phonetic Association.
- Butterworth, Brian. 1975. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research* 4: 75–87. [CrossRef]
- Campione, Estelle, and Jean Véronis. 2002. A large-scale multilingual study of pause duration. Paper presented at International Conference on Speech Prosody, Aix-en-Provence, France, April 11–13, pp. 199–202.
- Candea, Maria, Ioana Vasilescu, and Martine Adda-Decker. 2005. Inter- and intra-language acoustic analysis of autonomous fillers. Paper presented at Workshop on Disfluency in Spontaneous Speech Workshop (DiSS 2005), Aix-en-Provence, France, September 10–12; pp. 47–52.
- Cataldo, Violetta, Loredana Schettino, Renata Savy, Isabella Poggi, Antonio Origlia, Alessandro Ansani, Isora Sessa, and Alessandra Chiera. 2019. Phonetic and functional features of pauses, and concurrent gestures, in tourist guides’ speech. *Audio Archives at the Crossroads of Speech Sciences, Digital Humanities and Digital Heritage* 6: 205–231.
- Clark, Herbert H., and Jean E. Fox Tree. 2002. Using uh and um in spontaneous speaking. *Cognition* 84: 73–111. [CrossRef]
- Clark, John, Colin Yallop, and Janet Fletcher. 2007. *An Introduction to Phonetics and Phonology*, 3rd ed. Malden: Blackwell Publishing.
- Corley, Martin, and Oliver W. Stewart. 2008. Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass* 2: 589–602. [CrossRef]
- Crystal, Thomas H., and Arthur S. House. 1990. Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America* 88: 101–12. [CrossRef] [PubMed]
- de Jong, Nivja H., and Ton Wempe. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41: 385–90. [CrossRef] [PubMed]
- de Leeuw, Esther. 2007. Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics* 19: 85–114. [CrossRef]
- Fox Tree, Jean E. 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language* 34: 709–38. [CrossRef]

- Fuchs, Susanne, and Amélie Rochet-Capellan. 2021. The respiratory foundations of spoken language. *Annual Review of Linguistics* 7: 1–18. [\[CrossRef\]](#)
- Garnier, Maëva, Lucie Bailly, Marion Dohen, Pauline Welby, and H el ene L evenbruck. 2006. An acoustic and articulatory study of Lombard speech: Global effects on the utterance. Paper presented at Annual Conference of the International Speech Communication Association (INTERSPEECH), Pittsburgh, PA, USA, September 17–21, pp. 2246–49. [\[CrossRef\]](#)
- Gerstenberg, Annette, Susanne Fuchs, Julie Marie Kairet, Claudia Frankenberg, and Johannes Schr oder. 2018. A cross-linguistic, longitudinal case study of pauses and interpausal units in spontaneous speech corpora of older speakers of German and French. Paper presented at International Conference on Speech Prosody, Pozna n, Poland, June 13–16, pp. 211–15. [\[CrossRef\]](#)
- Gick, Bryan, Ian Wilson, Karsten Koch, and Clare Cook. 2004. Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica* 61: 220–33. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gold, Erica, and Peter French. 2011. International practices in forensic speaker comparison. *International Journal of Speech, Language and the Law* 18: 293–307. [\[CrossRef\]](#)
- Gold, Erica, Peter French, and Philip Harrison. 2013. Clicking behavior as a possible speaker discriminant in English. *Journal of the International Phonetic Association* 43: 339–49. [\[CrossRef\]](#)
- Goldman-Eisler, Frieda. 1972. Pauses, clauses, sentences. *Language and Speech* 15: 103–13. [\[CrossRef\]](#)
- Goodwin, Charles. 1981. *Conversation Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.
- G osy, M aria, and Vered Silber-Varod. 2021. Attached filled pauses: Occurrences and durations. Paper presented at Workshop on Disfluency in Spontaneous Speech (DiSS 2021), Paris, France, August 25–26; pp. 71–76.
- Gully, Amelia J, Paul Foulkes, Peter French, Philip Harrison, and Vincent Hughes. 2019. The Lombard effect in MRI Noise. Paper presented at International Congress of Phonetic Sciences (ICPhS), Melbourne, Australia, August 5–9, pp. 800–4.
- Harrington, Lauren, Richard Rhodes, and Vincent Hughes. 2021. Style variability in disfluency analysis for forensic speaker comparison. *International Journal of Speech Language and the Law* 28: 31–58. [\[CrossRef\]](#)
- Hay, Jennifer, Ryan Podlubny, Katie Drager, and Megan McAuliffe. 2017. Car-talk: Location-specific speech production and perception. *Journal of Phonetics* 65: 94–109. [\[CrossRef\]](#)
- Hughes, Vincent, Sophie Wood, and Paul Foulkes. 2016. Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech, Language and the Law* 23: 99–132. [\[CrossRef\]](#)
- Ibrahim, Omnia, Ivan Yuen, Marjolein van Os, Bistra Andreeva, and Bernd M obius. 2022. The combined effects of contextual predictability and noise on the acoustic realisation of German syllables. *The Journal of the Acoustical Society of America* 152: 911–20. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jessen, Michael. 2007. Forensic reference data on articulation rate in German. *Science and Justice* 47: 50–67. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jessen, Michael. 2012. *Phonetische und Linguistische Prinzipien des Forensischen Stimmenvergleichs*. M unchen: Lincom.
- Jessen, Michael. 2018. Forensic voice comparison. In *Handbook of Communication in the Legal Sphere*. Edited by Jacqueline Visconti and Monika Rathert. Berlin: Mouton de Gruyter, pp. 219–55.
- Jessen, Michael, Olaf K oster, and Stefan Gfroerer. 2005. Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech Language and the Law* 12: 174–213. [\[CrossRef\]](#)
- Keating, Patricia, Marc Garellek, and Jody Kreiman. 2015. Acoustic properties of different kinds of creaky voice. Paper presented at International Congress of Phonetic Sciences (ICPhS). Number 1, Glasgow, UK, August 10–14, pp. 2–7.
- Kelley, Matthew C., and Benjamin V. Tucker. 2020. A comparison of four vowel overlap measures. *The Journal of the Acoustical Society of America* 147: 137–45. [\[CrossRef\]](#)
- Kjellmer, G oran. 2003. Hesitation. In defence of ER and ERM. *English Studies* 84: 170–98. [\[CrossRef\]](#)
- Klug, Katharina, and Marie K onig. 2012. Untersuchung zur sprecherspezifischen Verwendung von H asitationspartikeln anhand der Parameter Grundfrequenz und Vokalqualit at. In *Erforschung und Optimierung der Callcenterkommunikation*. Edited by Ursula Hirschfeld and Baldur Neuber. Berlin: Frank & Timme, pp. 175–93.
- Kohler, Klaus J. 1994. Glottal stops and glottalization in German: Data and theory of connected speech processes. *Phonetica* 51: 38–51. [\[CrossRef\]](#)
- Krech, Eva Maria. 1968. *Sprechwissenschaftlich-Phonetische Untersuchungen zum Gebrauch des Glottisschlageinsatzes in der Allgemeinen Deutschen Hochlautung*. Basel and New York: Kager.
- K unzel, Hermann J. 1987. *Sprechererkennung. Grundz uge Forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik Verlag.
- Kuznetsova, Alexandra, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2017. ImerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82: 1–26. .: 10.18637/jss.v082.i13. [\[CrossRef\]](#)
- Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2: 205–54. [\[CrossRef\]](#)
- Li, Xiaoting. 2020. Click-initiated self-repair in changing the sequential trajectory of actions-in-progress. *Research on Language and Social Interaction* 53: 90–117. [\[CrossRef\]](#)
- Lindblom, Bj orn E. F. 1968. Temporal organization of syllable production. In *Speech Transmission Lab. Quarterly Progress Status Report*. Stockholm: KTH Department of Speech, Music, and Hearing, vol. 9, pp. 1–5.
- Lo, Justin J. H. 2020. Between  ah(m) and euh(m): The distribution and realization of filled pauses in the speech of German-French simultaneous bilinguals. *Language and Speech* 63: 746–68. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lombard, Etienne. 1911. Le signe de l' levation de la voix. *Annales des Maladies de L'oreille et du Larynx* 37: 101–19.

- Maclay, Howard, and Charles E. Osgood. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word* 15: 19–44. [\[CrossRef\]](#)
- McDougall, Kirsty, and Martin Duckworth. 2017. Profiling fluency: An analysis of individual variation in disfluencies in adult males. *Speech Communication* 95: 16–27. [\[CrossRef\]](#)
- McDougall, Kirsty, and Martin Duckworth. 2018. Individual patterns of disfluency across speaking styles: A forensic phonetic investigation of Standard Southern British English. *International Journal of Speech, Language and the Law* 25: 205–30. [\[CrossRef\]](#)
- Muhlack, Beeke. 2020. L1 and L2 production of non-lexical hesitation particles of German and English native speakers. Paper presented at Workshop on Laughter and Other Non-Verbal Vocalisations, Bielefeld, Germany, October 5; pp. 44–47.
- Niebuhr, Oliver, and Kerstin Fischer. 2019. Do not hesitate!—Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance. Paper presented at Annual Conference of the International Speech Communication Association (INTERSPEECH), Graz, Austria, September 15–19; pp. 544–48. [\[CrossRef\]](#)
- O’Connell, Daniel C., and Sabine Kowal. 2005. Uh and um revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research* 34: 555–76. [\[CrossRef\]](#)
- O’Connell, Daniel C., and Sabine Kowal. 2008. *Communicating with One Another: Toward a Psychology of Spontaneous Spoken Discourse*. New York: Springer.
- Ogden, Richard. 2013. Clicks and percussives in English conversation. *Journal of the International Phonetic Association* 43: 299–320. [\[CrossRef\]](#)
- Ogden, Richard. 2020. Audibly not saying something with clicks. *Research on Language and Social Interaction* 53: 66–89. [\[CrossRef\]](#)
- Oliveira, Miguel. 2002. The role of pause occurrence and pause duration in the signaling of narrative structure. In *Advances in Natural Language Processing*. Berlin and Heidelberg: Springer, pp. 43–51. [\[CrossRef\]](#)
- Pätzold, Matthias, and Adrian Simpson. 1995. An acoustic analysis of hesitation particles in German. Paper presented at International Congress of Phonetic Sciences (ICPhS), Stockholm, Sweden, August 13–19, vol. 3, pp. 512–15.
- Pistor, Tillmann. 2017. Prosodische Universalien bei Diskurspartikeln. *Zeitschrift für Dialektologie und Linguistik* 84: 46–76. [\[CrossRef\]](#)
- Quené, Hugo. 2007. On the just noticeable difference for tempo in speech. *Journal of Phonetics* 35: 353–62. [\[CrossRef\]](#)
- R Core Team. 2022. R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*. Vienna: R version 4.1.3.
- Reitbrecht, Sandra. 2017. *Häsitationsphänomene in der Fremdsprache Deutsch und ihre Bedeutung für die Sprechwirkung*. Berlin: Frank & Timme.
- Roach, Peter J. 2009. *English Phonetics and Phonology: A Practical Course*, 4th ed. Cambridge, New York and Melbourne: Cambridge University Press.
- Rose, Philip. 2002. *Forensic Speaker Identification*. London: Taylor & Francis.
- Schmidt, Jürgen Erich. 2001. Bausteine der Intonation? *Germanistische Linguistik* 157–158: 9–32.
- Schulman, Richard. 1989. Articulatory dynamics of loud and normal speech. *Journal of the Acoustical Society of America* 85: 295–312. [\[CrossRef\]](#)
- Shriberg, Ee. 1994. Preliminaries to a Theory of Speech Disfluencies. Ph. D. thesis, University of California, Berkeley, CA, USA.
- Shriberg, Elizabeth. 2001. To ‘errrr’ is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* 31: 153–69. [\[CrossRef\]](#)
- Šimko, Juraj, Štefan Beňuš, and Martti Vainio. 2016. Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue. *The Journal of the Acoustical Society of America* 139: 151–62. [\[CrossRef\]](#) [\[PubMed\]](#)
- Simpson, Adrian P. 2007. Acoustic and auditory correlates of non-pulmonic sound production in German. *Journal of the International Phonetic Association* 37: 173–82. [\[CrossRef\]](#)
- Smiljanić, Rajka, and Ann R. Bradlow. 2009. Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass* 3: 236–64. [\[CrossRef\]](#)
- Smith, Vicki L., and Herbert H. Clark. 1993. On the course of answering questions. *Journal of Memory and Language* 32: 25–38. [\[CrossRef\]](#)
- Swerts, Marc. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30: 485–496. [\[CrossRef\]](#)
- Trouvain, Jürgen. 2004. *Tempo Variation in Speech Production. Implications for Speech Synthesis*. Phonus 8. Saarbrücken: Saarland University.
- Trouvain, Jürgen, and Zofia Malisz. 2016. Inter-speech clicks in an Interspeech keynote. Paper presented at Annual Conference of the International Speech Communication Association (Interspeech), San Francisco, CA, USA, September 8–12; pp. 1397–401. [\[CrossRef\]](#)
- Trouvain, Jürgen, and Raphael Werner. 2022. A phonetic view on annotating speech pauses and pause-internal phonetic particles. In *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion*. Edited by Cordula Schwarze and Sven Grawunder. Tübingen: Narr, pp. 55–73.
- Tuomainen, Outi, Linda Taschenberger, Stuart Rosen, and Valerie Hazan. 2021. Speech modifications in interactive speech: Effects of age, sex and noise type. *Philosophical Transactions of the Royal Society B* 377: 20200398. [\[CrossRef\]](#)
- Van Summers, W., David B. Pisoni, Robert H. Bernacki, Robert I. Pedlow, and Michael A. Stokes. 1988. Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America* 84: 917–28. [\[CrossRef\]](#) [\[PubMed\]](#)
- Whalen, D. H., Wei-Rong Chen, Christine H. Shadle, and Sean A. Fulop. 2022. Formants are easy to measure; resonances, not so much: Lessons from Klatt (1986). *The Journal of the Acoustical Society of America* 152: 933–41. [\[CrossRef\]](#) [\[PubMed\]](#)

- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain Francois, Garrett Golemund, Alex Hayes, Lionel Henry, Jim Hester, and et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4: 1686. [[CrossRef](#)]
- Wieling, Martijn, Jack Grieve, Gosse Bouma, Josef Fruehwald, John Coleman, and Mark Liberman. 2016. Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change* 6: 199–234. [[CrossRef](#)]
- Wohlert, Amy B., and Vicki L. Hammen. 2000. Lip muscle activity related to speech rate and loudness. *Journal of Speech, Language, and Hearing Research* 43: 1229–39. [[CrossRef](#)] [[PubMed](#)]
- Zellers, Margaret. 2022. An overview of discourse clicks in Central Swedish. Paper presented at Annual Conference of the International Speech Communication Association (Interspeech), Incheon, Republic of Korea, September 18–22; pp. 3423–27. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.