# The effect of Dutch (L1) and English (L2) bilingualism on the stability of voice quality parameters in cross-language voice comparison

Graaff, Donna de

# The effect of Dutch (L1) and English (L2) bilingualism on the stability of voice quality parameters in cross-language voice comparison

Donna de Graaff (4024451)

Faculty of Humanities, Leiden University

Master Thesis Linguistics: Applied Linguistics

Dr. Willemijn Heeren

June 22, 2025

## Abstract

Bilingualism and cross-language voice comparisons are becoming more prevalent in forensic cases. An international survey on forensic voice comparison practices among experts shows that voice quality (VQ) is mentioned most often as most useful for discriminating speakers. Studies on the discriminatory potential of VQ parameters are limited. Studies that compare VQ parameters across languages show contradictory results. The current study focused on how L1 Dutch and L2 English bilingualism influence the stability of voice quality parameters in cross-language voice comparison. This is done by comparing the schwa-like vowel in filled pauses (*uh, um*) of 35 speakers. The VQ parameters investigated are F0, jitter, shimmer, mean spectral energy and spectral tilt. The research question is: How does L1 Dutch and L2 English bilingualism influence the stability of voice quality parameters in cross-language voice comparison? The expectations are that jitter, shimmer, mean spectral energy and spectral tilt are unstable VQ parameters and that F0 is a stable VQ parameters when compared across Dutch and English voice comparisons. Linear mixed-effects models showed that spectral tilt is significantly influenced by language when compared across Dutch and English. Moreover, the VQ parameters F0, jitter, shimmer and mean spectral energy stay stable when compared across Dutch and English. In the discussion, the findings are discussed in relation to previous research and theories. Further research with other language combinations is necessary to better understand the influence of bilingualism and language on VQ parameters in cross-language voice comparisons.

## Contents

Introduction ..... 4
Theory ..... 6
Voice quality parameters in voice comparisons ..... 6
Bilingualism and VQ parameters in voice comparisons ..... 9
*Fundamental frequency* ..... 9
*Jitter and shimmer* ..... 11
*Glottal tension* ..... 12
Problem definition and relevance ..... 14
Methodology ..... 17
Speakers and recordings ..... 17
Acoustic analysis ..... 18
Statistical analysis ..... 20
Results ..... 23
Descriptive statistics ..... 23
*F0, jitter and shimmer* ..... 23
*Intensity, slope, tilt-linear and tilt-log* ..... 23
Correlations between variables ..... 26
Results of the linear mixed-effects models ..... 26
*F0, jitter and shimmer* ..... 26
*ST and MSE* ..... 27
*The influence of sex on VQ parameters* ..... 27
Discussion ..... 29
The effect of language on mean spectral energy ..... 29
The effect of language on fundamental frequency ..... 30
The effect of language on jitter and shimmer ..... 31
The effect of language on spectral tilt ..... 31
The influence of sex on VQ parameters ..... 33
Limitations and strengths ..... 33
Recommendations ..... 35
*Future research* ..... 35
*Practical implications* ..... 36
References ..... 37
Appendix ..... 43

## Introduction

In 2020, the Tamil Nadu police in India investigated an abduction case involving threatening phone calls made to the victim's family (Sundaram & Kannan, 2023). The police had difficulty identifying the caller among six suspects, so they turned to the forensic voice analysis team for help. The forensic voice experts noticed the use of the Tamil word "lanthu" which is jargon for "nuisance", commonly used in the Madurai region. The six suspects were recorded saying this word and the experts compared these samples with the disputed material. The experts focused on the pronunciation of this word, as the usage of the regional Madurai accent plays a key role in the identification. The experts specifically looked at the formants in the vowel sounds in "lanthu" (Sundaram & Kannan, 2023). Formants are frequencies in the vocal tract when vowels are pronounced (Rietveld & van Heuven, 2009). The voice analysis revealed that the pronunciation of one of the six suspects matched the caller's. This result became an important piece of evidence in this case that, in combination with other evidence, eventually led to a likely identification of the perpetrator (Sundaram & Kannan, 2023).

This case is an example of the use of forensic voice comparison in criminal cases and shows the intersection of bilingualism with forensic phonetics where pronunciation of vowels is influenced due to speaking different language varieties. In forensic voice comparison, the linguistic features are analysed and compared across the recordings of the offender's and the suspect's voice. In the case above, the results of the language analysis caused a breakthrough in the possible identification of the perpetrator. The pronunciation of "lanthu" by the suspects who do not speak Tamil from the Madurai region can be influenced by their first language variety. Sounds from the second language (L2) that are absent in the speaker's first language (L1), are often substituted with the closest L1 category, resulting in L2 productions that mirror similar but non-equivalent L1 productions (Lo, 2021). For example, the dental fricative /θ/ as the "th" as pronounced in the English word "thing" is not present in the Dutch language and L1 Dutch speakers are prone to pronounce this as a /t/ sound (Wester et al., 2007). Language can also influence certain acoustic parameters, for example the pitch of a speaker's voice when they speak one language versus another. A study by Theelen (2017) found that people speak Dutch with a higher fundamental frequency (F0) compared to English. F0 is often described as the pitch of a person's voice. Even though the two are closely related, they are not the same. The pitch represents what the ears and brains perceive, whereas the F0 is the actual physical phenomenon (Algemene Fonetiek, 2009, p. 215). Parameters like F0 are also influenced by sex. Male speakers have a lower average F0 than female speakers. Personal characteristics of speakers, like their sex, have to be taken into account when conducting

research on what is influencing certain acoustic parameters. The influence of language and bilingualism on certain acoustic parameters is not clear as previous research shows contradicting results (see Ng et al., 2012 and Zhu et al., 2022), which will be discussed in the following paragraphs.

Bilingualism and cross-language voice comparisons are becoming more prevalent in forensic cases (Lo, 2021). In the current study, the term *bilingualism* refers to someone speaking multiple languages and who learned those languages one after the other in different settings. These speakers are called sequential bilinguals (Edwards, 2012). For example, the mother tongue, or L1, is learned at home and the second language, or L2, is learned at school. In the current study, the term *bilingualism* does not refer to a balanced fluency in all languages but refers to the use of multiple languages that are learned to various levels of proficiency. More people are becoming bilingual and attitudes towards bilingualism are very positive among EU citizens with 86 percent in 2023 agreeing that everyone should speak at least one other language next to their mother tongue (European Commission, 2024). In 2023, English as a second language was spoken by 47 percent of Europeans and 93 percent of Dutch people (European Commission, 2024). The growing bilingualism and prevalence of bilingualism in cross-language voice comparisons indicate that research on forensic voice comparison should adopt a multilingual perspective and requires more knowledge on the influence of language on acoustic parameters (de Boer & Heeren, 2023).

The broad aim of the current study is to contribute to the knowledge on the stability of acoustic parameters across languages. In the next section, some important voice quality parameters used in voice comparisons will be discussed. Subsequently, the influence of language, bilingualism, and sex on voice quality parameters are discussed in more detail. The study focuses on five voice quality parameters: F0, jitter, shimmer, spectral tilt, and mean spectral energy. These are discussed in further detail in the next paragraph. Finally, the problem definition and relevance are discussed, and the research question and expectations are presented.

# Theory

## Voice quality parameters in voice comparisons

In forensic voice comparison you study and analyse the features of a voice in different recordings. Then, you compare for each feature what in the disputed versus comparison material happens with the feature and this will be categorised as a similarity or difference. Finally, this is interpreted against the same-speaker-hypothesis and the different-speaker-hypothesis (Nederlands Forensisch Instituut, 2016). The court seeks to answer the question if the offender's voice came from the suspect or from someone else. The forensic voice expert does not and cannot decide if the offender's voice in the disputed recording is the same as the suspect's voice on the comparison material or if it is someone else, however the expert expresses the probability of the results with a likelihood ratio. In the Netherlands this is done on a scale of probability: "about equally probable, a little more probable, more probable, much more probable, very much more probable, extremely much more probable" (Nederlands Forensisch Instituut, 2016). Each verbal term corresponds with a likelihood ratio. For example, the term "a little more probable" corresponds to a likelihood ratio of two to ten. This means that the probability of observing the research results is considered two to ten times higher when one hypothesis is true then when the other is true (Nederlands Forensisch Instituut, 2017). The judge decides, in combination with other evidence, if there is enough evidence to prove that the perpetrator is the same person as the suspect.

In forensic voice comparison, different acoustic parameters are used to compare voices on audio recordings. Voice quality (VQ) is found to be a robust and important feature in forensic voice comparison (Hughes et al., 2019; Lo, 2021). According to Nolan (1987), a feature is robust when it is speaker-specific. This means that it has low within-speaker variation and high between-speaker variation (Nolan, 1987). It also means that a feature stays stable over different circumstances, for example recording quality. An international survey on forensic voice comparison practices among experts shows that VQ is mentioned most often as most useful for discriminating speakers (Gold & French, 2011). However, when recording qualities are moderate to bad, it can become difficult to determine the VQ under the noise (Hughes et al., 2019).

Voice quality is a broad term and there exists some ambiguity around the assessment of this acoustic feature (San Segundo & Gomez-Vilda, 2014). VQ is mostly associated with the phonetic description of Laver (1980) who defines VQ as the combination of laryngeal and supralaryngeal characteristics in an individual's voice that create a lasting impression on perception which makes that voice distinguishable from others. The ambiguity lies in the

assessment of VQ as this can be done with a perceptual, articulatory or acoustic assessment (San Segundo & Mompean, 2017). The acoustic assessment consists of featural analysis (San Segundo & Mompean, 2017), however, there is no consensus about which features are the most robust for voice comparisons. A study by Jessen (1997) investigated the speaker-specificity of VQ parameters in the vowel /a/ produced by twenty male German speakers. The results show that the VQ parameters that are most speaker-specific are fundamental frequency (F0) and the amplitude differences between the first and second harmonic (H1-H2) (Jessen, 1997). The features are reflections of what the speaker produces with their speech organs (vocal tract, larynx, vocal folds) (Jessen, 1997).

F0 is, among others, one of the features that is most commonly used by experts in forensic voice comparison (Gold & French, 2011). F0 represents the rate of vibration of the vocal folds. When a speech sound is produced, an air stream flows from the lungs to the mouth or nose and this air stream can be manipulated by vibrating the vocal folds (Theelen, 2017). The length of a person's vocal folds affects their F0. Typically, longer vocal folds lead to a lower F0 and shorter vocal folds lead to a higher F0 (Algemene Fonetiek, 2009, p. 39). This is one of the physiological aspects that influence the difference in average F0 between males and females. When boys and girls grow into adulthood, males' F0 drops to an average of 120 Hz and females' F0 drops to an average of 220 Hz (Kreiman & Sidtis, 2011). As explained before, F0 is related to the pitch of a person's voice, but they are not the same. F0 and pitch are both expressed in Hertz (Hz). The amplitude differences between the first and second harmonic represent a ratio of energy between the lower and upper harmonic. Harmonics are frequencies that are a multiple of the F0 (Algemene Fonetiek, 2009, p. 141). When the F0 is 100 Hz, the first harmonic is 200 Hz and the third harmonic is 300 Hz etc. More prominent upper harmonics than lower harmonics correspond to hyperadduction of the vocal folds (Ng et al., 2012). The adduction of vocal folds means that they move closer together, which means that the vocal folds are stiff and this indicates glottal tension which can produce a creaky voice sound (Algemene Fonetiek, 2009, p. 40; Chan, 2023). The abduction means that the vocal folds move away from each other, which indicates less glottal tension, and this can produce a breathy voice sound (Algemene Fonetiek, 2009, p. 40; Chan, 2023).

Glottal source features, like glottal tension, larynx position and phonation types, appear to be robust features (San Segundo & Gomez-Vilda, 2014). Anatomical and physiological variation between speakers show that there is a lot of room for between-speaker and within-speaker variation in laryngeal VQ (van Hugte & Heeren, 2024). The Laryngeal Articulator Model explains the different components and variations of the larynx and

describes the larynx as a complex and multifaceted articulator (Esling et al., 2019). Not only the vibration of the vocal folds is at play, but there is also an interplay of six laryngeal components that lead to phonation and there is between-speaker variation in the anatomy of these components (Esling et al., 2019). Variation is possible in the size of the larynx, movement of the pharyngeal walls, or speakers can have asymmetrically shaped ventricular folds (Casper et al., 1987). A laryngeal feature of VQ is creaky voice. One study on creaky voice shows that it has distinctive profiles (van Hugte & Heeren, 2024). The results of this study show that there is overall variation between Dutch male speakers in creaky voice, meaning that this parameter is speaker-specific. However, this variation showed to be very low between any pair of speakers (van Hugte & Heeren, 2024).

Additionally, Hughes and colleagues (2019) tested the robustness of VQ parameters in forensic voice comparisons across different recording qualities: studio, landline telephone, and mobile telephone recordings. They found that F0 together with spectral tilt (ST) and additive noise measures show discriminatory ability and are robust to variation between high quality studio recordings, landline telephone recordings and mobile phone recordings (Hughes et al., 2019). ST is a parameter of glottal tension and it is a ratio of energy between the lower harmonics and the upper harmonics and represents the rate at which amplitude of the harmonics declines in the LTAS contour (Ng et al., 2012). An LTAS is a long-term average spectrum. It represents the logarithmic of the average power in a sound during a certain time range and in a certain frequency range (Boersma & Weenink, 2024). This means that a low ST indicates more prominent upper harmonics than lower harmonics and corresponds to hyperadduction of the vocal folds (Ng et al., 2012). A lower ST means more glottal tension and indicates a louder voice, a more prominent syllable or a creakier voice, while a higher ST means less glottal tension and indicates a breathier voice (Chan, 2023). While the quality of the parameters decreased when the quality of the type of recording also decreased, this decrease stayed relatively low (Hughes et al., 2019). The most robust parameter across different types of recordings appeared to be ST (Hughes et al., 2019). Generally, females have a breathier voice than males (Kreiman & Sidtis, 2011; Mendoza et al., 1996). Due to physiological differences between the vocal folds of males and females, most females' vocal folds do not close completely during each phonatory cycle and this causes a glottal gap which leads to an aspiration noise (Kreiman & Sidtis, 2011).

In contrast to the findings by Hughes and colleagues (2019), a study by Chan (2023) states the opposite. Chan (2023) looked into the same VQ parameters as Hughes and colleagues (2019), however he tested the robustness of VQ regarding speech style mismatch

and non-contemporaneous recordings instead of recording quality and method. These VQ parameters are the amplitude differences between the first and second harmonics (H1-H2) and second and fourth harmonics (H2-H4). Also, he looked at the amplitude differences between the first harmonic (H1) and the spectral magnitude at the first formant (A1), second formant (A2) and third formant (A3) which are written as H1-A1, H1-A2, and H1-A3. The last two VQ parameters tested are cepstral peak prominence (CPP) and harmonics-to-noise ratio (HNR). A larger CPP is an indication of a more modal voice and a smaller CPP means a breathier voice. HNR measures the spectral noise level and shows the degree of perceived breathiness (Chan, 2023). With the involvement of non-contemporaneous recordings and speech style mismatch these VQ parameters have low speaker-specificity. Chan (2023) advises forensic analysts to be cautious when using these VQ parameters as speaker discriminants in certain circumstances.

**Bilingualism and VQ parameters in voice comparisons**

It is unclear how robust VQ parameters are when compared across languages (Zhu et al., 2022). To be a robust feature, the VQ parameters have to be more speaker-dependent than language-dependent. This means that there should be almost no variation between the VQ parameters in different languages. In this paragraph, some contested VQ parameters in cross-language voice comparison are discussed.

***Fundamental frequency***

F0 has been perceived as a discriminatory feature in voice comparison analysis (Jessen, 1997; Hughes et al., 2019; Lo, 2021). Nonetheless, in voice comparison across languages there are some opposing results. Zhu and colleagues (2022) investigated the influence of language on the VQ of American and Chinese bilingual speakers. The results show no variation of F0 between speaking English and speaking Mandarin (Zhu et al., 2022). This indicates that F0 could be a robust VQ parameter when compared across languages. Yet, Zhu and colleagues (2022) argue that previous studies on the influence of language on VQ parameters show contradictory results. An earlier study by Ng and colleagues (2012) investigates the influence of language on the VQ of 40 L1 Cantonese and L2 English bilingual speakers using the same VQ parameters as Zhu and colleagues (2022) and their results show variation in F0 between speaking English and Cantonese (Ng et al., 2012). Another study on bilingual speakers of Cantonese and English found no variation in F0 between speaking Cantonese and English (Altenberg & Ferrand, 2006).

Moreover, research by Cantor-Cutiva and colleagues (2021) on the influence of bilingualism on VQ parameters comparing bilingual Spanish-English speakers with monolingual Spanish speakers show that bilingual Spanish-English speakers have a lower F0 on both Spanish and English than monolingual Spanish speakers have. This finding is explained by the speech accommodation theory which argues that speakers accommodate their vocal behaviour, like intonation, accent, and speech rate, when interacting with native speakers to match their speech to build rapport and to reduce the social distance with the interlocutor (Giles et al., 1987). Cantor-Cutiva and colleagues (2021) go a step further and suggest that bilingual Spanish-English lower their F0 to match the F0 of native English speakers, and once this decreased F0 is incorporated in the speaker's muscle memory, it is also produced when speaking their first language (Spanish), suggesting influence of the L2 on the L1 (Cantor-Cutiva et al., 2021). This is an example of how bilingualism can influence acoustic parameters of bilingual speakers. Another study by Järvinen and colleagues (2013) shows variation in F0 between speaking Finnish and English by Finnish-English bilingual speakers and they suggest that bilingual speakers try to match the perceived pitch of native speakers of their L2, which aligns with the arguments of Cantor-Cutiva and colleagues (2021).

Cross-language research between Dutch and English is limited, except for the study by Theelen (2017). In this study, it was found that people speak Dutch with a higher F0 than English, compared between Dutch native speakers and English native speakers. English native speakers spoke Dutch with a significantly higher F0 than when they spoke English (Theelen, 2017). There is, however, no significant difference in F0 when a Dutch native speaks English compared to when they speak Dutch (Theelen, 2017). This could be an indication of vocal behaviour from the first language being transferred to the second language. Vocal behaviour of someone's L1 can have a large influence on their L2 (Lo, 2021). Dutch speakers might not lower their F0 when speaking English to try to match the perceived pitch of English natives because their F0 in their L1 is transferred to their L2, causing the F0 to stay stable across languages. Some studies found F0 to be a stable parameter across languages (Altenberg & Ferrand, 2016; Ng et al., 2012), but others argue that there could be an effect of language on F0 (Chan, 2023; Theelen, 2017; Zhu et al., 2022). Therefore, this parameter is included in the current study.

***Jitter and shimmer***

Not only is there uncertainty about the robustness of F0, but there is also uncertainty about the robustness of jitter and shimmer in cross-language voice comparisons. Jitter and shimmer reflect the regularity of vocal fold vibration during speech (Zhu et al., 2022). Jitter reflects the regularity of the cycle-to-cycle F0 of the voice and shimmer reflects the regularity of the amplitude of each cycle (Cantor-Cutiva et al., 2021). More jitter and shimmer indicate more variation in rate (F0) and excursion (amplitude) of vocal fold vibration (Zhu et al., 2022). A reason for fluctuations is because the rate of vibration of the vocal folds is not stationary. Since it originates from an organic structure, it can fluctuate (Collins & Mees, 2003). Another reason for fluctuations is because the speaker wants it to fluctuate to establish intonation and meaning, for example going up in pitch when asking a question (Collins & Mees, 2003).

Research on the stability of jitter and shimmer in bilingual speakers across languages is limited, although they are included more often in recent literature (Cantor-Cutiva et al., 2021; Zhu et al., 2022). The study by Cantor-Cutiva and colleagues (2021) on differences between bilingual speakers of L1 Spanish and L2 English and monolingual Spanish speakers shows variation in jitter and shimmer between bilingual and monolingual speakers. Male bilingual speakers had higher jitter than male monolingual speakers and female bilingual speakers had lower jitter and shimmer than female monolingual speakers. Cantor-Cutiva and colleagues (2021) argue that this variation in jitter and shimmer between bilingual and monolingual speakers is because language differences affect laryngeal characteristics when speakers switch between languages. They do note that, because there is an effect of sex on jitter and shimmer, it is important for future studies to take this into account and investigate if this affects the influence of language on the VQ parameters (Cantor-Cutiva et al., 2021).

Moreover, Zhu and colleagues (2022) found that there is an increased jitter and shimmer between speaking English and Mandarin when American English speakers and Mandarin speakers were speaking their mother tongue compared to their L2 (Zhu et al., 2022). This aligns with the speech accommodation theory as speakers might try to match the perceived pitch variation and intonation of native speakers of their L2 to build rapport and reduce social distance (Giles et al., 1987). The use of intonation is known to differ between languages and is used to create meaning (Almbark et al., 2014; Collins & Mees, 2003). This suggests that the jitter and shimmer of bilingual speakers are different between their L1 and their L2 because the L2 is adjusted to the perceived pitch variation of native speakers of the L2 language, to establish the correct use of intonation and meaning.

***Glottal tension***

In addition to the uncertainty about the robustness of glottal tension as a VQ parameter in voice comparisons within a language (Chan, 2023; Hughes et al., 2019), there is also uncertainty about this parameter in voice comparisons across languages. Glottal tension can be measured by spectral tilt (ST) and mean spectral energy (MSE). ST is a ratio of energy between the lower harmonics and the upper harmonics and represents the rate at which the amplitude of the harmonics declines in the LTAS contour (Ng et al., 2012). The value of MSE shows the average energy (amplitude) within a certain frequency range and it measures the intensity of the speech when a certain vowel is pronounced (Zhu et al., 2022). When there is higher glottal tension, there is a higher MSE (Zhu et al., 2022).

In the studies by Ng and colleagues (2012) and Zhu and colleagues (2022), MSE showed to be an unstable parameter across languages. Also, research by Bahmanbiglu and colleagues (2017) shows variation in MSE between Farsi and Qashqai Turkish. The effect of language on MSE is explained by Bahmanbiglu and colleagues (2017) and Ng and colleagues (2012) by different patterns in resonance between languages. Earlier research by Kerr (2000) on accent modification, showed that more glottal tension leads to posterior resonance. Kerr (2000) found that people tend to resonate more in the back of the mouth when speaking English and more posterior resonance was used when people speak Cantonese, while Mandarin and English show a similar resonance pattern.

Not only is MSE argued to be influenced by different resonance patterns, but also ST can be influenced by this. ST was found to be a robust VQ parameter when compared between Mandarin and English bilingual speakers (Zhu et al., 2022). However, the study by Ng and colleagues (2012) showed that ST was not stable when compared between Cantonese and English bilingual speakers. Different resonance patterns between Cantonese and English is argued to explain the instability of ST (Ng et al., 2012), while similar resonance patterns between Mandarin and English is argued to explain the stability of ST in the study by Zhu and colleagues (2022).

Research on resonance patterns between Dutch and English is limited, but there is no indication that Dutch and English use posterior resonance patterns and both languages show similar anterior and mid-vocal resonance patterns in the formant frequencies of vowels (Collins & Mees, 2003). This suggests that Dutch and English might share comparable resonance patterns which could lead to similar MSE and ST values across both languages, however the use of laryngealization is more common in English than in Dutch (Collins & Mees, 2003). Laryngealization is the use of more glottal tension and produces creaky voice

(Nguyen & Kenny, 2009). Creaky voice is frequently used in English, whereas it is very uncommon in Dutch (Collins & Mees, 2003). According to Collins and Mees (2003, p. 86) “Dutch-speaking learners are advised to imitate creaky voice in order to make their English accents more convincing”. This could lead to a greater difference in MSE and ST between Dutch and English.

A creakier voice is caused by more glottal tension, however, glottal tension does not always indicate creaky voice. More glottal tension can also be measured when a voice is louder or when there is a more prominent syllable (Monsen et al., 1978; Titze, 2023). Creaky voice is a particular voice and often studied on its own. Aside from little research about glottal tension, there is more research about creaky voice and the differences between languages. Benoist-Lucy and Pillot-Loiseau (2013) found differences in creaky voice between speaking English and French. Creaky voice is argued to be language-dependent because of different social meanings and relevance of creaky voice in different languages (Benoist-Lucy & Pillot-Loiseau, 2013). For instance, a language-specific effect of creaky voice in American English is that it indicates taking an authoritative stance (Sicoli, 2015) and in British English it indicates vowel hiatus (Davidson & Erker, 2014). This is not the case in Dutch, which could lead to lesser use of creaky voice in Dutch and, thus, less glottal tension in Dutch than in English.

These findings support the theory of Bruyninckx and colleagues (1994), that different languages are associated with different voice qualities. Bruyninckx and colleagues (1994) studied the influence of bilingualism on LTAS parameters, like ST and MSE, between Spanish and Catalan. They argue for a language effect on VQ parameters. Bruyninckx and colleagues (1994) describe this effect as the association of different languages to different modes of phonation and vocal parameters, while using the same vocal apparatus. Modes of phonation refer to the typical vocal fold settings and surrounding structures during speech (Laver, 1980). These settings influence how voice is produced and have a direct effect on voice quality parameters such as ST, jitter, shimmer, and MSE (Gerrat & Kreiman, 2001; Kreiman & Sidtis, 2011). This suggests that an individual speaker using their vocal apparatus will produce a different ST and MSE when speaking one language than when speaking another language, because different languages are associated with a different range of glottal tension.

Additionally, research by Cantor-Cutiva and colleagues (2023) found variation in vocal fry between speaking English and Spanish with more vocal fry when speaking English than Spanish. Vocal fry is a type of creaky voice (Keating et al., 2023). Native bilingual American English speakers used more vocal fry when speaking English and Spanish than

native bilingual Latin-American Spanish speakers. Cantor-Cutiva and colleagues (2023) argue that this may be an indication of transferring vocal behaviour from the first languages to the second language which would reduce the language dependency of ST. Lo (2021) has also argued that there is VQ transfer from the more dominant language to the weaker language. This means that creaky voice and breathy voice are also notable in the L2 as they are in the L1 of the speaker (Lo, 2021).

Not only creaky voice and vocal fry could cause ST and MSE values to be similar in the L1 and the L2, but also stress and intonation patterns could lead to similar ST and MSE values across languages. Putting stress to get more prominent syllables or words is caused by more glottal tension (Monsen et al., 1978; Titze, 2023). The acquisition of the correct word-level and phrase-level stress in the L2 has been found to be very challenging (Almbark et al., 2014). Depending on the ability of the speaker, new vocal behaviour is acquired or it is substituted by a known vocal behaviour from the L1. According to the notion of vocal behaviour transfer from the L1 to the L2, the measurements of glottal tension should not vary much across languages, indicating that glottal tension could be a stable VQ parameter in cross-language voice comparisons. Even though van Hugte and Heeren (2024) found creaky voice to allow for moderate speaker distinction between Dutch speakers, it is uncertain if this speaker distinction can still be made when comparing speakers' VQ across languages.

It is notable that Zhu and colleagues (2022) found variation in MSE but no variation in ST while they are both measurements of glottal tension. Although both MSE and ST are associated with glottal tension, their sensitivity to cross-linguistic variation may differ. ST is more directly influenced by subtle phonatory and resonance shifts (Hanson, 1997; Ng et al., 2012), whereas MSE reflects overall vocal energy and is therefore less sensitive to resonance-related differences between languages or speaker-specific variation (Gobl & Ní Chasaide, 2003; Kreiman et al., 2021). Consequently, the current study measures glottal tension by using both parameters.

**Problem definition and relevance**

The increasing prevalence of bilingualism and cross-language voice comparisons in forensic cases (Lo, 2021), show that research on forensic voice comparison should adopt a multilingual perspective and requires more knowledge on the influence of language on acoustic parameters (de Boer & Heeren, 2023). Lo (2021) argues that studies on the discriminatory potential of VQ parameters are limited. Existing studies show that certain VQ parameters are robust in forensic voice comparisons (Hughes, 2019; Lo, 2021). However,

studies that compare VQ parameters across languages show contradictory and ambiguous results (Cantor-Cutiva, 2021; Zhu et al., 2022). Due to the contradictory findings, it is still unclear how language influences the VQ. The effect of language on VQ between speaking English and Mandarin was recently investigated by Zhu and colleagues (2022). They state that their study is limited by the small sample group and advise future research to look into the influence of language on VQ in different language combinations (Zhu et al., 2022). Furthermore, Zhu and colleagues (2022) focus specifically on differences between languages and not on differences between individuals, which is a relevant part for forensic voice comparison and is taken into account in the current study.

Subsequently, the current study focuses on the influence of bilingualism on the stability of VQ parameters. The languages included are L1 Dutch and L2 English. This is relevant information for forensic voice comparisons where bilingualism and cross-language voice comparisons are becoming more prevalent. The study aims to contribute to the existing literature about the robustness of VQ parameters in cross-language (forensic) voice comparisons. The VQ parameters investigated are F0, jitter, shimmer, MSE and ST.

Based on the findings of previous literature (see Altenberg & Ferrand, 2006; Ng et al., 2012; Theelen, 2017; Zhu et al., 2022), I expect Dutch-English bilingual speakers to transfer their F0 from their L1 to their L2, in the current study. This means that I expect the stability of F0 to not be influenced by Dutch-English bilingualism. Furthermore, there have been limited results about the stability of jitter and shimmer across languages (see Cantor-Cutiva et al., 2021; Järvinen et al., 2013; Zhu et al., 2022). Based on previous literature (see Giles et al.,1987; Järvinen et al., 2013; Zhu et al., 2022) I expect Dutch-English bilingual speakers to adjust the variation of their F0 in their L2 (English) to the perceived variation of F0 of native speakers of English. This means that the jitter and shimmer of Dutch-English bilingual speakers in Dutch are different from their jitter and shimmer in English. Therefore, in the current study I expect the stability of jitter and shimmer to be influenced by Dutch-English bilingualism.

Glottal tension has been measured by spectral tilt and mean spectral energy. ST shows contradicting results while MSE shows clear variation across languages (see Bahmanbiglu et al., 2017; Ng et al., 2012; Zhu et al., 2022). However, it is unclear why MSE differs across languages as English and tonal languages, like Mandarin, are both associated with laryngealization, which is caused by a more tense glottis (Kenny & Nguyen, 2009). Therefore, the current study includes both parameters to measure glottal tension. Based on the findings of previous literature (see Bahmanbiglu et al., 2017; Ng et al., 2012; Zhu et al., 2022) and the

difference in the use of laryngealization between English and Dutch (Collins & Mees, 2003), I expect more glottal tension when Dutch-English bilinguals speak English than when they speak Dutch. This means that I expect the measurements of ST and MSE to show variation across Dutch and English. Furthermore, as sex is an influential characteristic on the VQ, it is important to take this into account in the current study and to investigate if sex influences the effect of language on voice quality.

Subsequently, the research question of the current study is:

How does L1 Dutch and L2 English bilingualism influence the stability of voice quality parameters in cross-language voice comparison?

There are different expectations for F0 than for jitter, shimmer, ST and MSE. The expectation for F0 is that the stability of F0 is not influenced by Dutch-English bilingualism. The expectation for jitter, shimmer, ST and MSE is that the stability of jitter, shimmer, ST and MSE is influenced by Dutch-English bilingualism.

## Methodology

The stability of VQ parameters across languages was tested through an acoustic analysis followed by a statistical analysis. For this research, audio recordings from 35 speakers from the database of the D-LUCEA Accent Project were used (Orr & Quené, 2017). The D-LUCEA Accent Project collected data from speakers of L1 Dutch and L2 English to study the convergence of accents (Orr & Quené, 2017). The current study used the recordings of (semi)spontaneous English and Dutch monologues by the speakers. In the following sections the speaker sample and recordings, acoustic analysis and statistical analysis are described and explained.

### Speakers and recordings

The participants of the D-LUCEA Accent Project were students from University College Utrecht (UCU). They were recorded at multiple moments during their three year study and this was done in four consecutive cohorts over a six year period (Orr & Quené, 2017). The current study used recordings from participants from the first cohort which took place within one month after arrival on campus (Orr & Quené, 2017). Recordings from the first recording session of each speaker were used because the multilingual environment of UCU has most likely not had any effect on the speaker's accents yet (de Boer & Heeren, 2020). The students included in this study are L1 Dutch speakers and L2 English speakers. Their level of English is estimated to be at least B2 level according to the Common European Framework of Reference for Languages (Council of Europe, 2019), as the students are required to score an 8 out of 10 for English in high school to be accepted at UCU (Quené et al., 2017). A total of 35 speakers were included in this study of which 10 are male and 25 are female and from each speaker a recording of them speaking English and Dutch was used.

The recordings took place in August 2011 in a quiet, furnished office. The people present at the recordings were a speaker participant and one or more facilitators. Recordings were made on eight different channels (Orr & Quené, 2017). The recordings used in the current study were recorded on a headset microphone worn by the speaker participant. The speakers were asked to do different speaking tasks. The tasks in the recordings that were used in the current study were a two minute monologue in L1 (Dutch) on an informal free topic and a two minute monologue in English on an informal free topic (Orr & Quené, 2017). The decision for using recordings of these tasks was made because they were the only (semi)spontaneous speech tasks instead of read speech, and (semi)spontaneous speech is closer to real cases in forensic voice comparison than read speech. Most studies use read

speech or the researcher asks the participant to sustain a long vowel, however in forensic voice comparison cases, experts usually have to work with spontaneous speech (San Segundo & Gomez-Vilda, 2014).

**Acoustic analysis**

The acoustic analysis was done in the program Praat (Boersma & Weenink, 2024). The parameters that were analysed are F0, jitter, shimmer, mean spectral energy and spectral tilt. The script that was run to measure these parameters can be viewed in the Appendix (p.c., Heeren, 2025). Before measuring the parameters, the audio recordings were analysed and prepared. The parts of the audio recordings that were analysed were the vowels in filled pauses (*uh*, *um*) that lasted at least 160 milliseconds. In both (British) English and Dutch this vowel is described as a lengthened schwa vowel /ə/ (Hughes et al., 2016; Stouten & Martens, 2003). In total, there were 1051 tokens of this schwa-vowel from the 35 participants included in the dataset. Previous literature differ in what they included in their samples. Some only removed unvoiced segments and included all voiced segments (Ng et al., 2012; Zhu et al., 2022), others used vowel-only samples (Chan, 2023; Hughes et al., 2019), and Lo (2021) included vowels and semivowels because of their vowel-like qualities. According to San Segundo and Gomez-Vilda (2014) and Boersma and Weenink (2024), it is important to use long sustained vowels in order to measure VQ parameters. San Segundo and Gomez-Vilda (2014) took tokens of the /e:/ vowel that lasted around 160 milliseconds which were naturally sustained in pause fillers. Furthermore, measuring certain VQ parameters is typically only performed on long sustained vowels, like jitter and shimmer (Boersma & Weenink, 2024). In a previous study by Cantor-Cutiva and colleagues (2021), jitter and shimmer were measured from tokens of the long sustained vowel /a:/. These vowels are usually unnaturally sustained, except for the vowels in filled pauses like in "*uh"* and *"um*" and some vowels are naturally a bit prolonged. The intervals of these vowels were annotated in Praat and the intervals of the filled pauses (*uh, um*) were partially annotated in the recordings by de Boer for the study of the acoustics of "*uh"* and "*um"* in native Dutch and non-native English (de Boer & Heeren, 2020) and partially annotated by the researcher of the current study.

F0 can be measured in Praat by using various methods, however raw cross-correlation is the preferred method in voice analysis (Boersma & Weenink, 2024). All of these methods are based on the waveform's self-similarity. If the waveform is nearly identical when shifted by 10 milliseconds in time, it means that 100 Hz is suitable as the F0 (Boersma & Weenink, 2024). If the waveform is self-similar over a time-shift of 5 ms, it means that 200 Hz is

suitable as the F0 and if the waveform is self-similar over a time-shift of 2.5 ms it gives 400 Hz as an F0 candidate etc. In the current study, raw cross-correlation was used to measure F0. Aside from being the advised technique for voice analyses (Boersma & Weenink, 2024), this was also used in previous studies to measure F0 (see van Hugte & Heeren, 2024). The minimum and maximum thresholds of the F0 were set between 30 Hz and 600 Hz. This is a broad range to include both female and male speakers and is recommended by Boersma and Weenink (2024).

To measure jitter and shimmer, a script was run to get the voice report of the selected vowel sounds (see Appendix). The voice report shows multiple jitter and shimmer measurements. The different measures are all based on the computation of all periods by the wave-form matching procedure (Boersma & Weenink, 2024). According to Farrús and Ejarque (2007) these measures show a high correlation. The parameters that were looked at in the current study are jitter (local) and shimmer (local), as was done in previous studies (see Cantor-Cutiva et al., 2021; Zhu et al., 2022). By following previous studies, the results of the current study will be more easily comparable to previous results. Jitter (local) measures the average absolute difference between consecutive periods, divided by the average period (Farrús & Ejarque, 2007). It is expressed as a percentage. Shimmer (local) measures the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude (Farrús & Ejarque, 2007).

Glottal tension was measured by MSE and ST and they are measurements that quantify the spectral analysis. A long term average spectrum (LTAS) was done after a Hann band filter over the same vowels as was done with jitter and shimmer of each speaker in both English and Dutch to investigate how acoustic energy is distributed across frequency. Previous literature has measured MSE as the average amplitude across the frequency range of 0–8 kHz, and it was expressed as a negative value relative to the highest amplitude peak within the frequency range (Ng et al., 2012). In the current study, mean spectral energy was measured as the average intensity over the same frequency range of 0-8 kHz and expressed in decibels (dB). Spectral tilt has been measured in various ways. Ng and colleagues (2012) and Zhu and colleagues (2022) calculated ST by dividing the sum of amplitudes between 0 and 1 kHz and that between 1 and 5 kHz (Ng et al., 2012; Zhu et al., 2022). Other previous studies have measured ST by harmonic differences (Chan, 2021; Hughes et al., 2018; van Hugte & Heeren, 2024). In the current study ST was measured using the method and settings of Ng and colleagues (2012) and Zhu and colleagues (2022) by getting the slope of the LTAS in Praat (Boersma & Weenink, 2024). Additionally, the current study calculated the logarithmic and

the linear ST to see if this renders different results on the stability of ST across languages. The linear tilt measure looks at energy differences more evenly across all frequencies than the slope and log-transformed tilt measure. The slope and log-transformed tilt measure are more sensitive to the distribution of energy in higher frequency ranges than the linear tilt measure. This makes the slope and log-transformed tilt measure more sensitive to subtle articulatory and phonatory variations between languages than the linear tilt measure (Hillebrand et al., 1994; Stevens, 1998; Zhu et al., 2023).

**Statistical analysis**

All VQ parameters were tested in statistical models. The statistical analysis and calculations were performed using R Statistical Software (R Core Team, 2024). The effect of language on VQ were tested by performing a linear mixed-effects model (LME) in R using the *lme4* package (Bates, Maechler, & Bolker, 2012). The independent variables were language, which consisted of the levels 'Dutch' and 'English', and sex, which consisted of the levels 'male' and 'female'. The dependent variables were the VQ parameters. The five parameters were F0, jitter, shimmer, MSE, and ST. These were all continuous variables. For each parameter, an LME model was built which resulted in seven models with dependent variables: F0, jitter, shimmer, intensity, slope, ST-linear, and ST-log.

Before performing the analyses, all seven models had been tested to see whether they met the assumptions of linearity, absence of homoscedasticity, absence of multicollinearity and normality of residuals. The assumption of linearity was not explicitly checked, however if the other assumptions were met then it can be assumed to be unproblematic (Winter, 2017). A visual inspection of residual plots of all models did not show any obvious deviations from homoscedasticity. The Variance Inflation Factor (VIF) values of all models showed that there was an absence of multicollinearity. The VIF values show whether multiple independent variables measure the same underlying effect (Field, 2017). With a VIF value above five, there is a problem in estimating the regression coefficient of the variable (Field, 2017). The tables showed that the VIF values were below five, which means that there was no multicollinearity.

However, the Q-Q plots and histograms of the models indicated a small deviation from a normal distribution of residuals. The Shapiro Wilkinson test showed that some models have a p-value just below .05 which means that the residuals significantly deviate from a normal distribution. This can be solved by including the log of the independent variables in the analysis, however there are disadvantages to transforming data. Transforming variables results

in measuring a different construct from the original. This has implications for interpreting the data and hypotheses have to be changed. Also, the consequences of using the wrong transformation could be worse than the consequences of analysing untransformed data (Field, 2017, p. 270). Additionally, when comparing the Q-Q-plots and the results of the Shapiro Wilkinson test of the transformed data it became clear that for jitter and shimmer the transformed data resulted in a more normal distribution of the residuals, so these were used in the LME models of jitter and shimmer. However, for F0, intensity, slope, ST-linear, and ST-log it was better to not transform the data. The Q-Q-plots and Shapiro Wilkinson tests showed a worse distribution of residuals with the transformed data in these models. So in these models the untransformed data were used.

Furthermore, the normality of the scale variables was tested. The histograms showed that jitter, shimmer and F0 did not have a normal distribution and the boxplots showed that F0, jitter and shimmer had a few outliers. The outliers were not excluded from the data as they were not extreme outliers. The jitter and shimmer outliers were well below the threshold of pathology and the F0 outliers were within the ranges of a high-pitched average female voice and a low creaky voice according to Boersma and Weenink (2024). Measures of creaky voice were included in the dataset as they did not influence the effect of language on F0. This was tested by comparing the models of these VQ parameters with models excluding the creaky voice measurements. The threshold for creaky voice was a value of more than two standard deviations from the mean for females and males separately, and 3.3% of the dataset consisted of creaky voice measurements. The Shapiro Wilkinson test showed that all scale variables have a p-value smaller than .05. This means that the data were not normally distributed, so the correlation between the scale variables was tested using the Spearman test instead of the Pearson's test. There were two dichotomous variables, 'language' and 'sex', 'sex' was coded with the two levels 'male' and 'female' and 'language' was coded with the two levels 'Dutch' and 'English'. The relation between these variables could be tested with the Odds Ratio, however it was decided to not calculate the Odds Ratio. There would be no association between these variables, because the dataset used in the current study included English and Dutch recordings of all males and all females. This means that there was no association of a certain sex with a certain language.

Then, the descriptive statistics were calculated. This included the mean, standard deviation, minimum, and maximum value of all parameters of male and female speakers and each condition (Dutch and English). An LME was used to test the relationship between two variables. In this study, those variables were language and voice quality. The fixed effects

were language spoken (English vs Dutch) and sex of the speaker (male or female). The samples of the speech recordings were not all independent from each other because there were multiple samples from the same speaker. So, random effects by speaker were added to the LME with the random intercept and the random slope. The random intercept characterizes variation due to within-speaker differences. A random slope by speaker was added for the effect of language. This was done for each VQ parameter. The code in R looked like this for the F0 parameter: F0 ~ Language + Sex + Interaction + (1|Speaker) + (1+Language|Speaker) + ε. This model was compared to models that take out the fixed-factors one-by-one until an intercept-only model was left to check what the optimal model was for each VQ parameter. The addition of random effects were assessed as well to check if the random slope significantly differed between speakers. In addition to this, a Bonferroni correction was performed that adjusted probability (p) values to account for the increased risk of false positives when conducting multiple statistical tests (Field, 2017). Applying the Bonferroni correction reduced the risk of false significant results and ensured that the results were reliable (Field, 2017).

# Results

## Descriptive statistics

In this section, the variables included in the analysis are reported and described. The descriptive statistics are reported in Table 1. Then the correlations between these variables are shown in Table 2 and descriptions are given for them.

### *F0, jitter and shimmer*

As can be seen in Table 1, the average F0 of females speaking Dutch is 185 Hz and when speaking English it is 180 Hz. The average F0 of males speaking Dutch is 104 Hz and when speaking English it is 101 Hz. It is important to note that 29% of the participants are male and 71% of the participants are female, so there is no equal split. The overall minimum F0 is 40 Hz and the overall maximum F0 is 324 Hz. This minimum is very low and indicates creaky voice. The average jitter of males speaking Dutch is 0.017% and when speaking English is 0.018%. This means that the period length of the vibration rate fluctuates on average with 0.017% for males when speaking Dutch and with 0.018% when speaking English. The average jitter of females differs with 0.05% from males for both English and Dutch. The average shimmer of males when speaking Dutch and when speaking English is 0.062%. This means that the amplitude of the vibrations fluctuate with an average of 0.062% for males.

### *Intensity, slope, tilt-linear and tilt-log*

The variables intensity, slope, tilt-linear, and tilt-log all represent glottal tension. There is little difference in the average values of these variables between males and females. The average intensity of males when speaking Dutch is 69 dB and when speaking English is 71 dB. The average intensity of females when speaking Dutch is 70 dB and when speaking English is 75 dB. The lowest intensity measured is 51 dB which represents the lowest glottal tension measured and it indicates an "*uh, um*" that could be said with a breathier voice. The slope, tilt-linear and tilt-log all represent the ratio of energy between the lower and upper harmonics. The average slope of females speaking Dutch is –18 Hz and when speaking English is –17 Hz. The minimum slope is –30 Hz and the maximum slope is –8 Hz. The minimum slope of –30 Hz represent more prominent upper harmonics than lower harmonics which means that the vocal folds move closer together so there is more glottal tension and the maximum slope of –8 Hz represent more prominent lower harmonics than upper harmonics which means that the vocal folds move away from each other so there is less glottal tension.

**Table 1**

*The descriptive statistics are displayed as the mean, standard deviation, minimum, and maximum of the dependent variables in both Dutch and English and divided in groups of male and female participants.*

| | **Mean** | | **SD** | | **Min** | | **Max** | |
|---|---|---|---|---|---|---|---|---|
| | Dutch | English | Dutch | English | Dutch | English | Dutch | English |
| **F0** | | | | | | | | |
| Male | 104.38 Hz | 101.09 Hz | 26.49 Hz | 29.79 Hz | 40.13 Hz | 44.17 Hz | 209.54 Hz | 205.78 Hz |
| Female | 184.80 Hz | 179.89 Hz | 47.43 Hz | 46.89 Hz | 42.13 Hz | 51.17 Hz | 324.05 Hz | 295.37 Hz |
| **Jitter** | | | | | | | | |
| Male | 0.017 % | 0.018 % | 0.013 % | 0.015 % | 0.002 % | 0.003 % | 0.085 % | 0.125 % |
| Female | 0.012 % | 0.013 % | 0.009 % | 0.012 % | 0.002 % | 0.002 % | 0.065 % | 0.110 % |
| **Shimmer** | | | | | | | | |
| Male | 0.062 % | 0.062 % | 0.031 % | 0.038 % | 0.013 % | 0.011 % | 0.20 % | 0.34 % |
| Female | 0.050 % | 0.058 % | 0.028 % | 0.027 % | 0.016 % | 0.011 % | 0.21 % | 0.24 % |
| **Intensity** | | | | | | | | |
| Male | 69.05 dB | 71.46 dB | 6.33 dB | 6.02 dB | 54.10 dB | 51.20 dB | 89.00 dB | 88.40 dB |
| Female | 70.28 dB | 74.70 dB | 5.72 dB | 5.35 dB | 55.60 dB | 57.30 dB | 87.40 dB | 84.60 dB |
| **Slope** | | | | | | | | |
| Male | –18.88 Hz | –17.05 Hz | 3.932 Hz | 3.230 Hz | –30.10 Hz | –27.70 Hz | –8.20 Hz | –9.50 Hz |
| Female | –18.00 Hz | –17.11 Hz | 3.816 Hz | 3.719 Hz | –29.60 Hz | –27.70 Hz | –7.80 Hz | –8.00 Hz |
| **Tilt-linear** | | | | | | | | |
| Male | –0.008 Hz | –0.008 Hz | 0.001 Hz | 0.001 Hz | –0.010 Hz | –0.010 Hz | –0.005 Hz | –0.004 Hz |
| Female | –0.007 Hz | –0.007 Hz | 0.002 Hz | 0.002 Hz | –0.010 Hz | –0.010 Hz | –0.003 Hz | –0.003 Hz |
| **Tilt-log** | | | | | | | | |
| Male | –38.24 Hz | –40.00 Hz | 7.506 Hz | 8.500 Hz | –57.90 Hz | –66.20 Hz | –21.90 Hz | –17.50 Hz |
| Female | –36.45 Hz | –37.71 Hz | 8.352 Hz | 9.008 Hz | –62.20 Hz | –61.80 Hz | –14.70 Hz | –17.00 Hz |

**Table 2**

*The correlation matrix shows the bivariate analyses of the Spearman test which represents the correlation between two continuous variables or between a continuous variable and a dichotomous variable.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1. Language | - | | | | | | | | |
| 2. Sex | - | - | | | | | | | |
| 3. F0 | –0.017 | –0.652*** | 1 | | | | | | |
| 4. Jitter | 0.003 | 0.261*** | –0.449*** | 1 | | | | | |
| 5. Shimmer | –0.027 | 0.230*** | –0.419*** | 0.653*** | 1 | | | | |
| 6. Intensity | 0.116** | –0.041 | 0.448*** | –0.451*** | –0.414*** | 1 | | | |
| 7. Slope | 0.147*** | –0.053 | 0.238*** | –0.232*** | –0.190*** | 0.478*** | 1 | | |
| 8. Tilt-linear | –0.050 | –0.289*** | 0.313*** | –0.150*** | –0.104* | 0.120*** | –0.025 | 1 | |
| 9. Tilt-log | –0.079 | –0.111** | 0.166*** | –0.082 | –0.051 | 0.086* | –0.047 | 0.845*** | 1 |

*p < .006. **p < .0006. ***p <.0001

*Note*. The cells with '-' are not measured with the Spearman test. The p-values are adjusted with the Bonferroni correction.

**Correlations between variables**

In Table 2 the correlation matrix is shown and the Spearman correlation shows a significant strong positive correlation between tilt-linear and tilt-log ($r = .85$; $p < .0001$). A positive correlation means that, as tilt-linear increases, the tilt-log also tends to increase and vice versa. Table 2 also shows a significant medium negative correlation between sex and F0 ($r = –0.65$; $p < .0001$) and between jitter and shimmer ($r = –0.65$; $p < .0001$). A negative correlation means that as jitter increases, shimmer tends to decrease and vice versa. Table 2 also shows a significant positive correlation between language and intensity ($r = .12$; $p <.0006$) and between language and slope ($r = .15$; $p < .0001$), although these correlations are so low they can be interpreted as hardly any to no correlation. The Spearman correlation in Table 2 shows that all non-significant correlations between variables are also weak correlations. Table 2 also shows that the relation between language and sex is not tested with the Spearman test. As explained in the methodology section, a certain sex would not associate with a certain language because there are Dutch and English samples of all speakers.

**Results of the linear mixed-effects models**

This section presents and describes the results of the linear mixed-effects models. The results are reported in Table 3 and descriptions are given in the following paragraphs.

***F0, jitter and shimmer***

Table 3 shows the results of all linear mixed-effects models. Model 1 takes the F0 as the dependent variable and shows the effects of language, sex, and their interaction effect on the stability of F0. A coefficient of –4.343 for language on F0 means that when a person speaks Dutch, their F0 lowers by about 4 Hz compared to when they speak English. However, this is a non-significant effect. This means that the difference in F0 is coincidental and that there is no effect of language on the stability of F0. Table 3 shows in model 2 and 3 that language also has a non-significant effect on jitter and shimmer. This means that the differences of jitter and shimmer between Dutch and English are not influenced by the language someone is speaking. Moreover, when comparing the models of F0, jitter and shimmer with the models without the random slope, the results showed a non-significant difference between these models. This shows that the random slope does not differ between speakers. So, the possible effect of language on F0, jitter and shimmer does not differ significantly between speakers.

***ST and MSE***

Table 3 shows that language also has a non-significant effect on intensity. This means that there is no effect of language on the stability of intensity. However, language does have a significant effect on slope and on tilt-log. Language affected slope ($\chi 2$ (1)=14.04, $p<0.001$), adding about 0.904 % ± 0.329 (standard errors). Language affected tit-log ($\chi 2$ (1)= 10.01, $p=0.002$), lowering by about 1.394 % ± 0.539 (standard errors). Model 7 shows that language does not have a significant effect on tilt-linear. So, two out of three measures of ST show an effect of language on ST. Moreover, when comparing the intensity model with the intensity model without the random slope, the results showed a significant difference between these models. This means that the random slope of the effect of language on intensity differs significantly between speakers. The difference between the models with and without the slope of the effect of language on slope shows no significant difference. This means that the effect of language on slope does not differentiate significantly between speakers. This is the same for the effect of language on tilt-linear and on tilt-log.

***The influence of sex on VQ parameters***

In addition to the effect of bilingualism, Table 3 shows that sex has a large effect on F0. Sex affected F0 ($\chi 2$ (1)=44.27, $p<0.001$), adding by about 79.66 Hz ± 9.8 (standard errors). This means that a male speaker has an F0 of about 80 Hz higher than a female speaker. Sex also has a significant effect on jitter and shimmer. Sex affected jitter ($\chi 2$ (1)=13.93, $p<0.001$), adding by about 0.136% ± 0.051 (standard errors). This means that a female speaker has 0.136% more jitter in their voice than a male speaker. Sex affected shimmer ($\chi 2$ (1)=12.10, $p<0.001$), adding about 0.106% ± 0.039 (standard errors). This means that a female speaker has 0.106% more shimmer in their voice than a male speaker. Sex also affected tilt-linear ($\chi 2$ (1)=6.20, $p=0.0128$), lowering by about 0.0011 ± 0.0004 (standard errors). However, sex does not have a significant effect on intensity, slope and tilt-log. Lastly, Table 3 shows that the interaction effect of language and sex is non-significant in all models. This means that sex does not influence the effect of language on the VQ parameters.

**Table 3**

*The model outputs of the effect of language and sex on voice quality parameters displayed with the b-coefficient, the standard error of the coefficient, and the lower and upper bounds of the confidence interval.*

| | | | 95% CI | |
|---|---|---|---|---|
| | b | SE b | 2.5% | 97.5% |
| **Model 1: F0** | | | | |
| Language | –4.343 | 3.513 | –11.279 | 2.473 |
| Sex | –79.655*** | 9.802 | –98.772 | –60.491 |
| Language*Sex | 2.998 | 6.287 | –9.295 | 15.336 |
| **Model 2: Jitter** | | | | |
| Language | –0.014 | 0.026 | –0.065 | 0.038 |
| Sex | 0.136*** | 0.051 | 0.037 | 0.235 |
| Language*Sex | 0.048 | 0.047 | –0.045 | 0.140 |
| **Model 3: Shimmer** | | | | |
| Language | –0.005 | 0.017 | –0.037 | 0.028 |
| Sex | 0.106** | 0.039 | 0.031 | 0.182 |
| Language*Sex | –0.007 | 0.029 | –0.064 | 0.049 |
| **Model 4: Intensity** | | | | |
| Language | 0.180 | 0.680 | –1.157 | 1.506 |
| Sex | –0.623 | 1.609 | –3.766 | 2.528 |
| Language*Sex | 1.858 | 1.251 | –0.596 | 4.304 |
| **Model 5: Slope** | | | | |
| Language | 0.904*** | 0.329 | 0.258 | 1.545 |
| Sex | –0.676 | 1.119 | –2.857 | 1.519 |
| Language*Sex | 0.909 | 0.600 | –0.259 | 2.088 |
| **Model 6: Tilt-linear** | | | | |
| Language | –0.0002 | 0.0001 | –0.0004 | 1.4666e-06 |
| Sex | –0.0011* | 0.0004 | –0.0020 | –0.0003 |
| Language*Sex | 0.0001 | 0.0002 | –0.0002 | 0.0005 |
| **Model 7: Tilt-log** | | | | |
| Language | –1.394* | 0.539 | –2.441 | –0.330 |
| Sex | –2.234 | 2.199 | –6.545 | 2.070 |
| Language*Sex | –0.265 | 0.942 | –2.117 | 1.579 |

*p < .017. **p < .0017. ***p < .0003

*Note*. The p-values are adjusted with the Bonferroni correction.

## Discussion

The research question of the current research is: How does L1 Dutch and L2 English bilingualism influence the stability of voice quality parameters in cross-language voice comparison?

It was expected to find that the stability of jitter, shimmer, ST and MSE are influenced by Dutch-English bilingualism and that the stability of F0 is not influenced by Dutch-English bilingualism. The results correspond partially with the expectations. The answer to the research question is that L1 Dutch and L2 English bilingualism influences the stability of spectral tilt as a VQ parameter in cross-language voice comparisons between Dutch-English bilingual speakers and does not influence the stability of F0, jitter, shimmer and MSE as VQ parameters in cross-language voice comparisons between Dutch-English bilingual speakers.

In the following paragraphs, the findings of the current study are explained and related to previous research and theories. Then, the limitations and strengths of the current study are described. Finally, recommendations for future research are presented and practical implications are discussed.

### The effect of language on mean spectral energy

In contrast to the expectations, mean spectral energy appears to be a stable VQ parameter across Dutch and English voice comparisons of the schwa-like "*uh*" and "*um*" filled pause vowels. This contradicts previous research by Bahmanbiglu and colleagues (2017), Ng and colleagues (2012) and Zhu and colleagues (2022), who found that MSE is an unstable parameter for cross-language voice comparisons. In these studies, the instability of MSE is explained by different resonance patterns and laryngealization between languages. Resonator cavities affect voice quality (Hynes, 1993) and the use of posterior resonance is argued to be caused by glottal tension (Kerr, 2000). More posterior resonance use was found when people speak tonal languages than when people speak English (Kerr, 2000). The observed stability of MSE across Dutch and English in the current study, may be explained by the phonetic and phonatory similarities between the two languages. Unlike tonal languages, such as Cantonese, Dutch and English might share comparable resonance patterns as there is no indication that Dutch and English use posterior resonance patterns and both languages show similar anterior and mid-vocal resonance patterns in the formant frequencies of vowels (Collins & Mees, 2003).

Moreover, Dutch speakers may transfer their L1 vocal behaviour and use of glottal tension into English, leading to consistent vocal energy output between both languages. As a

result, MSE stays stable despite the language switch, unlike in language pairs with greater phonatory differences (see Ng et al., 2012; Zhu et al., 2022). When learning a second language it is difficult to acquire sounds that are not know in someone's first language (Lo, 2021). According to the Speech Learning Model (Flege, 1995; Flege & Bohn, 2021) and the Perceptual Assimilation of Second Language Speech Learning (Best & Tyler, 2007), new sounds from the L2 are categorized with reference to L1 categories and, depending on the ability of the speaker, a new sound category is created or the new sound is submitted by a known sound from the L1. This was also described in the case of the Tamil Nadu Police that was explained in the introduction. There was L1 transfer of the regional Madurai language variety to the L2 (Tamil) when pronouncing the word "lanthu". Transfer of the vocal behaviour from the first language to the second language could also be the case in the use of glottal tension of Dutch-English bilinguals. In Dutch, it is uncommon to use laryngealization while it is common in English (Collins & Mees, 2003), therefore, it can be difficult to acquire the ability or the knowledge on when to use it correctly. This could lead to the transferring of the L1 use of glottal tension into the L2.

**The effect of language on fundamental frequency**

It was expected to find that F0 is not influenced by Dutch-English bilingualism. The results of the current study correspond with this expectation and with the finding by Theelen (2017) that the F0 of native Dutch speakers stays stable when they speak English. According to Cantor-Cutiva and colleagues (2021), an explanation for the stability of F0 across languages can be found in the speech accommodation theory (Giles et al., 1987). Speakers adjust their F0 to match the F0 of native speakers to build rapport with the interlocutor and to reduce the social distance. Cantor-Cutiva and colleagues (2021) suggest that the L2 adjustments in vocal behaviour are then incorporated in the speaker's muscle memory which influences the L1 vocal behaviour. Thus, native Dutch speakers might decrease their F0 on order to match the perceived pitch of native English speakers and once this decreased F0 is incorporated in the speaker's muscle memory, it is also produced when speaking Dutch. However, this long-term incorporation of the L2 vocal range in the speaker's muscle memory is caused by long-term adjustment of the vocal behaviour (Laver, 1987). The current study only included speakers in their first semester of UCU to prevent influences of an international setting on their pronunciation and accent. This makes it unlikely that the accommodation of their speech to sound more native had already affected their L1. Another explanation, is that Dutch-English bilingual speakers do not accommodate their speech to sound more native, but

that there is a transfer of vocal behaviour from the L1 to the L2. This suggests that Dutch-English bilinguals maintain similar F0 values in both languages, rather than actively adjusting their F0 when switching. This could cause the F0 to stay stable across languages.

**The effect of language on jitter and shimmer**

The results for jitter and shimmer contradict the expectation that jitter and shimmer are influenced by Dutch-English bilingualism, as no effect of language on jitter and shimmer was found. An explanation for these findings can be that, aside from L1 (Dutch) transfer of the F0 into the L2 (English), the variation of the F0 may also be transferred across languages. This would mean that bilingual speakers show similar jitter and shimmer values in both Dutch and English. This contradicts the findings of Cantor-Cutiva and colleagues (2021), who suggest that bilingual speakers try to match the perceived pitch variation in their L2 to match native speakers. When explaining the stability of jitter and shimmer in the current study, the same reasoning can be applied as with the explanation of F0 stability across Dutch and English. The absence of a language effect on jitter and shimmer in the current study, may be attributed to differences in the amount of L2 exposure and the stage of bilingual development. While previous studies like Cantor-Cutiva and colleagues (2021) and Järvinen and colleagues (2013) included speakers who were likely to have undergone long-term phonetic and phonatory adaption through frequent interaction in native L2 environments, the participants in the present study were in the early phase of their university education and had limited experience in English-speaking contexts. As a result, they may not yet have adjusted their phonatory control, which is reflected in jitter and shimmer and F0, to align with native-like L2 patterns. Instead, they likely maintained the precise phonatory behaviour of their L1 (Dutch), leading to comparable jitter and shimmer values in both languages. Moreover, intonation patterns in L2, which can be caused by variation in F0, can be difficult to acquire (Almbark et al., 2014). This could also cause Dutch-English bilingual speakers to maintain similar F0, jitter and shimmer when speaking English.

**The effect of language on spectral tilt**

The current study shows contradicting results of the stability of spectral tilt. ST was measured in three various methods and two out of three methods showed that ST is not a stable VQ parameter across Dutch and English voice comparisons. The different results across the three ST measures in the current study can be attributed to methodological differences in how ST is measured. Both the slope of the LTAS and the log-transformed tilt measure (tilt-log) are more sensitive to the distribution of energy in higher frequency ranges than the linear

tilt measure, which makes them more susceptible to subtle articulatory and phonatory variations between languages (Hillebrand et al., 1994; Stevens, 1998). In contrast, the linear tilt measure (tilt-linear) looks at energy differences more evenly across all frequencies, which might make it less sensitive to subtle differences between languages (Stevens, 1998; Zhu et al., 2023). As a result, there is a significant effect of language on slope and tilt-log, while there is no significant effect of language on tilt-linear.

Previous literature on ST as a VQ parameter across languages also gives contradicting results. Ng and colleagues (2012) found ST to be an unstable parameter across Cantonese and English, while Zhu and colleagues (2022) found ST to be a stable parameter across Mandarin and English. The method to measure ST used in these previous studies was to measure the slope of the LTAS. In the current study, there was an effect of language found on this measure of ST. This confirms the findings of Ng and colleagues (2012), but contradicts the findings of Zhu and colleagues (2022). The current study, also found significant differences of the effect of language on slope between speakers. This means that there is between-speaker variation of the effect of language on their slope. However, these differences between speakers were not found for the other two measures of ST.

The variation in the results of ST across languages between various studies can be explained by language-specific influences on vocal qualities as described by Bruyninkcx and colleagues (1994). In their study in Spanish-Catalan bilingual speakers, they showed that the same vocal apparatus can produce different vocal qualities depending on the language being spoken. They argue that different languages are associated with distinct phonatory settings and vocal behaviour, which may include variation in parameters such as ST. In addition to this, Zhu and colleagues (2022) and Ng and colleagues (2012) suggest that differences in resonance patterns shape glottal tension, which is reflected in ST values. The stability of ST across Mandarin and English is likely due to similar resonance patterns between these languages (Zhu et al., 2022). In contrast, Cantonese was found to have different resonance patterns than English (Kerr, 2000) and this might lead to the instability of ST between Cantonese and English (Ng et al., 2012).

Applying this reasoning to the current study, the instability of ST across Dutch and English may similarly result from subtle differences in resonance patterns between the two languages. These difference could influence glottal tension and, consequently, ST. Moreover, individual speakers may implement language-specific phonatory settings in varying was, even when switching between closely related languages. This could explain the observed between-speaker variation in the effect of language on slope in the current study. Although both MSE

and ST are associated with glottal tension, their sensitivity to cross-linguistic variation may differ. ST is more directly influenced by subtle phonatory and resonance shifts (Hanson, 1997; Ng et al., 2012), whereas MSE reflects overall vocal energy and is therefore less sensitive to resonance-related differences between languages or speaker-specific variation (Gobl & Ní Chasaide, 2003; Kreiman et al., 2021). This might explain the stability of MSE and the instability of ST across Dutch and English.

**The influence of sex on VQ parameters**

Lastly, sex was found to be significantly influencing F0, jitter, shimmer and tilt-linear. This confirms findings of earlier research that F0 is influenced by sex due to physiological differences (Kreiman & Sidtis, 2011; Ng et al., 2012) and research of Cantor-Cutiva and colleagues (2021) about jitter and shimmer in bilingual versus monolingual speakers. They found that jitter and shimmer differ between males and females. Male bilingual speakers had higher jitter than male monolingual speakers and female bilingual speakers had lower jitter and shimmer than female monolingual speakers. The influence of sex on tilt-linear is just below the alpha adjusted with Bonferroni correction. There is no significant effect of sex on the other ST measures. This shows that it is more credible that sex is not significantly affecting the ST. This contradicts the findings of Mendoza and colleagues (1996) and Kreiman and Sidtis (2011) who both found lower ST measures for females than for males and argue that females are more likely to speak with a more breathy voice quality than males. Moreover, the interaction effect between sex and language appeared to be non-significant on all VQ parameters. This means that the significant effect of sex on F0, jitter, shimmer, and tilt-linear does not influence the effect of language on these parameters.

Altogether, these findings lead to the conclusion that L1 Dutch and L2 English bilingualism affects the stability of spectral tilt as a VQ parameter in cross-language voice comparisons, while it does not affect the stability of F0, jitter, shimmer, and MSE as VQ parameters in such comparisons between Dutch-English bilingual speakers.

**Limitations and strengths**

In addition to the findings, it is important to acknowledge the limitations of the current study. The voice quality parameters were analysed using tokens of the schwa-like /ə/ vowel taken from the filled pauses "*uh*" and "*um*", with durations of at least 160 milliseconds in both Dutch and English. Although this method provides comparability, it limits the generalizability of the results. Including a broader range of naturally long-sustained vowels, such as the /i/ vowel common in both Dutch and English (Ladefoged & Johnson, 2015), may

lead to different outcomes. Different vowels involve different articulatory settings and places of articulation, which influence the shape and movement of the vocal tract, thereby affecting VQ parameters (Kent & Read, 2002).

Moreover, lexical vowels, which are vowels occurring within words, differ from filled pause vowels in their phonetic context. Lexical vowels are typically preceded and/or followed by other vowels or consonants, making them more subjective to coarticulation effects compared to filled pause vowels, which are typically preceded and/or followed by silence (Swerts, 1998). Coarticulation can affect the acoustic properties and voice quality (Gick et al., 2013), meaning that lexical vowels may show more variation in VQ parameters than filled pause vowels. Therefore, the findings on VQ stability in this study, which are based on schwa-like filled pause vowels, might not be directly applicable to other vowels or lexical schwa-like vowels. For this reason, caution is needed when generalizing these results to other contexts or studies that use different vowel material in cross-language VQ comparisons. Future research should consider including a broader range of vowels to better understand the effects of bilingualism on voice quality across languages.

Another methodological limitation, is the unbalanced sample group regarding sex. Out of 35 participants, 25 are female and 10 are male. This could have influenced the results regarding the effect of sex on the parameters. The current findings show that sex significantly affects F0, jitter, shimmer and one of the ST measures. With a more balanced sample group regarding sex, the ambiguity of the effect of sex on ST might become clearer. It could also render different results regarding the interaction effect of sex and language on the VQ parameters. The current results show that there is no interaction effect on any VQ parameter, even though there is a significant effect of sex on various VQ parameters. It is a possibility that a more balanced group of males and females could render different results regarding the interaction effect. It is also a possibility that it confirms the current findings of no interaction effect on any VQ parameter.

Apart from the differences in sex, the sample was a homogeneous group regarding age and language background, with Dutch being their only L1 and spoken without an audible accent. Although there was no detailed information about their L2 English proficiency, it is of at least B2 level and because the participants were recorded within one month after arrival on campus, it was expected that the multilingual environment of the campus had not influenced their English proficiency yet. With the homogeneity of the sample group, the aim was to exclude speaker characteristics that could influence the findings. This strengthens the validity and reliability of the research.

Altogether, the current study contributes to the literature on cross-language voice comparison with the language combination of Dutch and English. Previous literature on cross-language voice comparison between Dutch and English is limited and the current study adds to earlier research that have mostly focused on English versus tonal languages and English versus Spanish or French. However, additional research incorporating different languages should be conducted on order to further understand the influence on VQ parameters. It will help to better comprehend how language affects VQ if further research is conducted on the VQ characteristics of bilinguals who speak other languages.

## Recommendations

### *Future research*

Following the current study, the recommendations for future research on the effect of language on VQ parameters in cross-language voice comparison is to look at the effect of different methods of measuring ST and to incorporate a broader vowel range. The current study measured ST using three different methods and showed different results. Two out of three ST measurements showed instability across Dutch and English voice comparisons. This contradicts the finding of stability of MSE across Dutch and English even though they both measure glottal tension. More research about the differences and similarities of the methods to measure ST and the effects on cross-language voice comparisons will help to better comprehend the influence of vocal features on ST and how this is affected by language and bilingualism. Including a broader vowel range will help to understand whether the stability of VQ parameters, such as ST observed in filled pauses, generalizes across different speech sounds.

Additionally, the VQ parameters studied in the current research are not the only VQ parameters that can be affected by language. For example, first spectral peak (FSP) and harmonics-to-noise ratio (HNR) are parameters investigated in previous literature. These parameters were not incorporated in the current study in light of time constraints and as they are not as contested as other parameters. FSP was found to be a stable parameter across English and Mandarin and across English and Cantonese. HNR showed to be a stable parameter across different recording qualities (Hughes et al., 2019). However, when involving non-contemporaneous speech and speech style mismatch, HNR was found to be an unstable parameter (Chan, 2023). Literature on HNR and FSP compared across languages is limited, so it is difficult to make strong conclusions about the stability of these parameters in cross-language voice comparisons. Further research on the stability of FSP and HNR with various

language combinations can contribute to the knowledge about the robustness of these parameters in cross-language voice comparisons.

Finally, this research also contributes to a more multilingual perspective in research on forensic voice comparisons. As mentioned before, it is recommended to investigate other languages in order to better understand the effect of language on VQ parameters and the validity of VQ parameters in certain cross-language voice comparisons.

***Practical implications***

The findings of the current research can be brought into practice by adding to the knowledge base about the stability of certain VQ parameters in voice comparisons across Dutch and English bilingual speakers. As bilingualism and cross-language voice comparisons are becoming more prevalent in forensic cases and there is limited knowledge on the discriminatory potential of VQ parameters in cross-language voice comparisons (Lo, 2021), forensic experts can draw on this knowledge base to validate the robustness of the VQ parameters they use to distinguish between speakers. However, it is advised to be cautious when using these VQ parameters as speaker discriminants in cross-language voice comparisons, because various components could influence the robustness of the parameters. For example, when comparing across languages other than Dutch and English or when the recording quality is not optimal. Notably, the majority of the forensic experts in the research by Gold and French (2011) stated that, while some individual parameters hold significant weight, it is the total combination of parameters that they consider essential in distinguishing speakers. As Gold and French (2011) put this in Aristotelian terms, “The whole is greater than the sum of the parts” (Aristotle).

**References**

Almbark, R., Bouchhioua, N., & Hellmuth, S. (2014). Acquiring the phonetics and phonology of English word stress: Comparing learners from different L1 backgrounds. *Concordia Working Papers in Applied Linguistics, 5*, 19–35. https://doe.concordia.ca/copal/documents/3_Almbark_etal_Vol5.pdf

Altenberg, E. P., & Ferrand, C. T. (2006). Fundamental Frequency in Monolingual English, Bilingual English/Russian, and Bilingual English/Cantonese Young Adult Women. *Journal of Voice*, *20*(1), 89–96. https://doi.org/10.1016/j.jvoice.2005.01.005

Bahmanbiglu, S. A., Mojiri, F., & Abnavi, F. (2017). The Impact of Language on Voice: An LTAS Study. *Journal of Voice*, *31*(2), 249.e9-249.e12. https://doi.org/10.1016/j.jvoice.2016.07.020

Bates, D.M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.

Benoist-Lucy, A., & Pillot-Loiseau, C. (2013). The Influence of language and speech task upon creaky voice use among six young American women learning French. In *Proceedings of Interspeech 2013* (pp. 2395-2399). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2013-558

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Boersma, P., & Weenink, D. (2024). Praat: doing phonetics by computer (version 6.4.25) [Computer program]. Retrieved May 15, 2025, from https://www.fon.hum.uva.nl/praat/

Bruyninckx, M., Harmegnies, B., Llisterri, J., & Poch-Oiivé, D. (1994). Language-induced voice quality variability in bilinguals. *Journal of Phonetics*, *22*(1), 19–31. https://doi.org/10.1016/S0095-4470(19)30265-7

Cantor-Cutiva, L. C., Bottalico, P., Nudelman, C., Webster, J., & Hunter, E. J. (2019). Do Voice Acoustic Parameters Differ Between Bilingual English-Spanish Speakers and Monolingual English Speakers During English Productions? *Journal of Voice*, *35*(2), 194–202. https://doi.org/10.1016/j.jvoice.2019.08.009

Cantor-Cutiva, L. C., Bottalico, P., Webster, J., Nudelman, C., & Hunter, E. J. (2023). The Effect of Bilingualism on Production and Perception of Vocal Fry. *Journal of Voice*, *37*(6), 970.e1-970.e10. https://doi.org/10.1016/j.jvoice.2021.06.002

Cantor-Cutiva, L. C., Jiménez-Chala, E.-A., Bottalico, P., & Hunter, E. J. (2021). Bilingualism and Voice Production. Differences Between Bilingual Latin-American Spanish- English Female Speakers and Monolingual Spanish Female Speakers During Spanish Productions. *Journal of Voice*, *37*(5), 716–721. https://doi.org/10.1016/j.jvoice.2021.04.026

Casper, J. K., Brewer, D. W., & Colton, R. H. (1987). Variations in normal human laryngeal anatomy and physiology as viewed fiberscopically. *Journal of Voice, 1*(2), 180-185. https://doi.org/10.1016/50892-1997(87)80043-7

Chan, R. K. W. (2023). Evidential value of voice quality acoustics in forensic voice comparison. *Forensic Science International*, *348*, 111725–111725. https://doi.org/10.1016/j.forsciint.2023.111725

Collins, B. D., & Mees, I. (2003). *The Phonetics of English and Dutch* (5th rev. ed.). BRILL.

Council of Europe (2019). "Common European framework of reference for languages: Learning, teaching, assessment (CEFR),". Retrieved March 13, 2025 from https://www.coe.int/en/web/common-european-frameworkreference-languages

Davidson, L., & Erker, D. (2014). Hiatus resolution in American English: the case against glide insertion. *Language*, *90*(2), 482–514. https://doi.org/10.1353/lan.2014.0028

de Boer, M. M., & Heeren, W. F. L. (2023). The language dependency of /m/ in native Dutch and non-native English. *The Journal of the Acoustical Society of America*, *154*(4), 2168–2176. https://doi.org/10.1121/10.0021288

de Boer, M. M., & Heeren, W. F. L. (2020). Cross-linguistic filled pause realization: The acoustics of uh and um in native Dutch and non-native English. *The Journal of the Acoustical Society of America*, *148*(6), 3612–3622. https://doi.org/10.1121/10.0002871

Edwards, J. (2012). Bilingualism and multilingualism: Some central concepts. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (2nd ed., pp. 5–25). John Wiley & Sons, Ltd. https://doi.org/10.1002/ 9781118332382.ch1

Esling, J. H., Moisling S. R,. & Benner, A., & Crevier-Buchman, L. (2019). *Voice Quality: The Laryngeal Articulator Model*. Cambridge University Press. http://refhub.elsevier.com/S0892-1997(24)00162-0/sbref22

European Commissions (2024). *Europeans and their languages*. (Report). https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2686

Farrús, M., Hernando, J., & Ejarque, P. (2007). Jitter and shimmer measurements for speaker recognition. In *Proceedings of Interspeech 2007* (pp. 7787-781). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2007-147

Field, A. (2017). *Discovering statistics using IBM SPSS Statistics*. Sage Publications std.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross language research* (pp. 233–277). York Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Way land (Ed.), *Second language speech learning* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Gerratt, B. R., & Kreiman, J. (2001). Measuring vocal quality with spectral tilt. *The Journal of the Acoustical Society of America, 110*(4), 2586. https://doi.org/10.1121/1.4744819

Gick, B., Wilson, I., Derrick, D., & Cook, C. (2013). *Articulatory Phonetics* (2nd ed.). Wiley-Blackwell.

Gobl, C., & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication, 40*(1–2), 189–212. https://doi.org/10.1016/S0167-6393(02)00082-1

Gold, E. & French, P. (2011). International practices in forensic speaker comparison. *The international journal of speech, language and the law, 18*(2), 293-307. https://doi.org/10.1558/ijsll.v18i2.293

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America, 101*(1), 466–481. https://doi.org/10.1121/1.417991

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*(4), 769–778. https://doi.org/10.1044/jshr.3704.769

Hughes, V., Cardoso, A., Harrison, P., Foulkes, P., French, P., & Gully, A. J. (2019). Forensic voice comparison using long-term acoustic measures of laryngeal voice quality. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 1455-1459). Australasian Speech Science & Technology Association Inc. https://doi.org/10.21437/ICPhS.2019-1504

Hughes, V., Foulkes, P., and Wood, S. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. The *International Journal of Speech, Language and the Law 23*, 99–132. https://doi.org/10.1558/ijsll.v23i1.29874

Hynes, W. (1993). The results of pharyngoplasty by muscle transplantation in “failed cleft palate” cases, with special reference to the influence of the pharynx on voice

production. *British Journal of Plastic Surgery*, *46*(5), 430–439. https://doi.org/10.1016/0007-1226(93)90051-C

Järvinen, K., Laukkanen, A. M., & Aaltonen, O. (2012). Speaking a foreign language and its effect on F0. *Logopedics Phoniatrics Vocology*, *38*(2), 47–51. https://doi.org/10.3109/14015439.2012.687764

Jessen, M. (1997). Speaker-specific information in voice quality parameters. *International Journal of Speech, Language and the Law, 4*(1), 84–103. https://doi.org/10.1558/ijsll.v4i1.84

Keating, P. A., Garellek, M., Kreiman, J., & Chai, Y. (2023). Acoustic properties of subtypes of creaky voice. *The Journal of the Acoustical Society of America*, *153*(3), A297-A297. https://doi.org/10.1121/10.0018918

Kent, R. D., & Read, C. (2002). *The Acoustic Analysis of Speech* (2nd ed.). Singular Publishing Group.

Kerr, J. (2000). Articulatory setting and voice production: Issues in accent modification. *Prospect*, *15*(2), 4-14. https://search.informit.org/doi/10.3316/aeipt.102461

Kreiman, J., Gerratt, B. R., & Ito, M. (2021). The relative contribution of source and filter characteristics to voice quality. *The Journal of the Acoustical Society of America, 149*(5), 3438–3450. https://doi.org/10.1121/10.0004757

Kreiman, J. & Sidtis, D. (2011). *Foundations of voice studies: an interdisciplinary approach to voice production and perception*. Wiley-Blackwell.

Ladefoged, P., & Johnson, K. (2015). *A Course in Phonetics* (7th ed.). Cengage Learning.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge University Press.

Laver, J. (1987). *Individual features in voice quality* [PhD thesis, University of Edinburgh]. http://hdl.handle.net/1842/6732

Lo, J. H. (2021). *Issues of bilingualism likelihood ratio-based forensic voice comparison* [PhD thesis, University of York]. https://etheses.whiterose.ac.uk/30007/1/Lo_JJH_Thesis_Final.pdf

Mendoza, E., Valencia, N., Muñoz, J., & Trujillo, H. (1996). Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *Journal of Voice*, *10*(1), 59–66. https://doi.org/10.1016/S0892-1997(96)80019-1

Monsen, R. B., Engebretson, A. M., & Vemula, N. R. (1978). Indirect assessment of the contribution of subglottal air pressure and vocal-fold tension to changes of

fundamental frequency in English. *The Journal of the Acoustical Society of America*, *64*(1), 65–80. https://doi.org/10.1121/1.381957

Nederlands Forensisch Instituut. (2016). *Vergelijkend Spraakonderzoek*. https://www.forensischinstituut.nl/publicaties/publicaties/2020/02/03/vakbijlage-vergelijkend-spraakonderzoek

Nederlands Forensisch Instituut. (2017). *Waarschijnlijkheidstermen*. https://www.forensischinstituut.nl/publicaties/publicaties/2017/10/18/vakbijlage-waarschijnlijkheidstermen

Ng, M. L., Chen, Y., & Chan, E. Y. K. (2012). Differences in Vocal Characteristics Between Cantonese and English Produced by Proficient Cantonese-English Bilingual Speakers—A Long-Term Average Spectral Analysis. *Journal of Voice*, *26*(4), e171–e176. https://doi.org/10.1016/j.jvoice.2011.07.013

Ng, M. L., Chen, Y., & Kreiman, J. (2012). Voice quality evaluation across languages: A cross-linguistic comparison. *Folia Phoniatrica et Logopaedica, 64*(1), 1–9. https://doi.org/10.1159/000329604

Nguyen, D. D., & Kenny, D. T. (2009). Effects of Muscle Tension Dysphonia on Tone Phonation: Acoustic and Perceptual Studies in Vietnamese Female Teachers. *Journal of Voice*, *23*(4), 446–459. https://doi.org/10.1016/j.jvoice.2007.12.004

Nolan, F. (1987). The phonetic bases of speaker recognition. *Speech Communication*, *6*(2), 171–175. https://doi.org/10.1016/0167-6393(87)90039-2

Orr, R. & Quené, H. (2017). D-LUCEA: Curation of the UCU Accent Project Data. In: Odijk, J., &, van Hessen, A. (eds.) *CLARIN in the Low Countries*, 181-193. Ubiquity Press. https://doi.org/10.5334/bb1.15

Quené, H., Orr, R., & van Leeuwen, D. (2017). Phonetic similarity of /s/ in native and second language: Individual differences in learning curves. *The Journal of the Acoustical Society of America*, *142*(6), 519–524. https://doi.org/10.1121/1.5013149

R Core Team (2024). “R: A language and environment for statistical computing” https://www.R-project.org/ (Last viewed 1 June 2025).

Rietveld, T., & van Heuven, V. J. J. P. (2009). *Algemene fonetiek* (3e, herz. dr.). Coutinho

San Segundo, E., & Gómez-Vilda, P. (2014). Evaluating the forensic importance of glottal source features through the voice analysis of twins and non-twin siblings. *Language and Law / Linguagem e Direito, 1*(2), 22–41. https://ojs.letras.up.pt/index.php/LLLD/article/view/2430

San Segundo, E., & Mompean, J. A. (2017). A Simplified Vocal Profile Analysis Protocol for the Assessment of Voice Quality and Speaker Similarity. *Journal of Voice*, *31*(5), 644.e11-644.e27. https://doi.org/10.1016/j.jvoice.2017.01.005

Sicoli, M. A. (2015). Voice registers. In: D. Tannen, H.E. Hamilton, & D. Schiffrin (eds.), *The Handbook of Discourse Analysis* (pp. 105-126). John Wiley & Sons.

Stevens, K. N. (1998). *Acoustic phonetics*. MIT Press.

Stouten, F., & Martens, J. P. (2003). A feature-based filled pause detection system for Dutch. In *Proceedings of the 2003 IEEE Workshop on Automatic Speech Recognition and Understanding* (pp. 309-314). IEEE. https://doi.org/10.1109/ASRU.2003.1318459

Sundaram, R., & Kannan, S. (2023, September 13). Cops add voice analysis as forensic tool to crack crime. *The Times of India*. Retrieved May 22, 2025, from https://timesofindia.indiatimes.com/city/chennai/tn-cops-add-voice-analysis-as-forensic-tool-to-crack-crime/articleshow/103623936.cms

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*(4), 485–496. https://doi.org/10.1016/S0378-2166(98)00014-9

Theelen, M. (2017). Fundamental frequency differences including language effects. *Junctions: Graduate Journal of the Humanities, 2*(1), 9-24. http://junctionsjournal.org/

Titze, I. R. (2023). Simulation of vocal loudness regulation with lung pressure, vocal fold adduction, and source-airway interaction. *Journal of Voice*, *37*(2), 152–161. https://doi.org/10.1016/j.jvoice.2020.11.030

van Hugte, T. B. R., & Heeren, W. F. L. (2024). Exploring Interspeaker Variation in Creaky Voice in Dutch. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2024.05.011

Wester, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences, 29*(2-3), 477–491. https://doi.org/10.1016/j.langsci.2006.12.029

Zhu, S., Chong, S., Chen, Y., Wang, T., & Ng, M. L. (2022). Effect of Language on Voice Quality: An Acoustic Study of Bilingual Speakers of Mandarin Chinese and English. *Folia Phoniatrica et Logopaedica*, *74*(6), 421–430. https://doi.org/10.1159/000525649

Zhu, X., Lee, A., & Wang, S. (2023). Cross-language spectral tilt analysis: Stability and variability in voice quality measures. *Journal of Phonetics, 93*, 101160. https://doi.org/10.1016/j.wocn.2023.101160

## Appendix

Praat Script

```
#########################################################################
#       W F L Heeren, December 2016
#       procedures written by Jos Pacilly, Leiden University
#               adapted April 2025 for D. de Graaff
#
#
#########################################################################

# input directories, audio file and text grid should have the same names
audioDir$   = "C:\Users\name\data\ recordings wav"
txtDir$  = "C:\Users\name\data\textgrids"

# create output file and header
extractData$ = "ExtractData.analysis.txt"
fileappend 'extractData$' fileName 'tab$' duration 'tab$' language 'tab$' intensity-(dB) 'tab$'
slope-(dB) 'tab$' tilt-linear 'tab$' tilt-log 'tab$' jitter 'tab$' shimmer 'tab$' f0 'newline$'

files = Create Strings as file list: "fileList", txtDir$ + "\*.TextGrid"
lengthList = Get number of strings

for iRow from 1 to lengthList
        selectObject: files
        name$ = Get string: iRow
        Read from file: txtDir$ + "\" + name$
        fileName$ = selected$("TextGrid")

        wavFileName$ = fileName$
        Read from file: audioDir$ + "\" + fileName$ + ".wav"
        idSnd = selected("Sound")
        idPitch = To Pitch (raw cc): 0, 30, 600, 15, "no", 0.03, 0.45, 0.01, 0.35, 0.14
        idPitch = selected("Pitch")

        select idPitch
        plus idSnd
        idPp = To PointProcess (cc)

        selectObject: "TextGrid 'fileName$'"

        # extract number and duration of filled pause intervals that are annotated in tier 4
        numberOfIntervals = Get number of intervals: 4
        for interval from 1 to numberOfIntervals

                selectObject: "TextGrid 'fileName$'"
                label$ = Get label of interval: 4, interval

                if label$ == "uh"
                        startTarget = Get start point: 4, interval
```

```
                    endTarget = Get end point: 4, interval
                    duration = endTarget - startTarget
                    fileappend 'extractData$' 'fileName$' 'duration:6' 'tab$'

                    # which language?
                    languageInterval = Get interval at time: 1, startTarget
                    languageLabel$ = Get label of interval: 1, languageInterval
                    fileappend 'extractData$' 'languageLabel$' 'tab$'

                    # do spectral measurements across the segment's duration
                    selectObject: "Sound 'fileName$'"

                    idInt2 = noprogress To Intensity... 100 0 yes
                    idInt2 = selected("Intensity")
                    idInt2$ = selected$("Intensity")
                    call appendIntensity idInt2 startTarget endTarget
                    call appendSpectrum idSnd startTarget endTarget
                    call voiceAnalysis idSnd startTarget endTarget idPp idPitch
                    select idPitch
                    pitch = Get mean: startTarget, endTarget, "Hertz"
                    fileappend 'extractData$' 'pitch:2' 'tab$'

                    selectObject: "Intensity 'idInt2$'"
                    Remove

                    fileappend 'extractData$' 'newline$'

                endif

        endfor

        selectObject: "Sound 'fileName$'"
        plusObject: "TextGrid 'fileName$'"
        plusObject: "Pitch 'fileName$'"
        Remove

        select idPp
        plus idPitch
        plus idInt2
        Remove
endfor

# procedures written by Jos Pacilly, Leiden University
procedure appendIntensity .id .t1 .t2
  select .id
  .iMean = Get mean... .t1 .t2 dB
  fileappend 'extractData$' '.iMean:1' 'tab$'
endproc

procedure appendSpectrum .id .t1 .t2
```

```
  select .id
  .idSndFil  = Filter (pass Hann band)... 0 8000 100
  .idSndPart = Extract part... .t1 .t2 rectangular 1 yes
  .idSpec    = To Spectrum... no
  select .idSndPart
  .idLtas  = To Ltas... 100
  .slope   = Get slope... 0 1000 1000 5000 energy
  .report$ = Report spectral trend... 100 5000 linear Robust
  .a1 = extractNumber(.report$, "Slope: ")
  select .idLtas
  .report$ = Report spectral trend... 100 5000 logarithmic Robust
  .a2 = extractNumber(.report$, "Slope: ")

  select .idSndFil
  plus .idSndPart
  plus .idSpec
  plus .idLtas
  Remove
  fileappend 'extractData$' '.slope:1' 'tab$' '.a1:1'  'tab$' '.a2:1' 'tab$'
endproc

# procedure from Praat Help
procedure voiceAnalysis .id .t1 .t2 .pp .pi
  select .id
  plus .pp
  plus .pi
  .voiceReport$ = Voice report: .t1, .t2, 30, 500, 1.3, 1.6, 0.03, 0.45
  .jitter = extractNumber (.voiceReport$, “Jitter (local): ”)
  .shimmer = extractNumber (.voiceReport$, “Shimmer (local): ”)

  fileappend 'extractData$' '.jitter:5' 'tab$' '.shimmer:5' 'tab$'

endproc
```