A Forensic Speaker Identification Study: An Auditory-Acoustic Analysis of Phonetic Features and an Exploration of the 'Telephone Effect'

Universitat Autònoma de Barcelona

Treball de Fi de Grau

Grau d'Estudis Anglesos

15th June 2017

Name: Wolf De Witte

Supervisor: Dr. Núria Gavaldà Ferré

Acknowledgements

I would like to thank my supervisor, Dr. Núria Gavaldà, for her support, help, encouragement, and constructive feedback throughout the whole process, aiding me with her expertise and knowledge in the area of forensic phonetics. She initially sparked my interest in phonetics two years ago, which has grown during the remainder of my degree, consequently resulting in the discovery of the field of forensic phonetics, my desire to pursue research in this area, and the creation of this thesis.

I would also like to thank my partner and best friend, Jessica McDaid, for supporting me and listening to the problems I encountered. Thank you for showing interest in my thesis, and giving me a push whenever I needed it.

Table of Contents

Index of Tables

Index of Figures

Abstract

This study investigates the formant, fundamental frequency, and speech tempo parameters in a forensic speaker identification setting and whether these are adequate features to use in an auditory-acoustic analysis. Furthermore, the ‘telephone effect’ as described by Künzel (2001) is examined and analysed in terms of whether it applies to the aforementioned phonetic features. A closed set of samples, made up of recordings from the DyViS database, was used in order to determine the adequacy of the chosen parameters and the potential acoustic effect of intercept recordings on the correct identification of a disputed and a non-disputed sample. Measurements of F1, F2, and F3 were taken from between 11 and 26 tokens per speaker for three stressed vowels, along with measurements of the mean and standard deviation of F0, and articulation rate. The results showed that all three parameters proved to be efficient and appropriate for forensic speaker identification practices, but that the articulation rate of the disputed sample was heavily affected by a task-effect. In terms of the intercept recordings, F1 values, especially those of close vowels, were found to be affected, consistent with Künzel’s findings (2001). Mean fundamental frequency values were not altered by the intercept sample, but the standard deviation was, resulting in values twice as high compared to direct recordings.

1. Introduction

Forensic Phonetics is one of many examples of the ways in which phonetics can be applied. This specific branch of applied linguistics can be defined as being part of one of the three sub-branches of the field of forensic linguistics, that of Language as Evidence – the other two being Language and the Law, and Language and the Legal Process. More specifically, it deals with the analysis of spoken evidence which can be used as such in a court of law. This area involves many different tasks and the present study will be focusing on that of forensic speech comparison, in which a disputed and a non-disputed sample are subjected to analysis with the aim of finding out whether they have been produced by the same individual or not. Nolan defines the term, which was at the time known as 'forensic speaker recognition' as "any activity whereby a speech sample is attributed to a person on the basis of its phonetic-acoustic properties" (Nolan 1994: 328). Over a series of disputes regarding the terminology attached to this task, both the terms 'forensic voice comparison' and 'forensic speaker comparison' can be found, as the debate has yet to be settled.

Regardless of the terminology that is used, a distinction still has to be made between the different tasks that 'forensic voice/speaker comparison' entails. Nolan, in reference to technical speaker comparison, makes the distinction between that of speaker identification and verification. According to Nolan, speaker verification is that task in which "an identity claim by an individual is accepted or rejected by comparing a sample of his speech against a stored reference sample spoken by the individual whose identity he is claiming" (Nolan 1983: 8). Speaker identification, on the other hand, refers to that task in which "an utterance from an unknown speaker has to be attributed, or not, to one of a population of known speakers for whom reference samples are available" (Nolan 1983: 9).

The present study will concern itself with the latter and will try to show that, through the application of forensic speaker identification in the form of an auditory-acoustic analysis of specific features and parameters, a disputed sample can be attributed to a non-disputed sample in order to identify its speaker. Moreover, the disputed sample will also be in the form of an intercept telephone call which will allow for a further examination of the 'telephone effect' that has been proven to affect the signal and quality of speech samples by research in the past (Künzel, 2001, 2002; Lawrence, Nolan, & McDougall, 2008; Nolan, 2001; Rose, 2003). The existence of the same disputed sample in the form of both a direct recording and an intercept recording will, therefore, allow for an immediate comparison between the two, in order to examine the effect even further.

Hence, this study has two research questions. First of all, it is concerned with whether the formants, fundamental frequency, and speech tempo parameters are appropriate features to be used in common forensic speaker identification practices. Secondly, this study also concerns itself with the effect of intercept telephone recordings and wants to explore what this effect consists of, to which features it extends, and whether it impedes the application of the affected parameters in order to reach similarity or dissimilarity conclusions regarding disputed and a non-disputed sample in a closed set of samples. For the first research question, the hypothesis, then, would be that the parameters being explored can be used in order to come to a similarity judgement for an unknown and a known speaker in a forensic auditory-acoustic analysis. As for the second research question, the hypothesis would be that the intercept recording has an effect on the acoustic analysis and that it impedes or confounds the similarity judgement for a non-disputed sample and a disputed sample being produced by the same speaker.

2. Literature Review

The study carried out by Gold and French (2011) provides a survey of the international practices of forensic speaker comparison. The paper reveals information regarding the preferred method of analysis and the frameworks in which the results are described, however of particular interest to the present study are the results it shows of the features that are most frequently examined – both phonetic and non-phonetic. As for the former, they found that "all respondents analyse vowel and consonant sounds" (300), the majority of which evaluated the auditory quality of vowels and the formants. The study further states that all respondents "undertaking formant examinations […] measure the second resonance (F2); 87% […] reported measuring F1 and an equal percentage reported measuring F3" (300). The survey further points out that fundamental frequency is routinely measured and, more specifically, its mean and standard deviation (301). Interestingly, it is indicated that "a large proportion [of respondents] point out that [the analysis of the fundamental frequency] is usually of little help" and that it is "'usually used as an elimination tool rather than an identification tool'" (301). The study concludes by stating that the results obtained indicate that there is certainly a lack of consensus regarding the type of analysis that is used, the type of features that are examined and the way in which the results are presented. This dissensus can be due to personal preference of the expert, but also due to restrictions and/or regulations imposed on the professionals by the organisations in which they work or the government of the country in which they operate (303-304).

Also of interest is the study by Künzel (2001) regarding what he refers to as the 'telephone effect'. The findings suggest that F1 is significantly altered in recordings of speech over telephone. More specifically, it was found that "the F1 centres of each vowel measured higher in the telephone-transmitted data as compared to the data recorded directly" and that "the difference is largest for close vowels […], medium for vowels such as [e] and [o] and smallest or zero for open vowels" (89). The reason for this difference is that the transmission channel creates a slope which affects frequencies – and, thus, formants – below 400-500 Hz (82). Because of this, "the amplitude […] of these harmonics, will decrease if they come within the slope" and "[t]hus, the relative weight of the higher harmonics of a formant […] will be increased" (82-83). Künzel then concludes that "[e]rrors due to unnoticed formant shifts may ultimately lead to false identifications or false rejections" and that, consequently, "in cases containing both types of speech recordings [direct and telephone] centre frequency and bandwidth of F1 should not be used for the analysis" (93-94).

In response to the study by Künzel (2001), Nolan published an article in 2002 in order to object to the former's strong claims against the use of formant analysis in forensic speaker identification, since "the conclusion he draws […] could be read as endorsing a complete exclusion of formants from the FSI [forensic speaker identification] process, at least when one of the samples to be compared is telephone speech and the other is not" (74). In reference to Künzel's data, Nolan argues that the difference in the raising of the F1 estimates is very similar for both males and females, but that the amount of cut-off should, in fact, be much lower for females, "since all their formant frequencies tend to be higher than those of male speakers […] and therefore further from the low frequency cut-off" (75). Nolan, therefore, concludes that "proximity of the real formant frequency to the low frequency cut-off is not the only factor at work" (75), and that this phenomenon might be explained by the higher fundamental frequency and the greater separation of harmonics which are characteristic of female speech. Despite the slight issue, Nolan agrees on the fact that we can "expect an overestimate of the frequency of F1 if the speech being measured is telephone speech" (76). As for Künzel's claims regarding the exclusion of formant analysis from FSI practices due to the fact that they "may ultimately lead to false identifications or false rejections" (Künzel, 2001: 93), Nolan points out that – whichever the parameters are that are being analysed – "this danger of being misled by apparent correspondences or discrepancies is […] ever present in FSI" (77). In order to argue in favour of the use of formant analysis in FSI, Nolan suggests that first formant measurements should be excluded and that second formant measurements should be particularly focused on instead, considering that the latter seem to be unaffected by telephone cut-off. This conclusion is even more favourable due to the fact that "F2 is probably more sensitive than F1 to speaker characteristics" (77) anyway. Nolan further argues that realistic forensic speaker identification methods – as opposed to automatic speaker recognition techniques – involve human intervention on behalf of an expert, which works in favour of the general use of formants in FSI practices, since it allows for obvious errors to be observed and for referring to standard acoustic-phonetic models to determine the rough frequencies of particular formants in particular phonetic contexts (79). That is, experts have at their disposal an entire range of "established models of allophonic, stylistic, dialectal, sociolinguistic, and indeed random, variation" (79) to which they can refer and Nolan argues that Künzel, with his research, "has now added […] a model for variation due to telephone transmission characteristics", if ironically, which will "allow safer interpretation of telephone samples" (79). As for the use which can be given to F1, Nolan argues that, if it were the case that a given sample consistently shows lower F1 values in a telephone sample when compared to a direct sample, this would certainly tip the balance in favour of a 'different speakers' hypothesis, thus giving a role to F1 in forensic speaker identification practices (80). Nolan concludes his response by stating that "[i]f the cry is 'formants are unreliable …', [he is] content with that – as long as it is followed by '…just like all methods of speaker identification'" (82).

Künzel (2002), in turn, added to Nolan's (2002) article by publishing a rejoinder to the latter's response. Künzel claims that he does not "conceive of any way in which [his] wording might be regarded as a global interdiction on the use of F1, let alone of spectral information in general, for FSI" (83). As for the fact that the F2 remains unaffected by the telephone effect and Nolan's suggestion to pay particular attention to it for this reason, Künzel states that "one should not forget that even F2 […] may well be affected by factors such as increased loudness of speech, which is quite common when there is heavy ambient noise" (84). In relation to Nolan's claims regarding the apparent superiority of expert human intervention as opposed to automatic speech recognition, Künzel reminds us that the reason for using – or, at least, trying to use – formant tracking algorithms is so as to avoid the variability of the criteria applied, which is the negative side of formant analysis by a skilled human observer (84). In regard to the artefact found in one of the speaker's telephone recordings in Künzel's previous study (2001), he argues that – while this is an issue that can easily be spotted – there might be others which might potentially affect speech formants and that might prove harder to identify (84). Künzel further suggests that "data compression and other characteristics of GSM transmission may involve even worse effects" (84), thus predicting the results obtained in subsequent research (see Guillemin, & Watson, 2008). Künzel concludes that his many years of experience with spectrograms in a forensic speech context make him argue in favour of a critical use of 'static' parameters such as bandwidths and average centre frequencies – and, especially, F1 – for the direct comparison of speakers whenever telephone intercept recordings are being used (85).

In regard to the importance of speech tempo for forensic phonetic practices, Künzel's (1997) study is of particular interest. Participants were tested in three speaking conditions, each of which was recorded in two recording conditions – talking face-to-face to the researcher and talking over the phone with the researcher who was situated in another room – in order to examine whether the speaking and recording conditions had an influence on a number of temporal variables. The three speaking conditions were 'spontaneous', 'semi-spontaneous' and 'read' (54). The results that were obtained in this study suggested that "eliciting and recording a speech sample via the telephone rather than directly in a face-to-face situation has had no influence upon parameters of speaking tempo" (77) and that the analysis of both syllable rate –

or speech rate – and articulation rate "also revealed that the percentage of pausing in speech remains unaffected" (77), therefore establishing that the recording condition did not alter the results at all. One difference that was observed, however, was that of an increased usage of filled pauses in the telephone recording (77). In light of these findings, Künzel comes to the conclusion that "there is no clear-cut basis for arguing for or against the use of telephone recording of reference samples for forensic speaker identification" (78). Regarding the use of speech tempo as part of forensic phonetic practices, Künzel concludes that "SR [syllable rate] is of lesser speaker-specific value [compared to articulation rate or AR] due to the different types, numbers and durations of pauses that this parameter may contain" and that the "speaker-specific power of AR is of special interest" (79) because it appears not to be affected by the speaking condition.

A study carried out by Lawrence, Nolan and McDougall (2008) highlights the acoustic and perceptual effects of telephone transmission on three Standard Southern British English vowels. Considering the small amount of research that has been conducted in terms of whether auditory techniques of forensic analysis are affected by the telephone effect similarly to acoustic analysis, their study investigated how reliable phoneticians' auditory analysis of telephone recordings was and its implications for forensic phonetics (162). The researchers, thus, attempted to examine whether the 'telephone effect' as described by Künzel (2001) also affected the perception of different vowels over the telephone (164). The authors continue by establishing that, nowadays, there is a general agreement over the fact that "techniques from [both the auditory and the acoustic approach to forensic speaker comparison] should be used in combination" (164). Considering that when each of these approaches is used alone shows some clear disadvantages, using them both together they can complement each other (164). The importance of auditory practices is stressed further in the study by drawing on the fact that these provide information about "segmental features, intonation, rhythm and fluency, and geographical, accentual and idiosyncratic features of a speaker's voice" (164). This kind of approach also has its flaws, such as the fact that it is based on the hypothesis of idiolects; that is, the idea that "each individual can be characterised by a unique set of pronunciations of the words of his or her language" (165). This misconception is not unlike that of 'voiceprints' as being equivalent to fingerprints in the acoustic approach, an idea that was particularly prevalent when this kind of analysis was first used for identifying speakers (166). The use of acoustic analysis has also been questioned due to the large amount of potentially extraneous factors affecting the signal, such as "the quality of the recording, background noise, or other distortions" (166). Lawrence, Nolan and McDougall point out, though, that these problems can easily be accounted for by means of an auditory analysis, which is needed in order to select the material and to analyse spectrograms – so as to identify any errors committed by automatic formant trackers, for instance (166). The researchers conclude, then, that "[l]istening should always be supplemented by acoustic analysis, and acoustic analysis should always be monitored through auditory processes and skilled interpretation" (166). The results that the experiment returned in regard to whether the 'telephone effect' was shown perceptually apart from being shown acoustically were not so clear-cut (184). No significant differences between auditory judgements of direct and telephone tokens of the vowels /iː/ and /uː/ were found, but were for the vowel /æ/ (184). These findings suggest that "listeners may not hear much difference in vowels transmitted over the telephone" (184). Lawrence, Nolan and McDougall account for this by suggesting that the listeners may have been working with pre-theoretical intuitions about the quality of the vowels; that is, they may have been plotting the vowel they were presented with where they would expect them to be found in the vowel space (185). Another possibility is given, which is that the listeners might have had an idea of the area in which the vowels would be, but that, within that area, they were merely guessing (186). This would explain the little difference being observed between how directly-recorded and telephone-transmitted vowels were plotted (186). The researchers conclude, therefore, that “the effect of the telephone on perceptual judgements appears to have less of an impact than might have been expected” (186). Considering, however, the acoustic evidence for the shift in F1 of close vowels due to the ‘telephone effect’, the findings presented in the study at hand raises questions about how reliable auditory analysis of vowels is in direct and in telephone recordings, but they “do not detract from the many advantages offered by auditory techniques […] and are […] not being offered as evidence that auditory techniques should be abandoned from a combined auditory-acoustic approach to forensic speaker comparison cases” (187). What is certainly the case is that caution needs to be taken when dealing with intercepted telephone recordings and when forensic speaker comparison practices are being carried out (188).

3. Methodology

Four recordings were taken from the DyViS database to serve as samples for this study. Three recordings served as non-disputed (or ‘known’) samples and one served as a disputed (or ‘unknown’) sample. All samples were produced by male speakers of Standard Southern British English aged 18-25. The study, thus, mimics the set-up that one could find in a real case. It is concerned with a closed set of samples in which “it is known that the speaker to be identified is among the population of reference speakers” (Nolan, 1983). The three non-disputed samples were taken from the interview task, in which the participants were asked to take part in a mock interrogation by the police. The disputed sample, on the other hand, was taken from the telephone task, in which the participants were asked to have a conversation with an ‘accomplice’ over the phone (Nolan, McDougall, de Jong, & Hudson 2009). The goal was, then, to compare the disputed sample to the non-disputed samples in order to show that, through forensic speaker identification, the latter could be attributed to one of the three ‘known’ samples. In both the interview and the telephone task the same kind of information was elicited, so that the samples could be used in direct comparison with each other by examining the same features in the same contexts.

The features that were examined in the auditory-acoustic analysis of the recordings were among those that have been pointed out as the most frequently analysed internationally in the survey by Gold and French (2011), and those discussed by Baldwin and French (1990). In particular, the phonetic features that the present study was concerned with were fundamental frequency – or F0 (see Hudson, de Jong, McDougall, Harrison, & Nolan, 2007 for an analysis of this feature on samples taken from the DyViS database) –, vowel formants and speech tempo. For the analysis of the former, 13-26 instances of the front vowels /æ/, /ɛ/ and /iː/ had been selected for each speaker – except for the vowel /æ/ of the second speaker, for which, due to limited availability of tokens, 11 samples were selected –, since the F1 and F2 formant values for these are further distanced from each other than is the case with back vowels and are, thus, easier to interpret. All instances that were analysed contained these vowels in stressed position.

Apart from the direct recording of the telephone task, an intercept recording was made of each of the participants and incorporated in the DyViS database, so the ‘telephone effect’ as described by Künzel (2001, 2002) could be evaluated both directly and indirectly. That is, the consequences of this effect can be examined in terms of whether it affects the successful analysis and consequent attribution of the disputed sample to one of the three non-disputed samples by comparing the telephone recording to the interview recordings, but it also allows for further exploration of the phenomenon by comparing the direct recording – referred to in the DyViS database as the ‘studio’ version – and the telephone sample directly. The relevance of this type of analysis to forensic speaker identification has already been revealed in a study by Lawrence, Nolan and McDougall (2008).

For the analysis of the formant values, once the different tokens for each vowel for each speaker had been obtained and a steady interval in the middle of the vowel had been selected to avoid any consonantal transitional effects as much as possible in Praat, a script was run in the same software which extracted the formant values of the intervals on the TextGrid files for each speaker. That is, each TextGrid contained three interval tiers – one for each vowel – and the script extracted the values for each interval in each tier individually. For the intercept version of the disputed sample, great care was taken to make sure that the tokens being selected for analysis corresponded as closely as possible to those from the direct version, so that results from the comparison of the two recordings would be comparable in an as objective way as possible. Once the values had been obtained, obvious outliers were removed.

For the analysis of the fundamental frequency of the speakers, the recordings were cut to a length of 3-4 minutes, removing most silences, instances of laughter and coughing, and the voice of the interviewer – in the case of the non-disputed samples, which were taken from the interview task – and the alleged accomplice – in the case of the disputed sample, which was taken from the telephone task. An extraction of the mean and standard deviation of F0 was then performed using the built-in "Analyse periodicity - To Pitch…" functionality in Praat, with the minimum pitch value set to 75 Hz and the maximum pitch value set to 300 Hz. The required values were then obtained by selecting the resulting Pitch file and selecting the option "Query - Get mean…" and "Query - Get standard deviation…".

Finally, for the speech tempo parameter, between 151 seconds and 164 seconds of speech were extracted for each speaker – this time also for the direct recordings of the telephone task of the first two speakers.[1] Then, the parts where the speaker was talking were manually

[1] The decision to incorporate the direct recordings of the telephone task for all speakers is one which resulted from some of the results that had been obtained; namely, a significantly higher articulation rate value had been observed for the unknown speaker's telephone recording while comparing it to the known speakers' interview recordings. In order to explore the possibility of a task-effect, the aforementioned samples were also analysed.

marked as ‘sounding’ intervals on a TextGrid and those where the speaker was pausing – whether these were filled or unfilled – were marked as ‘silent’. The ‘sounding’ intervals were then manually analysed in terms of the number of syllables produced. Finally, a Praat script was run in order to calculate the phonation time of each speaker, which was done by subtracting the ‘silent’ intervals from the total duration of the excerpt. From this information, the following measures were calculated:

- The articulation rate, by dividing the total number of syllables produced by the speaker by the phonation time of the excerpt.
- The average syllable duration, by dividing the phonation time of the excerpt by the total number of syllables produced by the speaker.

Articulation rate is commonly defined as “the number of syllables in an utterance divided by the utterance duration *excluding pauses*” (Rose, 2002: 169). Laver (1994) further states that articulation rate “[excludes] any silent pauses, but [includes] non-linguistic speech material such as filled pauses and prolongations of syllables” (1994: 539), yet Künzel (1997) calculates articulation rate as “number of syllables/ [duration - combined duration of *all pauses*] [emphasis added]” (1997: 56). It appears, therefore, that a consensus has yet to be reached on whether filled pauses are to be included in the analysis of articulation rate or not. In the present study, Künzel’s stance was taken and the duration of both silent and filled pauses combined was subtracted from the total duration of the utterances. Speech rate, while being a speech tempo parameter, has been shown to be of considerably less speaker-specific value (Künzel, 1997) and has, therefore, not been included in the present study as an analysed feature.

While carrying out the acoustic analysis, a simultaneous auditory examination was performed. The importance of this kind of analysis has been pointed out by Lawrence, Nolan and McDougall (2008) in their study on the effects of the ‘telephone effect’ on the perception of vowel quality. Auditory techniques allow for information to be made available to the phonetician about certain features of a speaker's voice, such as intonation, rhythm and fluency (2008: 164). Nowadays, however, considering that both auditory and acoustic techniques have their flaws when used in isolation, a combination of both in an auditory-acoustic approach should ideally be used when carrying out forensic speaker comparisons (Baldwin, & French, 1990; Lawrence, Nolan, & McDougall, 2008; Rose, 2002, 2003). Baldwin and French (1990) point out, however, that they would "accept the auditory/acoustic approach as helpful, so long as the balance were inclined to the auditory" (1990: 9). Rose (2002), in light of a more widely accepted point of view, argues that "both approaches are indispensable: the auditory analysis [...] is of equal importance to its acoustic analysis, which the auditory analysis must logically precede" (2002: 35). For this reason, the present study makes use of a combination of auditory and acoustic practices. Auditory techniques were used to initially identify the different vowels and to make sure that the tokens were comparable across speakers. Distortions or anomalies in the recordings were also accounted for and excluded from the subsequent acoustic analysis of the samples.

4. Results

4.1. Formant analysis using the studio recording

In order to interpret the different formant values obtained while performing the analysis of the data obtained from the five different recordings – the three known samples and both the direct and the intercept telephone recordings for the 'unknown' sample which belonged to one of the three known speakers –, nonparametric tests for independent samples were carried out.[2] More specifically, Mann-Whitney U tests were used in order to find out whether the differences

[2] In order to determine which kind of tests would have to be used, Shapiro-Wilk tests were carried out to evaluate the normality of the distribution of the data. Considering that most distributions were not normal, Mann-Whitney U tests were chosen and used for the statistical analyses.

between the F1, F2 and F3 values of the known samples were statistically significant or not when compared to the studio recording of the unknown sample. In other words, the aim of these tests was to explore whether the alternative hypothesis – which states that differences will be observed between the different speakers when an auditory-acoustic analysis is carried out – can be corroborated or not. The results of these tests can be observed in Table 1 below, in which the *p*-values of the comparisons are given. The unknown speaker will, from here on, be indicated as 'speaker 4', and the known speakers will, thus, be indicated as 'speaker 1', 'speaker 2' and 'speaker 3'. The values that have been marked in grey in Table 1 represent those that show statistically significant differences between the formant values of the two speakers that are being compared in each case.

Table 1. The *p*-values obtained from the Mann-Whitney U tests for the formant values of the three known samples compared to the unknown sample, indicated here as 'speaker 4'. The significance level is .05.

Vowel		***p*-value**		
		Speaker 1 vs 4	**Speaker 2 vs 4**	**Speaker 3 vs 4**
/æ/	F1	*p* = .062	*p* = .148	*p* = .382
	F2	*p* = .001	*p* = .000	*p* = .958
	F3	*p* = .000	*p* = .451	*p* = .345
/ɛ/	F1	*p* = .001	*p* = .005	*p* = .008
	F2	*p* = .529	*p* = .017	*p* = .227
	F3	*p* = .001	*p* = .000	*p* = .104
/iː/	F1	*p* = .126	*p* = .000	*p* = .001
	F2	*p* = .000	*p* = .732	*p* = .002
	F3	*p* = .000	*p* = .000	*p* = .003

The comparison between ‘speaker 1’ and ‘speaker 4’ shows six instances of statistically significant differences between the two, in which, overall, $p \leq .001$. The comparison between ‘speaker 2’ and ‘speaker 4’ equally shows six instances of statistically significant differences, in which $p \leq .005$, except for the F2 formant values for the vowel /ɛ/ of the two speakers, which has a *p*-value of .017, but is still significant at the .05 level. Finally, the comparison between ‘speaker 3’ and ‘speaker 4’ yielded four statistically significant formant results, with a value of $p < .05$. Thus, the comparison which shows the least amount of statistically significant formant values is the comparison between ‘speaker 3’ and ‘speaker 4’, which seems to suggest that these two speakers are the most similar in terms of their formant values. Taking into account that these two speakers are, in fact, the same, the formant analysis appears to correctly predict their similarity.

In order to further represent the significance of the formant values that have been obtained and to back up the results that have been obtained from the Mann-Whitney U tests, different scatterplots were created from the individual values of the three speakers that had been obtained. As can be observed in Figure 1 to Figure 3, the formant values visually corroborate the results.

Of interest are the *p*-values that have been obtained for the comparison between ‘speaker 3’ and the unknown speaker for the vowel /iː/ for all three formants, which are all statistically significant. From the boxplot in Figure 4, it becomes clear that the values for these two speakers are contained within more or less the same range, but that for ‘speaker 3’ these are found within the lower half of that range, and that for ‘speaker 4’ these are found within the higher half.[3] Bearing in mind that the unknown sample corresponding to ‘speaker 3’ was taken from a

[3] As can be observed in the boxplot in Figure 4, the four uppermost F1 values have been identified as outliers and the range, therefore, is not as clearly visible. Looking at all the individual F1 values, however, the researcher does not consider these to be outliers and simply a misjudgement on behalf of the statistics software used, SPSS.

different task – that of a telephone call with an accomplice task as opposed to a police interview task – it is possible that a task-effect is at play.

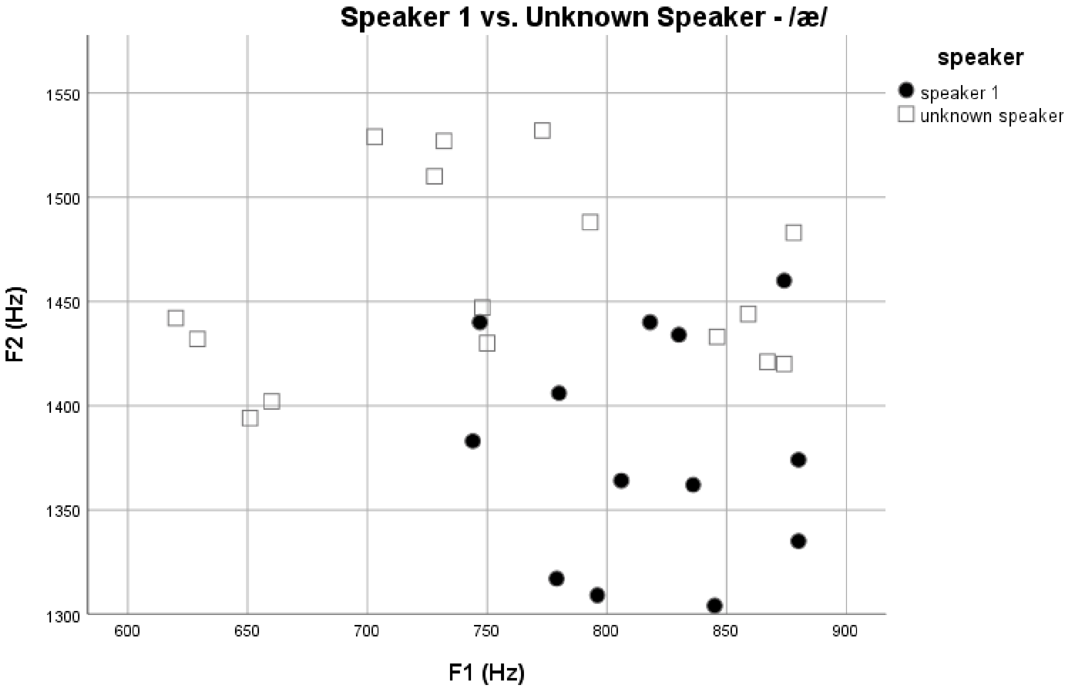

Figure 1. Scatterplot comparing the F1 and F2 formant values for 'speaker 1' and the unknown speaker for the vowel /æ/. The values for 'speaker 1' are represented by a circle, while the values for the unknown speaker are represented by a square.

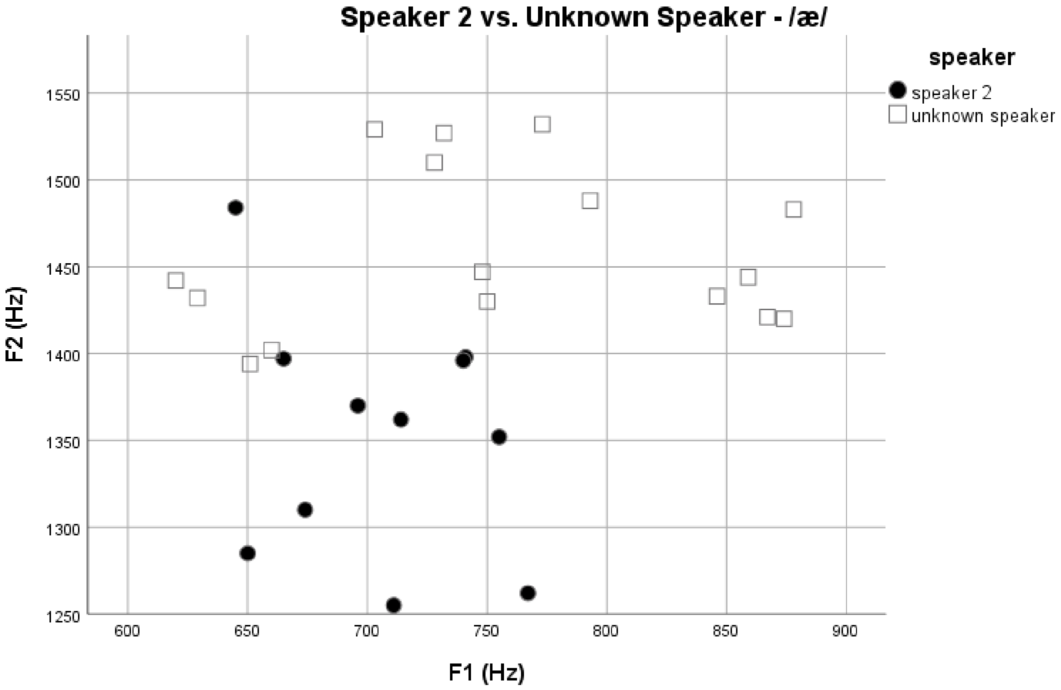

Figure 2. Scatterplot comparing the F1 and F2 formant values for 'speaker 2' and the unknown speaker for the vowel /æ/. The values for 'speaker 2' are represented by a circle, while the values for the unknown speaker are represented by a square.

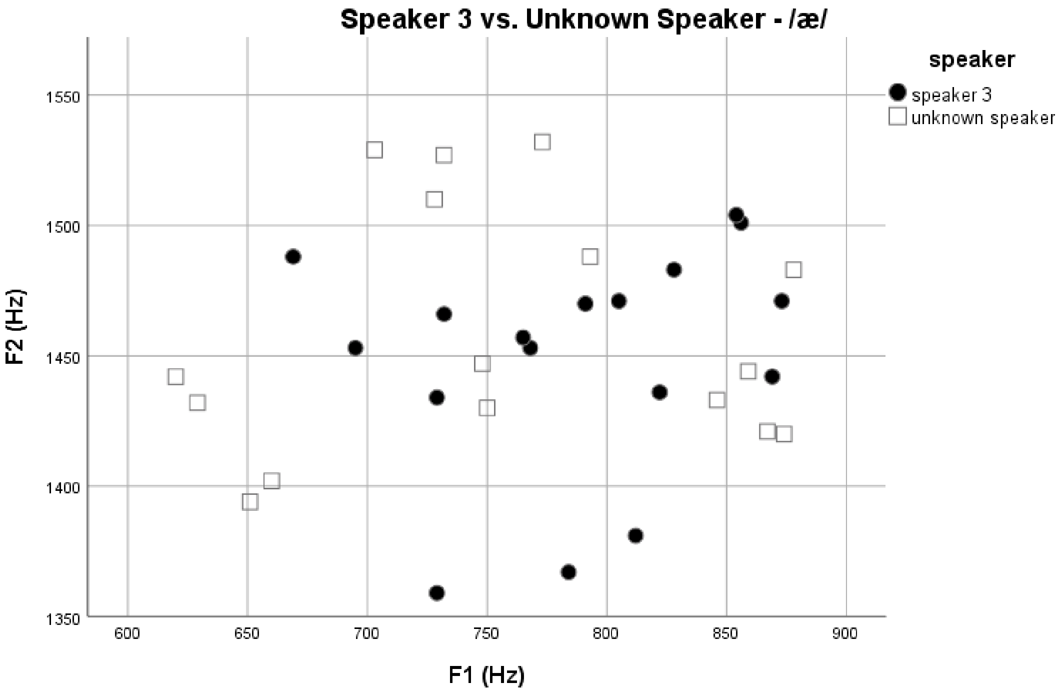

Figure 3. Scatterplot comparing the F1 and F2 formant values of ‘speaker 3’ and the unknown speaker for the vowel /æ/. The values for ‘speaker 3’ are represented by a circle, while the values for the unknown speaker are represented by a square.

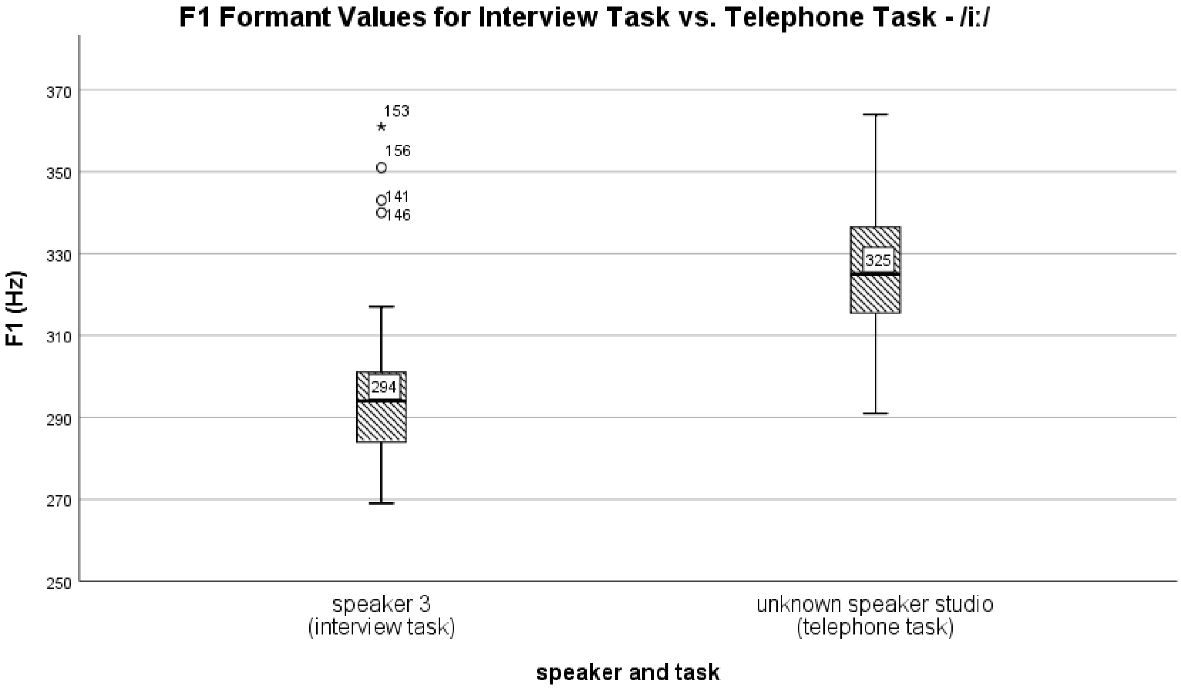

Figure 4. Boxplot showing the range of F1 values for the vowel /i:/ for ‘speaker 3’ in the police interview task and for the same speaker in the phone call with an accomplice task – which is functioning as the unknown sample in this study.

As can be observed in the boxplot in Figure 4, the range in which the F1 values occur is very similar in both tasks, with that of the interview task being 269 Hz - 361 Hz and that of the phone call task being 291 Hz - 364 Hz. The difference between the two conditions, then, lies in the fact that the number of occurrences of F1 values is much higher in the lower end of the range for the former, and that the number of occurrences of F1 values for the latter is concentrated around the higher end of that range. This causes the median for the interview task to be 294 Hz and that of the telephone task to be 325 Hz for the same speaker and the same vowel.

4.2. Formant analysis using the intercept recording

Apart from a formant analysis of the three known speakers and the studio recording of the unknown speaker, a second comparison was carried out between 'speaker 3' – that is, the speaker which corresponds to the unknown speaker – and the unknown speaker. This time, however, an intercept telephone recording was used in order to further investigate the telephone effect which has been found to affect F1 in particular and, more precisely, the F1 of close vowels such as /iː/ (Künzel, 2001, 2002; Lawrence, Nolan, & McDougall, 2008; Nolan, 2002; Rose, 2003). The tokens of the intercept sample used in this analysis match those of the studio recording entirely, in order to make sure that the formant values obtained would be as accurate as possible. Once again, Mann-Whitney U tests were carried out to observe the effect of the telephone recording on the correct identification of the unknown speaker. The results from this analysis can be observed in Table 2. The *p*-values that had been obtained from the comparison between 'speaker 3' and the studio recording of 'speaker 4' have also been included for reference, and the statistically significant results have been marked in grey.

The comparison between 'speaker 3' and the studio version of the unknown sample yielded four statistically significant results in total with a value of $p \leq .008$. The comparison between 'speaker 3' and the intercept version of the unknown sample, on the other hand, returned six statistically significant results in total with a value of $p < .05$.

Table 2. The *p*-values obtained from the Mann-Whitney U tests for the formant values of ‘speaker 3’ compared to both the studio version and the intercept version of the unknown sample, indicated here as ‘speaker 4’. The significance level is .05.

Vowel		***p*-value**	
		Speaker 3 vs 4 (studio)	**Speaker 3 vs 4 (intercept)**
/æ/	F1	*p* = .382	*p* = .040
	F2	*p* = .958	*p* = .331
	F3	*p* = .345	*p* = .005
/ɛ/	F1	*p* = .008	*p* = .482
	F2	*p* = .227	*p* = .210
	F3	*p* = .104	*p* = .003
/iː/	F1	*p* = .001	*p* = .000
	F2	*p* = .002	*p* = .021
	F3	*p* = .003	*p* = .000

While these results show that there appear to be more differences when ‘speaker 3’ is being compared to the intercept recording than when this speaker is compared to the studio recording, the telephone effect as described by Künzel (2001) is yet to be seen. For the purpose of illustrating exactly how the F1 values of the unknown telephone recording have been affected, a comparison between the intercept version and the studio version of the same sample – that is, the unknown ‘speaker 4’ – was carried out. These results were then plotted visually on a scatterplot in Figure 5, in which the F1 and F2 values of both recordings for the vowel /iː/ can be observed. A boxplot has also been included in Figure 6, in which the difference in range of the values between the two versions of the recording can be appreciated.

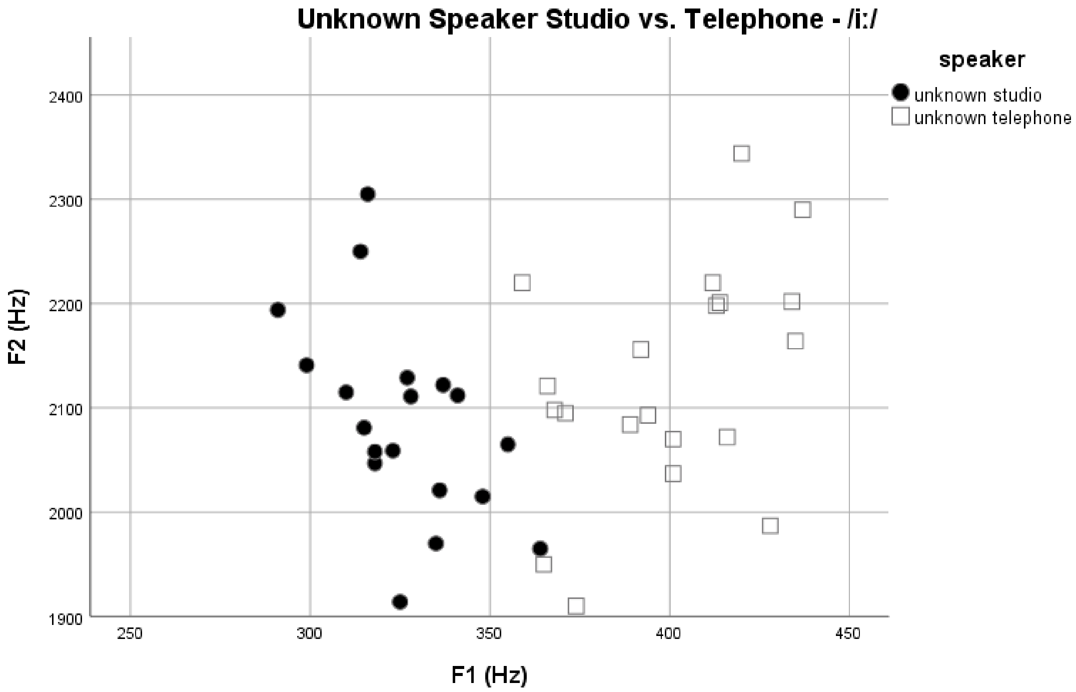

Figure 5. Scatterplot comparing the F1 and F2 formant values of the direct and intercept recording of the unknown speaker. The values for the direct recording are represented by a circle, while the values for the intercept recording are represented by a square.

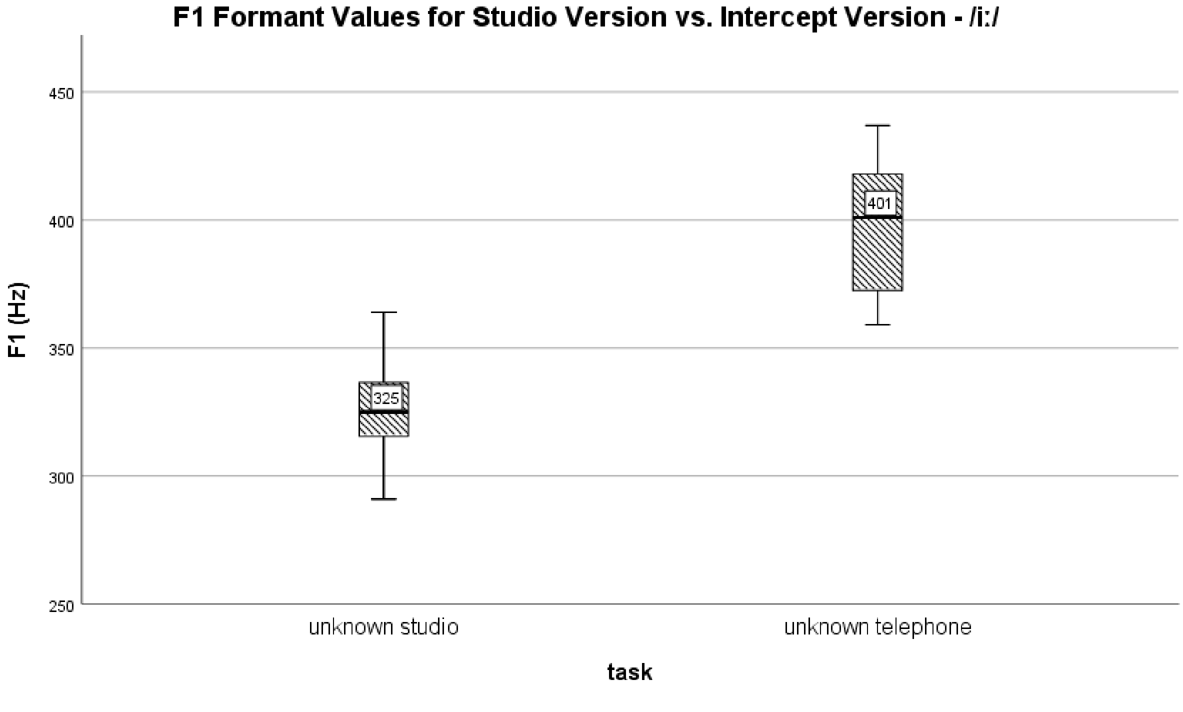

Figure 6. Boxplot illustrating the range of the F1 values of the studio version and the intercept version of the telephone recording for the unknown speaker.

Whereas the F2 values for both recordings of the same speaker for the vowel /iː/ are all very similar, the biggest difference can be observed in the F1 values of the intercept recording, which are clearly much higher than those of the direct recording. This finding is, thus, in line with those that have been obtained in previous studies (Künzel, 2001, 2002; Lawrence, Nolan, & McDougall, 2008; Nolan, 2002; Rose, 2003).

4.3. Fundamental frequency

Another phonetic feature that was examined in this study is that of fundamental frequency or F0. For this particular feature, the mean and standard deviation were calculated in line with the common forensic speaker comparison practices that have been described by Gold and French (2011) in their international survey. In Figure 7, the mean F0 values for all speakers can be observed, and in Figure 8 the standard deviation. Both the mean F0 value and the standard deviation for the intercept version of the unknown speaker have been added as well in their respective graphs, in order to explore any potential effect the telephone recording might have had on the speaker's fundamental frequency.

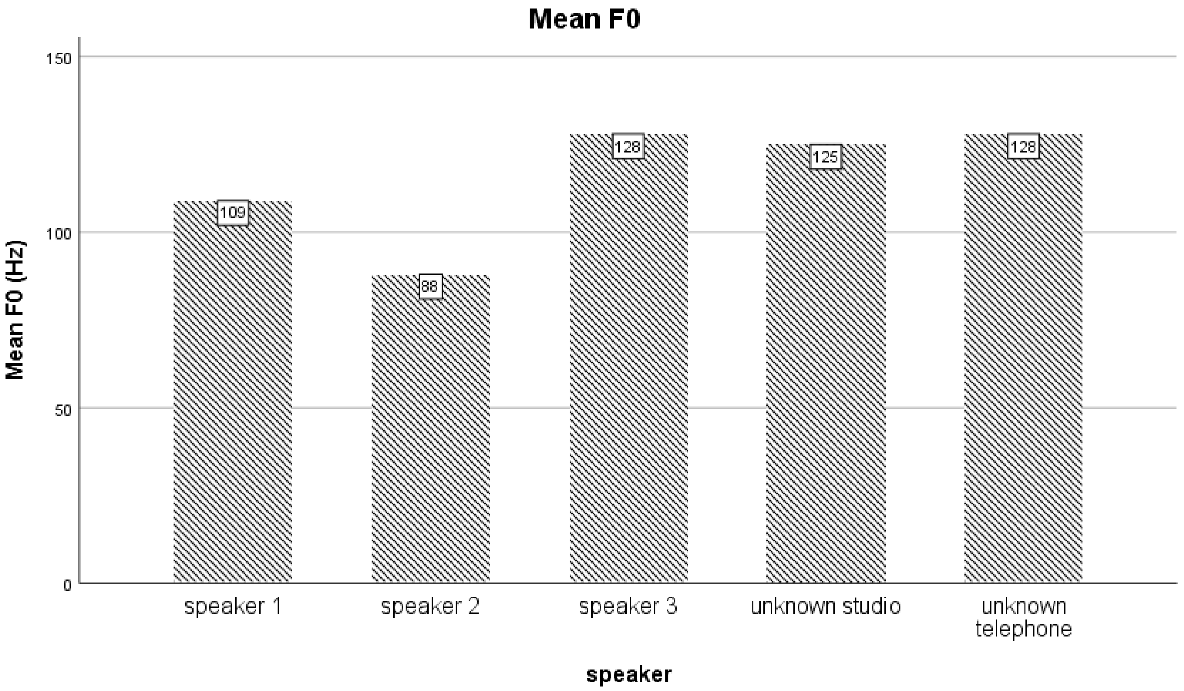

Figure 7. Bar graph of the mean F0 values for all the known speakers and both the unknown studio and the unknown telephone recording.

The lowest fundamental frequency mean belongs to 'speaker 2', which shows an F0 of 88 Hz, whereas the highest mean value can be found in 'speaker 3', whose F0 is very similar to that of the studio version of the unknown speaker and completely matches that of the telephone recording of the unknown speaker, correctly predicting their similarity. Upon comparing the direct and the intercept versions of the recording of the unknown sample, it seems that the telephone recording had no effect whatsoever on the mean fundamental frequency of the speaker.

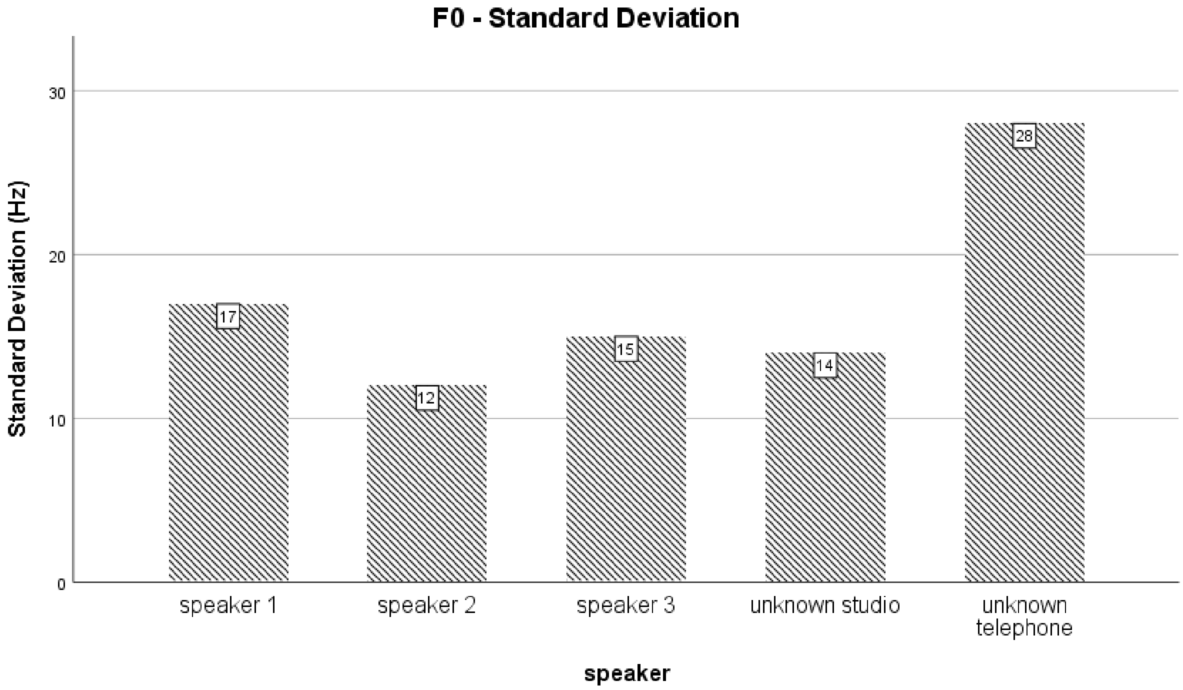

Figure 8. Bar graph of the F0 standard deviation values for all the known speakers and both the unknown studio and the unknown telephone recording.

Of the known samples, ‘speaker 1’ shows the highest standard deviation with a value of 17 Hz, and ‘speaker 2’ the lowest, with a value of 12 Hz. ‘Speaker 3’ shows the most similar result compared to the unknown recording, with a standard deviation of 15 Hz and 14 Hz respectively. This result would, then, accurately predict their similarity. When comparing the direct and the intercept versions of the telephone recording, however, it becomes apparent that the latter shows a much higher standard deviation of 28 Hz, one that is twice as high as that of the studio recording. It seems, therefore, that the mean fundamental frequency is a more reliable measure than the standard deviation when intercept telephone recordings are involved, since they appear to show a considerable upward shift.

4.4. Speech tempo

The last parameter that was examined was that of speech tempo, a suprasegmental feature related to fluency. Its importance for forensic speaker comparison practices has been pointed out by Künzel in a study carried out in 1997, in which the superiority of articulation rate of speech rate in terms of speaker-specific value is also highlighted. Moreover, in the survey carried out as part of their study in 2011, Gold and French state that, of those respondents that analysed tempo, 81% calculates speaking rate and/or articulation rate (2011: 302). Once the samples had been prepared as described in the Methodology section, an analysis of the number of syllables, the articulation rate, and the average syllable length of each speaker was performed. The values that have been obtained can be observed in Table 3 below.

Table 3. The number of syllables, the articulation rate, and the average syllable duration (ASD) for each known speaker and the unknown speaker, specified as 'speaker 4'. The values for 'articulation rate' are presented in syllables per second. The values for 'average syllable duration' (ASD) are presented in seconds.

Speaker	Syllables	Pauses	Articulation rate (syll./s)	ASD (s)
Speaker 1	516	75	4.81	0.21
Speaker 2	528	79	5.21	0.19
Speaker 3	509	71	5.00	0.20
Speaker 4	543	67	6.11	0.16

The unknown 'speaker 4' appears to have a much higher articulation rate than any of the known speakers, therefore not showing any direct similarity with any of these samples. Considering that the sample for the unknown speaker was taken from a different task than the known ones – a telephone call task as opposed to an interview task – an additional analysis was carried out in which the possibility of a task-effect altering the results was explored. In this comparison, the articulation rates of the interview task samples for 'speaker 1', 'speaker 2', and 'speaker 3' were compared directly to those of the telephone task samples for these same speakers. The results of this analysis can be seen in Figure 9 below.

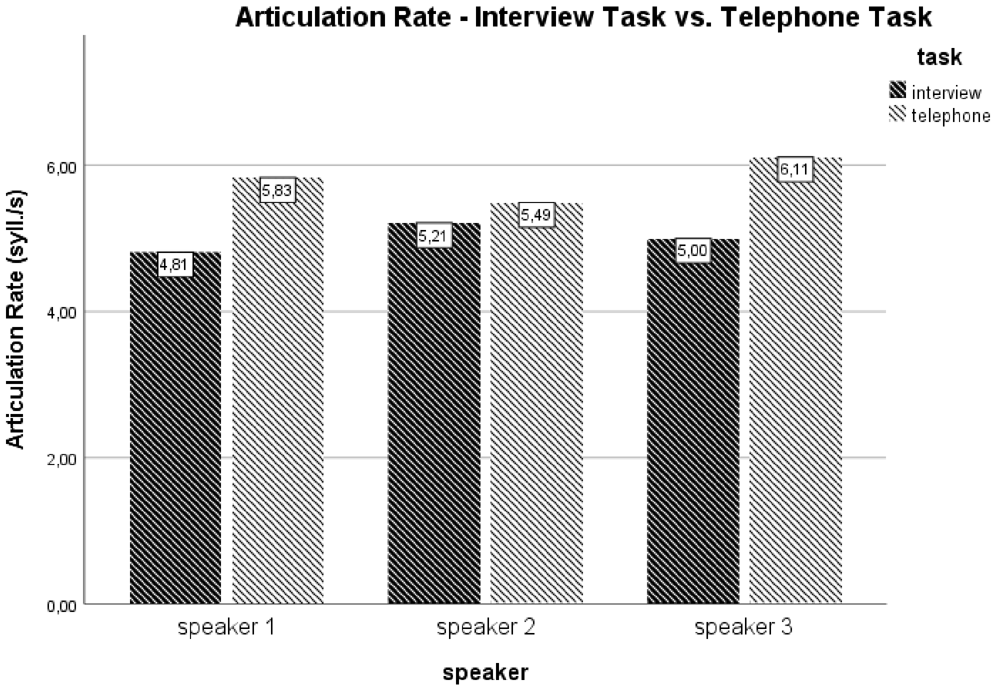

Figure 9. Bar graphs showing the articulation rate in syllables per second for each speaker in both the interview task and the telephone task.

The articulation rate is consistently higher in the telephone task, in which an increase of more than 20% can be observed for ‘speaker 1’ and ‘speaker 3’. ‘Speaker 2’, while still showing a slight increase in articulation rate during the telephone task as opposed to the interview task, does not show as much of an increase as the other two speakers. The feature articulation rate, therefore, seems to be altered by a task-effect, which hinders a similarity between known and unknown speakers to be observed.

5. Discussion

The aim of this study was to find out whether a sample of an unknown speaker could be attributed to another known speaker in a closed set of samples. For this purpose, three known speakers were compared to an unknown speaker in an auditory-acoustic analysis. The features that were analysed in this comparison were that of fundamental frequency, formant values and speech tempo. In common forensic speaker comparison practices, these are just three of many other features that can and tend to be analysed – segmental and suprasegmental phonetic features, and even non-phonetic features related to, for instance, discourse (2011: 300-302)–, as has been indicated in the survey carried out by Gold and French in 2011. Whether these features show similarities or differences between speakers, therefore, does not allow us to reach definite conclusions, but rather provides us with clues that, in turn, need to be verified by further analysis. That is, each individual analysis constituted a small part, and once all these parts were put together they would allow for a bigger picture to be constructed, which is that of being able to assert the partial similarity or dissimilarity of the different speakers. Only with a very complete analysis that includes an array of different analysed features can one reach a conclusion. The three parameters, then, that are being looked at in this study, again, are only a small part of this comparison.

First of all, a formant analysis was carried out. Once the formant values had been obtained as described in the Methodology section, they were subjected to Mann-Whitney U tests in order to find out whether any statistically significant differences could be found between the F1, F2 and F3 of the vowels /æ/, /ɛ/ and /iː/ of the three known speakers and the unknown speaker (see Table 1). Both the comparison between ‘speaker 1’ and the unknown speaker and ‘speaker 2’ and the unknown speaker yielded six significant differences with *p*-value ≤ .005, with the exception of the comparison of the F2 values of ‘speaker 2’ and the unknown speaker for the vowel /ɛ/, which has a *p*-value of .017 – a result which is still statistically significant. The comparison between ‘speaker 3’ and the unknown speaker, on the other hand, returned only four statistically significant differences, three of which can be considered significant due to their *p*-value being < .005. The fourth significant difference returned a value of p = .008. Considering that ‘speaker 3’ and the unknown speaker are the same person, these results, therefore, work in favour of the hypothesis that these two speakers are the same and that this can be determined by carrying out an auditory-acoustic analysis of different phonetic features – in this case the analysis and consequent comparison of formant values –, due to the fact that the comparison between these two speakers returned the least amount of statistically significant differences.

An aspect of the results of the tests that is worth mentioning is the statistically significant differences that have been obtained for all three formants of the vowel /iː/ for ‘speaker 3’ and the unknown speaker. Considering that they are the same speaker, these results stand in stark contrast with those that have been obtained for the formants of the vowels /æ/ and /ɛ/ in the same comparison. This occurrence will have to be explained, first and foremost, by the fact that intra-speaker variability does exist – a phenomenon which results such as the ones obtained, in fact, corroborate – and, moreover, that there might be a potential task-effect at play. Intra-speaker variability, also known as within-speaker variability, refers to the fact that “there will always be differences between speech samples, even if they come from the same speaker” (Rose, 2002: 10). That is, the same voice of a certain speaker will always show variability.

Regarding the potential task-effect, the three recordings of the known speakers all consisted of a mock police interrogation, whereas the unknown sample was taken from the second task which the participants performed, which was that of a phone call with an alleged accomplice. It is possible, therefore, that the change in recording condition had an effect on the speaker’s speech, which in this case seems to be the case for the three formants of the vowel /iː/ of the speaker. When observing the individual formant values that have been obtained for, for instance, the F1 of /iː/ of ‘speaker 3’ and the unknown speaker, one can observe that these all occur within more or less the same interval (see Figure 4). The difference lies, however, in the fact that, for the police interview task, these values are all concentrated around the lower half of the range, whereas for the telephone task, these values are all clustered around the higher half of that range. Bear in mind that the recording of the telephone task that is being referred to is the studio, or direct, version and the phenomenon under analysis at this point cannot be attributed to a potential telephone effect – an aspect which will be discussed next.

What is important about the statistical results in regard to how well one speaker can be judged to resemble another in the closed set of samples at hand is that the least amount of statistically significant differences were found between ‘speaker 3’ and the unknown speaker, which corroborates the fact that these two speakers are, indeed, the same. However, since a comparison of formant values is only one of the features that has been analysed as part of the analysis, due to the fact that a judgement of similarity of two voices cannot be passed by looking at one feature only, a second feature was examined; that of F0 or fundamental frequency, which will be discussed later on.

Concerning the formant analysis, still, the telephone effect which has been described to affect the F1 of the recording in previous research (Künzel, 2001, 2002; Lawrence, Nolan, & McDougall, 2008; Nolan, 2002; Rose, 2003) was also examined. First of all, an analysis was carried out of the amount of statistically significant differences that had been obtained from comparing ‘speaker 3’ to the unknown studio sample, and from comparing ‘speaker 3’ to the unknown intercept sample. The results showed that, while the former yielded only four significant differences, the latter yielded six. This already acts as an indicator of the fact that telephone recordings appear to be less optimal to use when carrying out a forensic speaker comparison, since these results make it more difficult to assert the unknown speaker’s dissimilarity from ‘speaker 1’ and ‘speaker 2’, which could be confirmed with more certainty when comparing these two speakers to the direct recording of the unknown speaker.

The relative inadequacy of the telephone recording became even more apparent when the same sample – with an extraction of the values of the exact same tokens – was compared to its studio counterpart, which is when the effect as it is documented by Künzel (2001) became clear. That is, the F1 of the intercept sample was clearly pulled up, resulting in considerably higher values than those encountered in the direct sample. From the scatterplot and the boxplot in Figure 5 and Figure 6 respectively, it becomes evident that the range of F1 values for the intercept recording is much higher than that of the studio recording. This finding, then, further confirms that intercept telephone recordings are not ideal as a sample from which to work in forensic speaker comparison practices, and it seems, then, that whenever a direct sample is available, the latter should be used and preferred over the recording that was intercepted.

The second analysis that was carried out was that of the fundamental frequency, considered "one of the most important parameters in forensic phonetics" (Rose, 2002: 244). Once the recordings had been cut and modified in the way described in the Methodology section, the mean and standard deviation were obtained for each speaker. The mean values could, indeed, be said to be quite decisive, as 'speaker 3' and the unknown speaker reported a mean F0 of 128 Hz and 125 Hz respectively, whereas 'speaker 1' showed a mean F0 of 109 Hz and 'speaker 2' a mean value of 88 Hz. 'Speaker 3' and the unknown speaker clearly, then, appear to have a very similar fundamental frequency. Since these two speakers are, in fact, the same, this parameter can, thus, the same way that the formant analysis did, be considered an adequate tool for determining the similarity and consequent identification of different speakers in a forensic phonetic context. The standard deviation also appears to be a very convincing measure, since the value obtained for 'speaker 3' was that of 15 Hz, and the value for the unknown speaker, 14 Hz. Considering that 'speaker 1' and 'speaker 2' respectively show standard deviations of 17 Hz and 12 Hz, however, means that these results are not as decisive as the mean values, since the former are all relatively similar.

The mean F0 and the standard deviation for the unknown speaker were also analysed for the intercept recording, in order to find out if the telephone effect that was found to affect the F1 values of close vowels (Künzel, 2001, 2002; Lawrence, Nolan, & McDougall, 2008;

Nolan, 2002; Rose, 2003) extended to the F0 of the speakers. The F0 for the intercept sample reported a mean value of 128 Hz – compared to the mean value of 125 Hz of the direct sample – which, thus, shows that the recording condition had little to no effect on this phonetic parameter, and did not impede the judgement of 'speaker 3' and the unknown speaker having very similar mean F0 values, since 'speaker 3' had a value of 128 Hz; that is, both these values coincided entirely. Moreover, these results also confirm that there is no task-effect which could potentially confound the findings – which might have been the case for the formant analysis. The standard deviation, on the other hand, shows a value of 14 Hz for the direct sample, but a value of 28 Hz – that is, one that is twice as high – for the intercept sample. The type of recording, then, seems to influence this parameter greatly, and would not allow for a correct identification of different speakers. It appears advisable, thus, to work with mean F0 values rather than the standard deviation when intercept telephone recordings are being used.

The third and final parameter that was explored was the feature speech tempo. Of particular interest was the analysis of the speakers' articulation rate. As has been described in the Methodology section, the articulation rate was calculated by dividing the number of syllables by the phonation time of the recording. The phonation time, in turn, was obtained by subtracting the pauses from the total duration of the recording. Having obtained the results thus, all individual measures of the known speakers were compared to those of the unknown speaker. From the results, it appears that none of the known speakers show any similarity with the unknown speaker, whose articulation rate of 6.11 syllables per second is much higher than that of the other speakers. It would seem, therefore, that it would be impossible to base our judgement on this analysis, which only emphasises a dissimilarity of the unknown speaker with all of the other speakers in the close set of samples. What is interesting, then, is the fact that the articulation rate for the telephone task sample – that is, the unknown speaker's sample – is considerably higher than for its interview task counterpart – that is, 'speaker 3's sample. It seems possible, therefore, that a task-effect is at play. For this reason and in order to confirm the presence of a potentially confounding variable, an analysis of the articulation rate was carried out for the remainder of the speakers in the telephone task as well (see Figure 9).

The results showed that the articulation rate for all speakers is significantly higher in the telephone task than in the interview task, except for ‘speaker 2’, whose value remains relatively unaffected, yet is still higher in the phone call task. For ‘speaker 1’ and ‘speaker 3’, an increase of more than 20% can be observed. This result, would, therefore, account for the difficulty of asserting the similarity of ‘speaker 3’ and the unknown speaker in terms of speech tempo, and clearly shows that the telephone recording condition has a significant effect on this parameter. This finding is particularly interesting since previous research (Künzel, 1997) has found that “eliciting and recording a speech sample via the telephone rather than directly in a face-to-face situation has […] no influence upon parameters of speaking tempo” (Künzel, 1997: 77). The results obtained in the present study seem to suggest that there is indeed a task-effect which is affecting the speaking tempo of the speakers, and it appears that articulation rate especially is altered drastically. This finding, therefore, appears to be more in line with what Byrne and Foulkes describe as ‘speaker effects’, which refer to the behavioural differences that speakers show when talking over the telephone, such as that of adopting a ‘telephone voice’, possibly resulting in changes to “voice quality, speaking rate, and/or the use of different segmental pronunciations” (2004: 84).

6. Conclusions

The aim of this study was to carry out an auditory-acoustic analysis of three features commonly used in forensic speaker identification practices, namely those of formant values, fundamental frequency and speech tempo, in order to find out whether these were adequate parameters and would result in a correct identification of a disputed speaker to that of a non-disputed speaker in a closed set of samples. Furthermore, an intercept recording of the disputed sample was used to explore the effect established by previous research (Künzel, 2001, 2002; Lawrence, Nolan, & McDougall, 2008; Nolan, 2002; Rose, 2003) that telephone transmission has on the recording itself and the consequences this has for the correct identification in forensic speaker comparisons. The results showed that both the analysis of fundamental frequency and of formant values are viable forensic speaker comparison practices which would result in correct judgements regarding a disputed sample in a closed set of samples. The third feature that was examined, that of speech tempo in terms of articulation rate, did not prove to be as conclusive as the other two parameters, and a direct comparison of these values resulted in the inability to identify the disputed sample with any of the non-disputed samples. The reason for this inconsistency was found to be attributable to a task-effect which was altering the articulation rate of the unknown speaker. This finding was corroborated once all the non-disputed samples' interview tasks recordings were compared to their corresponding telephone task recordings, the results of which showed that for all speakers the articulation rate was higher in the telephone task. This observation therefore stands in contrast with Künzel's statement that eliciting speech in a telephone task has no effect on parameters of speaking tempo (1997: 77).

The findings regarding the intercept telephone recording matched those observed in previous research (Künzel, 2001, 2002; Lawrence, Nolan, & McDougall, 2008; Nolan, 2002; Rose, 2003), and the 'telephone effect' (Künzel, 2001) was clearly visible. An upward shift in the F1 of the close vowel /iː/ was found, which made a formant-based forensic speaker identification using an intercept sample less adequate compared to when a direct recording is used. The mean fundamental frequency, on the other hand, did not appear to be affected by either the task or the recording condition, unlike the standard deviation, which showed a significant upward shift in the intercept sample of the unknown speaker. These results suggest, then, that caution should be exercised whenever one is working with intercept telephone samples, especially when conducting formant analyses, and phoneticians should opt for direct recordings for forensic speaker identification purposes whenever these are available.

While intercept landline recordings are still relevant and frequently used in a forensic phonetic setting, the use of mobile phones has increased significantly in recent years and they are being used more frequently than landline phones. The effect that the mobile phone signal has on the quality of recordings has already been explored in a preliminary study by Guillemin and Watson (2008), in which a significant impact on formant frequencies had been found, similar to Künzel's findings regarding the 'telephone effect'. A study prior to that, by Byrne and Foulkes (2004), had already explored the 'mobile phone effect' on vowel formants and had found that the impact on the F1 was even greater than that of landlines. F3 values were also found to be affected, while F2 values overall remained unaffected. Taking these findings into account, together with the fact that the use of mobile phones is becoming more and more widespread and common, a limitation regarding the experiments that have been carried out in the present study is that of restricting the analysis to intercept landline recordings. Only the latter were used, due to the unavailability of mobile phone recordings for the study. Further research, therefore, could be carried out comparing the same parameters across intercept landline and mobile samples and explore the impact each has on the different features and on the consequent judgements on the similarity or dissimilarity of the speakers.

7. Bibliography

Baldwin, J., & French, P. (1990). *Forensic Phonetics*. London: Pinter Publishers.

Byrne, C., & Foulkes, P. (2004). The 'Mobile Phone Effect' on Vowel Formants. *International Journal of Speech Language and the Law*, 11 (1): 83-102.

Gold, E., & French, P. (2011). International practices in forensic speaker comparison. *International Journal of Speech Language and the Law*, 18 (2): 293-307.

Guillemin, B. J., & Watson, C. (2008). Impact of the GSM mobile phone network on the speech signal: some preliminary findings. *International Journal of Speech Language and the Law*, 15 (2): 193-218.

Hudson, T., de Jong, G., McDougall, K., Harrison, P., & Nolan, F. (2007, August). *F0 Statistics for 100 Young Male Speakers of Standard Southern British English*. Paper presented at the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.

Künzel, H. J. (1997). Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics*, 4 (1): 48-83.

Künzel, H. J. (2001). Beware of the 'telephone effect': the influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguistics*, 8 (1): 80-99.

Künzel, H. J. (2002). Rejoinder to Francis Nolan's 'The "telephone effect" on formants: a response'. *Forensic Linguistics*, 9 (1): 83-86.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lawrence, S., Nolan, F., & McDougall, K. (2008). Acoustic and perceptual effects of telephone transmission on vowel quality. *International Journal of Speech Language and the Law*, 15 (2): 161-192.

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

Nolan, F. (1994). Auditory and acoustic analysis in speaker recognition. In J. Gibbons (Ed.), *Language and the Law* (pp. 326-345). London/New York: Longman.

Nolan, F. (2002). The 'telephone effect' on formants: a response. *Forensic Linguistics*, 9 (1): 74-82.

Nolan, F., McDougall, K., de Jong, G., & Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech Language and the Law*, 16 (1): 31-57.

Rose, P. (2002). *Forensic Speaker Identification*. London: Taylor & Francis.

Rose, P. (2003). The technical comparison of forensic voice samples. In F. Freckelton, & H. Selby (Eds.), *Expert Evidence* (Chapter 99). Sydney: Thompson Lawbook Co.