The ‘Mobile Phone Effect’ on Vowel Formants

Catherine Byrne and Paul Foulkes***

**University of Sheffield, **University of York*

ABSTRACT This study analyses the effect of mobile phone transmission on vowel formant frequencies, based on the study presented by Künzel (2001). Six male and six female speakers read a short passage into a mobile phone. Two simultaneous recordings were made, one at the far end of the phone line and the other via a microphone directly in front of the speaker. Measurements of F1, F2 and F3 were taken from between 15 and 25 stressed vowels per speaker in both sets of recordings. Due to the filtering effect of the phone transmission, F1 frequencies for most vowels were found to be higher than their counterparts in the direct recordings. The overall effect of the mobile phone on F1 frequencies was considerably greater than the landline telephone effect found by Künzel (2001): on average the F1 values in the mobile condition were 29 per cent higher than in the direct condition. On the whole F2 measures were not significantly affected, in line with Künzel’s findings. F3 frequencies were also generally unaffected by the mobile phone transmission. Exceptions were found, however, particularly for individual speakers with relatively high F3s. In these cases the mobile recordings tended to yield significantly lower values. The consequences of measurement errors arising from the different recording conditions are discussed with reference to forensic speaker identification.[1]

KEYWORDS formant analysis, mobile phone transmission, forensic speaker identification

INTRODUCTION

The commonest type of casework undertaken by forensic phoneticians involves comparison of recorded speech samples. A questioned voice in a criminal recording is compared with a reference sample from a suspect in order to assess the likelihood that both recordings were produced by the same person. In the great majority of such cases, perhaps as many as 90 per cent, the questioned sample involves speech that has been recorded after transmission via telephone (Hirson, French and Howard 1995: 230). It is well known that telephone transmission introduces various effects into the acoustic signal (see for example Moye 1979, Rose 2003). Since reference samples are often collected in face-to-face settings such as a police interview, it is likely that very different behavioural, environmental and transmission conditions exist for the questioned and reference samples. Forensic comparison is thus rendered more difficult, as allowances have to be made for acoustic variation resulting from the recording context itself.

© University of Birmingham Press 2004 1350-1771

Speech, Language and the Law 11(1) 2004

A good deal of research has therefore been carried out to assess the specific effects of telephones on the speech signal. The results of such research reveal three main types of effect which together characterize telephone speech.

First, effects may result from the physical setting in which a telephone call is made. We term these **environmental effects.** For example, phone calls may be made in the context of high degrees of background noise such as traffic. Recordings of telephone interactions will of course include sounds resulting from any background activity, which may in turn pose problems for forensic analysis of speech since the background noise may obscure some of the crucial information in the speech signal (see Hirson and Howard 1994).

Second, effects may result from speakers modifying their behaviour as a result of speaking into a telephone. We call these **speaker effects**. Behavioural differences may be conscious or subconscious depending on the individual and/or the interactional setting. A relatively extreme effect for some speakers may be the adoption of a different register or 'telephone voice'. The special register may result in changes to voice quality, speaking rate, and/or the use of different segmental pronunciations (typically, at least for English speakers, being closer to the perceived standard model; Wells 1982: 28). Forensic comparison with reference samples may therefore be made more difficult, since key phonetic variables may be radically different across the samples. Similarly, it has been shown experimentally that many individuals speak more loudly when using a telephone, probably as a subconscious reflex to circumvent environmental factors such as background noise (Summers *et al.* 1988, Summers *et al.* 1989, French 1998). One consequence of increasing loudness is an upward shift in a speaker's fundamental frequency (F0). This effect shows considerable cross-speaker variation, however. The degree of upward shift varies across individuals, and some speakers do not manifest the effect at all. Hirson, French and Howard (1995), for instance, investigated differences in F0 over the telephone and in face-to-face interaction. Thirteen of their twenty participants had a higher F0 in the telephone context, with an average increase of 4.33 Hz.

Finally we can also identify **technical effects** of telephone recordings, referring to those features which result from the act of transmitting speech through a handset and telephone line. Perhaps the most striking technical effect is the selective transmission of frequencies. Energy below *c.* 300 Hz and above *c.* 3,400 Hz is either attenuated or removed altogether (Künzel 2001, Rose 2003). The loss of high frequency components is particularly destructive for forensic speaker identification, since a great deal of potentially useful speaker-specific information (particularly voice quality information) is encoded in higher vowel formants (Nolan 1983, Künzel 1995).

The effect of attenuating low frequency energy has recently been investigated by Künzel (2001), specifically with reference to calculation of vowel formant values by acoustic analysis software programs. Calculations of the centre frequency of a formant are based on an averaging process over several neighbouring harmonics. However, the relative weighting of harmonics in these calculations decreases if they are attenuated, which is the case in phone transmission. Harmonics which fall outside the frequency ranges of the telephone filter function will be most affected. As a result, the relative weighting of those harmonics which fall inside the range of the filter function is increased. Where low formants are concerned (specifically below *c.* 500 Hz; Künzel 2001: 82), estimates of formant centre frequencies will be artificially high as a result of some lower frequency energy being damped or lost (the effect is illustrated graphically in Künzel 2001: 83 Fig. 2).

Künzel (2001) measured centre frequencies of vowel formants in recordings that had been simultaneously recorded in face-to-face context and via a phone line. For all twenty of his participants there were significant differences in F1 values measured across the two recording contexts, with F1 always higher after being recorded through the telephone. Similar results were obtained by Rose and Simmons (1996). Künzel's experiment further revealed that the average magnitude of F1 raising was very similar for both males and females (a point discussed by Nolan 2002). As predicted, low frequency formants were most affected, and hence close vowels (characterized by low F1 values) showed a greater F1 raising effect than more open vowels. The largest differences were in the region of 14 per cent for close vowels, diminishing to less than 1 per cent for some open vowels (Künzel 2001: 88 Table 2). By contrast, measurements for F2 were very closely matched across recording contexts, with only one speaker showing a statistically significant difference, and no individual vowel category varying by more than 2.2 per cent (Künzel 2001: 89). The forensic implications of Künzel's study are that we should expect F1 to be higher in telephone recordings than in direct recordings from the same speaker, while F2 can be assumed to remain relatively immune to the technical effects of the telephone (Künzel 2001, 2002, Nolan 2002).

Although there has been plenty of research carried out on telephone effects, the bulk of this work has concentrated on the characteristics of traditional landline telephone systems. Until recently little work has been performed to test the effects of mobile (cell) phones. This issue is of importance for many language-oriented fields including, for example, automatic speaker recognition (Sanders, Van den Heuvel and Choukri 1999). In particular there is a need to assess the ways in which mobile phone effects differ from landline effects. Environmental, speaker and technical effects are all subject to variability in respect of the type of telephone used. First, mobiles are subject to a wider range of environmental influences than landlines. The simple fact that mobile phones can be used almost anywhere means that many different types of background noise will be encountered with recordings made from mobiles. Second, aspects of speaker behaviour may differ when mobiles are used. It is clear from casual observation that mobile users have a tendency to speak particularly loudly, probably more so than is the case with landline users. The higher F0 effect of phone speech (noted above; Hirson *et al.* 1995) may therefore be exacerbated when mobiles are being used. McClelland (2000), for example, noted that F0 in mobile calls can be as much as 30 Hz higher than F0 in landline calls made by the same speaker. Third, the technical effects of mobiles differ from those of landlines. Recordings from mobile calls are often affected by GSM radio transmissions. These introduce an interference signal characterized by a fundamental frequency of 217 Hz, plus higher frequency harmonics overlapping the frequency range of speech (Harrison 2001; Grigoras 2003). It has furthermore been reported that the upper frequency cut-off for GSM transmission is lower than that for landline transmission at 3,200 Hz (Künzel 2001: 96). Sanders *et al.* (1999) also suggest that there is more variability in transmission quality across mobile networks, with some but not all providers using EFR ('enhanced full rate', which improves overall quality). A further indication of variability across systems is given by Rose, Osanai and Kinoshita (2003: 185), who found that data recorded via landlines in Japan had an upper cut-off frequency of around 4,500 Hz.

The importance of mobile phones for automatic speech recognition has led to a recent upsurge in research in this area (see Sanders *et al.* 1999 and references therein). Understanding the effects of mobiles on the speech signal is also, however, of urgent importance for forensic phonetics. The use of mobile phones is increasing at an exponential rate and it is expected that in a few years' time the number of mobile phone users will be higher than that of landline users. A 2003 study carried out on behalf of the UK government, for example, revealed that 75 per cent of adults owned or used a mobile, with that figure rising to 88 per cent for 15–34 year olds.[2] An inevitable consequence of this increase in mobile usage is an equally rapid increase in the amount of crime committed via mobiles. Forensic phonetic casework is therefore increasingly likely to involve recordings derived from mobiles.

EXPERIMENT

In the remainder of this article we describe an experiment carried out to test the effect of mobile phone transmission on vowel formant measurements. For the purposes of comparison the experimental design is based closely on that used by Künzel (2001) in his study of landline effects.

SUBJECTS

Künzel (2001) used 20 German subjects with an age range of 20 to 59. In our study 12 volunteers (six male and six female) were recruited. The males were aged between 20 and 34 (average 23) and the females ranged from 20 to 39 years (average 24). All were university-educated native speakers of English, and no speaker exhibited any speech or voice problems. The participants were all previously acquainted with one or other of the authors, which was helpful in eliciting relatively natural speech. The participants had a variety of accents, including RP and regional accents from the south, midlands and north of England. None of the regional accents was particularly strong, however.

MATERIALS

In Künzel's (2001) study participants read a German version of the text *The North Wind and the Sun*. In our study the subjects were asked to read a version of the text *The Story of Arthur the Rat* (see Appendix). This text was chosen as it is approximately twice as long as *The North Wind and the Sun*, and it includes most of the segmental phonological units of English (Abercrombie 1964: 38). However, the direct speech included in many published versions of the text was converted into narrative. We did this in light of the findings of Hirson *et al.* (1995: 236) that many speakers engaged in a dramatic reading style for the direct speech sections, which was therefore quite unlike their natural speech register. Participants were instructed to read the text as naturally as possible, and readings lasted on average about one minute.

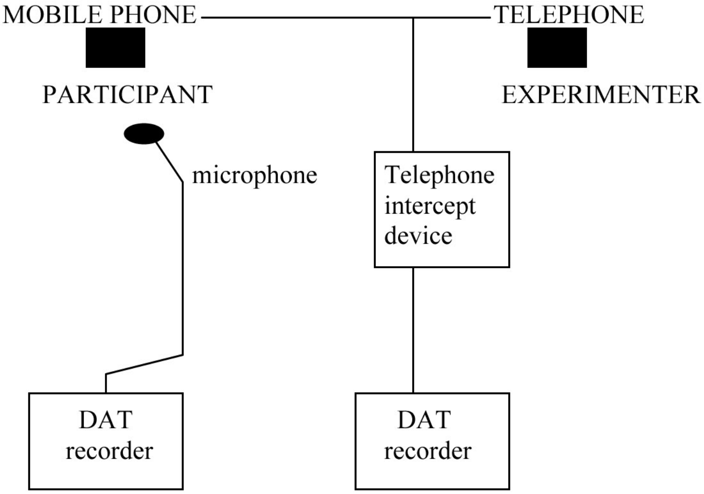

Figure 1 Schematic view of simultaneous recording environments

The mobile phone calls were made on a Nokia 3330 mobile phone operating on the BT Cellnet (since rebranded as O2) network. The participant was asked to place a call to the experimenter (the first author), who was located in another room. Two recordings of each speaker were then made simultaneously, as illustrated in Figure 1. The face-to-face recordings were made via an AKG microphone situated directly in front of the speaker, connected via a Rane pre-amplifier to a Sony digital audio tape (DAT) recorder. A second DAT recorder was situated in the experimenter's room, and was connected to a purpose-built intercept device to record the incoming phone call. Careful consideration was given to the position of the recording equipment during the experiment in order to reduce problems with GSM interference (Harrison 2001: 9).

ANALYSIS

The recorded data were resampled onto computer for acoustic analysis. The sampling rate was set to 22,050 Hz following Künzel (2001). Acoustic analysis was carried out using Sensimetrics *SpeechStation2* software.

Acoustic measurements were made of 11 vowel categories, as listed in Table 1. The vowels are categorized with reference to the lexical set notation used by Wells (1982). This system circumvents the need to use IPA symbols ambiguously to define phonological units as well as phonetic qualities. Table 1 also provides examples of the words analysed for each category. Note that some lexical items may be allotted to different categories depending on the speaker's accent. Northern varieties, for example, have TRAP rather than START (that is, [a] not [ɑː]) in *answer*, and no distinction between STRUT and FOOT words, both of which typically have [ʊ].

Following the usual practice in formant analysis measurements were taken at relatively stable points close to the centre of each vowel. Formant values were extracted from LPC spectra using the narrowest bandwidth setting available in the software. Between 15 and 25 vowels were analysed

Table 1 Vowel categories analysed

Keyword	typical quality	examples
FLEECE	iː	chief, three, least
KIT	ɪ	his, lived, given
DRESS	ɛ	never, friends, Helen
SQUARE	ɛː	there, care, hair
TRAP	a	carried, rat, (answer)
START	ɑː	Arthur, aunt, (answer)
NURSE	əː	learn, heard, search
LOT	ɒ	loft, rotten, horror
STRUT	ʌ	trouble, young, just
FOOT	ʊ	looked, good, shook
GOOSE	ʉː	who, do, room

per speaker across the two recording conditions, yielding a total of 250 measured tokens. As was coincidentally the case in Künzel's study, the number of tokens per vowel per condition per speaker ranged from one to seven.

Several problem cases occurred in the analysis, as is typical in acoustic work. For example, it was on occasion difficult to measure a formant with confidence because its amplitude was too low. In some cases (usually the most open or close back vowels) F1 and F2 were too close together to be distinguished from each other. In other tokens the target vowels were affected by coarticulatory or assimilatory effects such as elision or devoicing. In the case of F3 in the mobile recordings it was sometimes impossible to locate a formant peak at all, presumably because the high frequency cut-off of the telephone transmission had removed the relevant acoustic information. Problematic formants were not measured.

RESULTS

F1 frequencies

Table 2 summarizes the F1 measurements obtained for male and female speakers respectively. The data presented are the mean values for eleven vowel categories, averaged across the six male and six female speakers.

The first data column represents mean F1 values in the direct recordings, while the second shows the corresponding values from the mobile phone recordings. The third column shows the proportional difference between the previous two columns. A value of 100 per cent means that the formant frequencies were the same across the two conditions, and thus that telephone filtering has had no effect on the measured value. Figures larger than 100 per cent indicate that the mobile measurement is greater than that obtained in the direct condition. Statistical values were obtained via two-tailed Wilcoxon matched-pairs signed rank tests, following Künzel (2001).

The results in Table 2 demonstrate very clearly that the F1 increases in the mobile recordings. On average the mobile values are 29 per cent higher than the direct values for both males and females. This amounts to an average increase of 149 Hz for females and 141 Hz for males. The largest increase, 60 per cent, occurred in the females' LOT category. On average the tokens in this category showed a rise of 282 Hz. Most of the statistical comparisons proved to be highly significant. The four non-significant cases almost certainly result from very small numbers of tokens in the category.

Table 2 furthermore suggests that closer vowels (that is, those with lower F1 values) tend to be more markedly affected by the mobile transmission. This is clearest on examination of the male data, where FLEECE and GOOSE show relatively large increases (43 per cent and 40 per cent)

Table 2 Average F1 frequencies in direct and mobile phone recordings

Males	**F1 direct**	**F1 mobile**	**% difference**	**s.d.**	**Significance**
FLEECE	301	430	142.9	37.8	***
KIT	360	490	136.0	25.0	***
DRESS	518	639	123.5	28.7	***
SQUARE	520	617	118.7	8.6	***
TRAP	614	778	126.7	19.0	***
START	608	763	125.5	24.1	***
NURSE	448	621	138.6	22.6	**
LOT	470	609	129.4	24.3	ns
STRUT	460	678	147.3	33.4	*
FOOT	443	547	123.6	37.2	ns
GOOSE	355	499	140.4	59.4	*
all vowels	483	624	129.2	29.4	
Females	**F1 direct**	**F1 mobile**	**% difference**	**s.d.**	**Significance**
FLEECE	333	453	136.0	29.7	***
KIT	414	513	124.1	19.5	**
DRESS	591	781	132.1	29.1	***
SQUARE	586	746	127.2	19.4	**
TRAP	641	812	126.6	28.8	***
START	647	807	124.7	34.9	**
NURSE	446	678	152.1	35.9	*
LOT	467	748	160.3	39.2	**
STRUT	504	630	125.1	47.6	ns
FOOT	500	536	107.4	9.9	ns
GOOSE	448	514	114.7	21.4	**
all vowels	514	664	129.0	30.7	

Significance levels: *** < .001, ** < .01, * < .05, ns = not significant

compared with the most open vowels TRAP and START (27 per cent and 26 per cent). There was, however, substantial variability across tokens. The relatively high standard deviations for most of the vowel categories bear witness to this variability, which is also illustrated in Figure 2. This scatter graph represents all F1 measures in our sample. For each token the value in the direct recording is plotted against the x-axis. The y-axis represents the shift in Hz found in the corresponding mobile token (that is, the value of F1 in the direct condition subtracted from its counterpart in the mobile condition). The great majority of tokens (97 per cent) fall on or above the zero line on the y-axis, indicating that the mobile value was equal to or higher than the direct value. The variability within the sample is evident from the wide spread of data points, particularly around the centre of the graph. There is, however, some evidence in Figure 3 to support the observation that close vowels tend to be more affected by the telephone filtering than more open vowels. In general, the higher the F1 in the direct condition, the smaller the upward shift in the mobile condition (hence the low y values towards the right of the figure, and the negative trend line). The overall correlation between direct measurements and degree of upward shift is highly significant ($r = -0.304$, df = 248, $p < .001$).

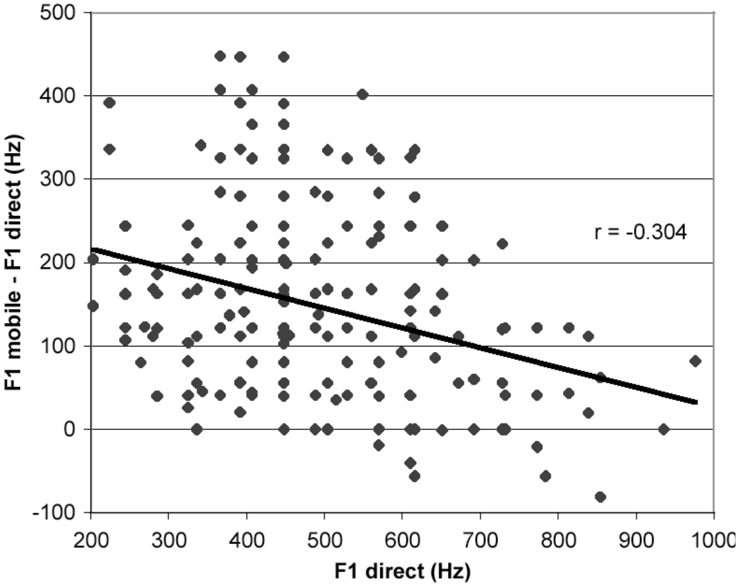

Figure 2 Scatter plot of F1 measures, showing value in direct recording (x axis) against shift observed in mobile recording (y axis). Values greater than 0 on the y axis indicate a higher value in the mobile condition.

Individual speakers also differed in the extent to which their F1 values rose in the mobile condition. Figure 3 shows the comparison of average F1 measures across the two conditions for each participant. Data from direct recordings are shown as black bars and the corresponding mobile data are shown as grey bars. The six males are shown to the left (m1, m2 etc) and the females to the right (f1, f2 etc). The number above each pair of bars represents the ratio between the two sets of data; thus for m1 the mobile values were on average 122.7 per cent of the direct values.

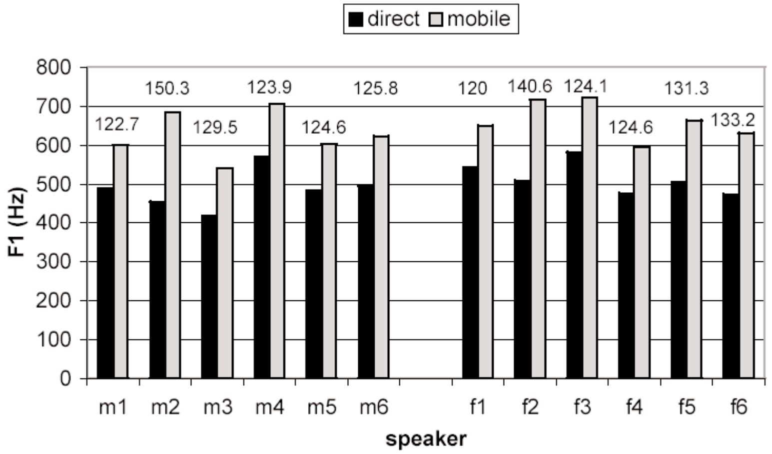

Figure 3 Comparison of average F1 values in direct and mobile conditions for individual speakers (m=male, f=female; the number above each pair of bars indicates the ratio between the data from the two recording conditions).

The majority of the speakers showed average rises of between 20 and 30 per cent. Two speakers, however, showed considerably higher rises, namely m2 (over 50 per cent) and f2 (40 per cent).

F2 frequencies

The results of F2 analysis are summarized in Table 3. In contrast with the F1 results, the F2 data show considerably closer correspondence between measurements taken across the two recording conditions. The proportional differences are generally much closer to the 100 per cent mark, with the largest deviation being the 90.8 per cent value for the KIT vowel by male speakers. The mobile results are in the main slightly lower than those in the direct condition, with average differences for both males and females of around 97 per cent.

Although the differences are much smaller than for F1, there are nonetheless some significant results identified by the statistical tests. In all but one case (the males' START) these significant results show lower values in the mobile condition. It is noteworthy also that the two categories which yielded significant differences for both males and females were those with the highest F2 values in the direct condition, namely the close front vowels FLEECE and KIT.

As with F1 measures, the mobile phone effect showed considerable variability across tokens. However, the variability in F2 measures derives largely from variation in the *direction* of the effect. Figure 4 represents the F2 data as a scatter plot with the same design as Figure 2.

Table 3 Average F2 frequencies in direct and mobile phone recordings

Males	**F2 direct**	**F2 mobile**	**% difference**	**s.d.**	**significance**
FLEECE	2417	2290	94.7	8.3	*
KIT	2129	1932	90.8	11.9	*
DRESS	1711	1647	96.3	9.2	ns
SQUARE	1773	1745	98.5	6.2	ns
TRAP	1502	1476	98.3	10.4	ns
START	1125	1204	107.0	10.8	*
NURSE	1544	1516	98.2	7.2	ns
LOT	1031	990	96.0	12.2	ns
STRUT	1334	1439	107.9	11.7	ns
FOOT	1466	1462	99.7	7.1	ns
GOOSE	1550	1637	105.6	9.9	ns
all vowels	1659	1623	97.8	10.5	

Females	**F3 direct**	**F3 mobile**	**% difference**	**s.d.**	**significance**
FLEECE	2622	2480	94.6	5.0	***
KIT	2277	2153	94.5	3.8	*
DRESS	1915	1857	96.9	9.4	ns
SQUARE	1832	1877	102.4	9.2	ns
TRAP	1727	1636	94.8	7.2	**
START	1410	1396	99.0	5.9	ns
NURSE	1702	1735	102.0	16.8	ns
LOT	1191	1288	108.1	19.6	ns
STRUT	1791	1674	93.5	0.6	ns
FOOT	1628	1588	97.5	9.3	ns
GOOSE	2105	2073	98.5	7.0	ns
all vowels	1898	1848	97.3	9.9	

Significance levels: *** < .001, ** < .01, * < .05, ns = not significant

The F2 data show a stronger correlation across the frequency range than was found for F1 ($r = -0.582$, df = 248, $p < .001$). We can see, though, that for F2 the effect of mobile transmission on formant values is complex: relatively low F2 values tend to be subject to an *upward* shift after filtering by the mobile (data points to the left), while high F2 values are usually recorded with *lower* values in the mobile context (data points to the right). In all, 35 per cent of tokens registered higher values in the mobile condition, while 57 per cent of tokens produced lower values.

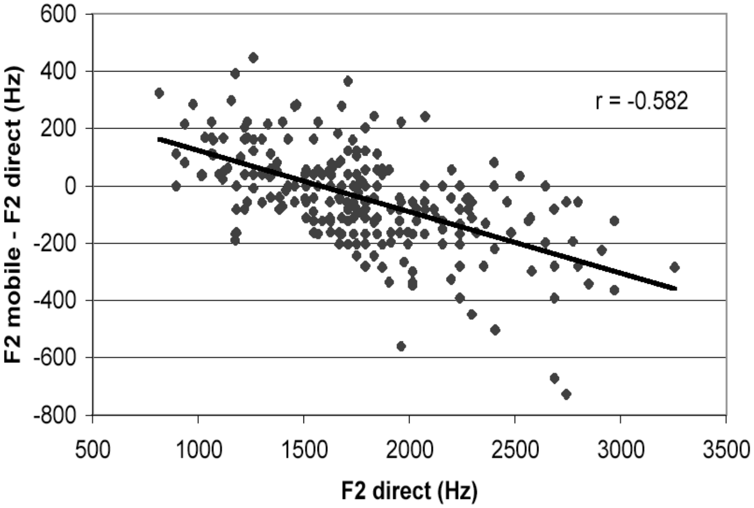

Figure 4 Scatter plot of F2 measures, showing value in direct recording (x axis) against shift observed in mobile recording (y axis). Values greater than 0 on the y axis indicate a higher value in the mobile condition.

F3 frequencies

The results for F3 measurements (Table 4 and Figure 5) are largely similar to those for F2. Across all tokens we found a small reduction in formant values in the mobile context (the average mobile:direct ratios were 96.6 per cent for males and 98.1 per cent for females). Significant differences were found only for three vowel categories – FLEECE for both sexes and KIT for males. It is notable that these three categories, all close front vowels, were characterized by particularly high F3 values relative to other categories. Measurements for these vowels (over 3 kHz) were therefore closest to the upper cut-off of the phone filter function. The overall correlation between direct measurements and the effect of filtering by the mobile phone was stronger than for either F1 or F2 (r = -0.713, df = 211, p < .001), supporting the observation that higher F3 values are subject to the greatest reduction in the mobile condition (Figure 5). This finding is particularly clear for those tokens with direct F3 values greater than 3,500 Hz (to the far right in Figure 5), for which the reduction in the mobile condition was particularly great (between 600 and 1200 Hz). Indeed, these values are so far removed from the corresponding F3 measures in the direct condition that they should probably be seen as errors on the part of the acoustic analysis program.

Table 4 Average F3 frequencies in direct and mobile phone recordings

Males	**F3 direct**	**F3 mobile**	**% difference**	**s.d.**	**significance**
FLEECE	3473	3089	88.9	8.6	**
KIT	3043	2853	93.8	6.7	**
DRESS	2836	2775	97.9	11.5	ns
SQUARE	2773	2718	98.0	10.0	ns
TRAP	2713	2686	99.0	16.1	ns
START	2888	2816	97.5	6.7	ns
NURSE	2713	2706	99.8	6.5	ns
LOT	3166	2947	93.1	12.6	ns
STRUT	2735	2690	98.3	4.5	ns
FOOT	2793	2796	100.1	5.4	ns
GOOSE	2727	2706	99.2	5.1	ns
all vowels	2900	2801	96.6	10.1	

Females	**F3 direct**	**F3 mobile**	**% difference**	**s.d.**	**significance**
FLEECE	3393	2970	87.5	10.9	***
KIT	2880	2906	100.9	3.8	ns
DRESS	2996	3030	101.1	5.3	ns
SQUARE	2767	2921	105.6	6.2	ns
TRAP	2808	2755	98.1	5.6	ns
START	3089	3073	99.5	7.7	ns
NURSE	2817	2935	104.2	2.8	ns
LOT	3043	2971	97.6	6.3	ns
STRUT	3022	2985	98.8	10.0	ns
FOOT	2815	2877	102.2	6.9	ns
GOOSE	2813	2737	97.3	8.7	ns
all vowels	2976	2918	98.1	8.4	

Significance levels: *** < .001, ** < .01, * < .05, ns = not significant

There was again a high degree of variability across tokens. Although the general trend was for the mobile condition to yield lower F3 values, the lowering was only found for just over half (53 per cent) of the individual tokens. A substantial number of tokens (39 per cent) registered a higher value in the mobile condition.

F3 is known to be fairly stable for individual speakers, and has thus often been regarded as a good speaker-specific diagnostic (for example Nolan 1990). In light of this we also analysed our F3 data with respect to the individual speakers in our sample. These results are shown in Figure 6. F3 values from the direct recordings, averaged across all tokens, are shown as black bars. The corresponding mobile data are shown as grey bars. The number above each pair of bars represents the ratio between the two sets of data; thus for m1 the mobile

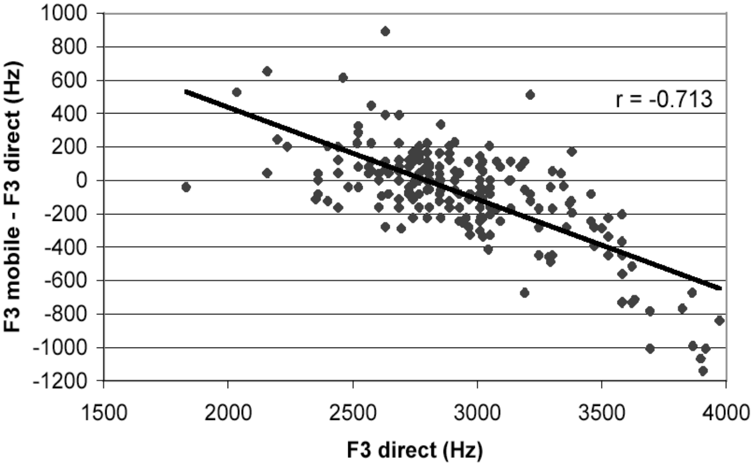

Figure 5 Scatter plot of F3 measures, showing value in direct recording (x axis) against shift observed in mobile recording (y axis). Values greater than 0 on the y axis indicate a higher value in the mobile condition.

values were on average 92.3 per cent of the direct values. The data in Figure 6 confirm our previous observation that higher F3 values are more affected by transmission through the telephone. The speakers with the highest average F3s (m1, m6, f5) are also those who show the greatest discrepancy in measurements across the two recording conditions. The 3 kHz mark appears to be a convenient dividing line: values above it are reduced by around 5 per cent or more in the mobile condition, while values below 3 kHz remain more stable across the two conditions.

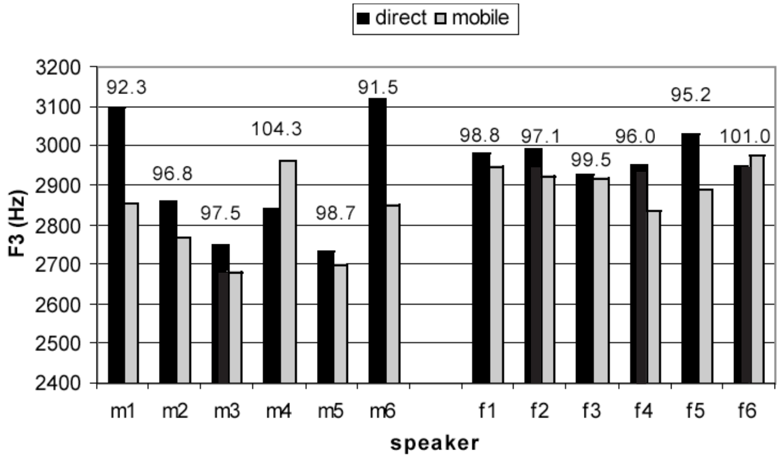

Figure 6 Comparison of average F3 values in direct and mobile conditions for individual speakers. (m=male, f=female; the number above each pair of bars indicates the ratio between the data from the two recording conditions.)

DISCUSSION

Overall our findings are compatible with those described by Künzel (2001), although there are also notable differences between the results of the two experiments.

F1 results in our study are generally subject to an upward shift after telephone transmission, and thus mirror Künzel's results. However, the differences across recording conditions are much greater in our data. The largest increase found by Künzel was 14 per cent (for [iː] in female speech), whereas in our mobile data the average increase was 29 per cent and specific vowel category increases reached as high as 60 per cent.

We suggest two possible reasons why the mobile data show a greater upward shift than Künzel's landline data. First, the specific filtering effect may be different in our study. The shape of the filter function and the precise location of the lower frequency cut-off may vary greatly depending on factors such as the type of telephone used and the method of transmission (Rose 2003). Second, when we compare mobile and landline telephones there are likely to be differences in the physical distance between the mouth and the telephone receiver. Mobiles are substantially smaller than landline handsets, and thus the distance from the mouth to the receiver is typically greater. The effect of this added distance may be to reduce even further the amplitude of lower frequency acoustic energy captured by the receiver.[3] The variation we saw in Figure 3, in which some speakers showed particularly large increases in F1 in the mobile condition, might also be explained if those individuals held the mobile phone in a markedly different position than the other speakers did. These suggestions, however, all require empirical testing.

Künzel (2001) furthermore found that F1 values of close vowels were the most markedly affected by phone transmission. This was predicted, since F1 values for close vowels are low, and thus more of their component harmonics are attenuated or removed by the low frequency cut-off of the telephone filter. Our data showed similar results, with lower F1 values being subjected to greater upward shift in the mobile condition (Figure 2). However, we also found a high degree of variability across tokens in terms of how large was the upward shift. A possible explanation for the variability can again be derived from the relationship between mouth and receiver: mobile users may exercise considerably more flexibility in both the distance between mouth and receiver, and also the angle between them. Variation in distance and/or angle may result in variability in the degree of attenuation of lower frequency energy. Once again, though, this suggestion requires empirical testing.

Our F2 data are also compatible with those reported by Künzel (2001). F2 measures are in general far less affected by telephone transmission, a finding which is predictable from the fact that the essential harmonic components of F2 measures fall well within the frequency range transmitted by telephone lines. We did, however, find that the majority of F2 measures were lower in the mobile recordings, with significant reductions found for a handful of vowel categories. The latter were mainly those with particularly high F2 values, that is, close front vowels.

F3 measurements were not taken by Künzel. Our results for F3 were comparable with those for F2, with generally a small downward shift in mobile recordings and significant shifts only for the highest values. This finding is again predictable from what is known of the filtering effects of telephone lines: high F3s are likely to be more greatly reduced because more of their harmonic components will fall outside the transmission range of the telephone.

Our results have clear implications for forensic speaker identification. First, we echo Künzel's warning that analysts should 'beware of the telephone effect'. Formant values, particularly those for F1, are subject to significant modification via the technical effects of the telephone. Indeed, our results show that the mobile phone effect on F1 is considerably greater than that found for landlines: average differences in the region of 30 per cent can be expected between recording conditions. Comparison of formants in telephone speech with formants in non-telephone speech is therefore problematic. As with Künzel's data, no algorithm can be proposed to normalize across recording conditions and undo the telephone effect, since there is a great deal of variability across tokens and across speakers. It is clear, though, that a situation in which phone speech F1s are consistently *lower* than those in a directly recorded reference sample is likely to be rare, and would thus provide strong evidence for the incompatibility of samples (compare with Nolan 2002).

F2 measures are more consistent across recording conditions, as noted by Künzel (2001) and Nolan (2002). However, our data do suggest a degree of caution should be exercised in comparing measurements across recording conditions. This is particularly the case with close front vowels, where the relatively high F2 values may be subject to significant lowering in mobile recordings.

F3 data are similarly more directly comparable across recording conditions than F1 data, but significant differences were found in our data for values at the higher end of the frequency scale. Since F3 remains reasonably stable for individual speakers, the interpretation of F3 measurements is therefore likely to vary across individual forensic cases. If a suspect's directly recorded reference sample indicates a low average F3, comparison with a phone sample is reasonably straightforward: we can anticipate that the phone effect will be small and probably insignificant. On the other hand, if the suspect's F3 is high (in particular over 3,000 Hz) we must allow for a significant telephone effect. This most likely means lower values in phone recordings, with the largest drop in our data being 8.5 per cent for one speaker (m6 in Figure 6) and 12.5 per cent for one vowel category (FLEECE for the female participants, Table 4). If larger differences were to be found in case materials then there would be weak evidence that the samples were more likely to be drawn from different speakers. Higher F3 values in phone speech than direct speech would be possible but less likely than the reverse, and would therefore also constitute evidence for incompatibility of samples.

CONCLUSION

The experiment reported here shows that mobile phones cause considerable interference with the acoustic properties of speech signals. It further shows that the effects of mobile phones may differ quite markedly from those of landlines (Künzel 2001). We therefore echo Künzel's call for further research on the differences between landlines and mobiles, and on the specific effects mobiles have on the speech signal.

The effects we found are furthermore both complex and variable: formant measures in mobile recordings may be higher or lower than in direct recordings, different formants are affected to different degrees, and effects vary across tokens, vowel categories and individual speakers. We have highlighted the main effects in our data to alert forensic phoneticians to the most likely effects they will find in casework. However, the most important conclusion to be drawn from our study is that caution should be taken in interpreting formant measures when different recording conditions are involved. This is particularly important in cases where it is not known whether a landline or a mobile was used in the production of a criminal telephone recording.

There have recently been suggestions, mainly from non-linguists, that instrumental phonetic analysis is inherently more valuable and reliable for forensic purposes than data derived from impressionistic auditory phonetic analysis (for example Ormerod 2002). Our data, however, serve to remind us that blind reliance on instrumental data is potentially dangerous. Formant measurements, just like auditory data, may present empirical difficulties, opportunities for differing interpretations, and thus varying degrees of forensic value.

APPENDIX

The Story of Arthur the Rat

There was once a young rat named Arthur, who would never take the trouble to make up his mind. Whenever his friends asked him if he would like to go out with them, he would only answer that he didn't know. He wouldn't say 'Yes' and he wouldn't say 'No' either. He could never learn to make a choice.

His aunt Helen said that no-one would care for him if he carried on like this. He had no more mind than a blade of grass. Arthur looked wise, but said nothing.

One rainy day the rats heard a great noise in the loft where they lived. The pine rafters were all rotten, and at least one of the joists had given way and fallen to the ground. The walls shook and the rats' hair stood on end with fear and horror. The old rat who was the chief said that this wouldn't do and so he sent out scouts to search for a new home.

Three hours later, the seven scouts came back and said that they'd found a stone house which was just what they wanted. There was room and good food for them all. There was a kindly horse named Nelly, a cow, a calf and a garden with an elm tree. Just then the old rat caught sight of young Arthur and asked whether he was coming with them. Arthur replied that he didn't know as he thought that the roof might not come down. The old rat said angrily that they couldn't wait all day for him to make up his mind. And they went off.

Arthur stood and watched the other rats hurry away. The idea of an immediate decision was too much for him. He said to himself that he would go back to his hole for a bit, just to make up his mind.

That night there was a great crash that shook the earth, and down came the whole roof. Next day some workers rode up and looked at the ruins. One of them moved a board, and under it they saw a young rat lying on his side, quite dead, half in and half out of the hole.

NOTES

1 We would like to record our thanks to Peter French, Philip Harrison and Lindy Rescorle for their help in the execution of the experiment reported here. Our thanks also to two anonymous referees, Peter French, Philip Harrison and Phil Rose for their helpful advice on drafts of the article.

2 <http://www.statistics.gov.uk/STATBASE/ssdataset.asp?vlnk=7202> (accessed 24 February 2004).

3 We are indebted to Peter French for this suggestion.

REFERENCES

Abercrombie, D. (1964) *English Phonetic Texts*, London: Faber & Faber.

French, J. P. (1998) 'Mr Akbar's nearest ear versus the Lombard reflex: a case study for forensic phonetics', *Forensic Linguistics*, 5(2): 58–68.

Grigoras, C. (2003) 'Digital audio recordings analysis – the electronic network frequency', paper presented at the International Association for Forensic Phonetics Conference, Vienna.

Harrison, P. (2001) 'GSM interference cancellation for forensic audio: a report on work in progress', *Forensic Linguistics*, 8(2): 9–23.

Hirson, A., French, J. P. and Howard, D. (1995) 'Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics', in J. Windsor Lewis (ed.), *Studies in General and English Phonetics in Honour of Professor J.D. O'Connor*, London: Routledge, 230–40.

Hirson, A. and Howard, D. M. (1994) 'Spectrographic analysis of a cockpit voice recorder tape', *Forensic Linguistics*, 1(1): 59–69.

Künzel, H. J. (1995) 'Field procedures in forensic speaker recognition', in J. Windsor Lewis (ed.), *Studies in General and English Phonetics in Honour of Professor J.D. O'Connor*, 68–84.

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(2): 83–6.

McClelland, E. (2000) 'Familial similarity in voices', paper presented at the BAAP Colloquium, University of Glasgow.

Moye, L. S. (1979) *Study of the Effects on Speech Analysis of the Types of Degradation Occurring in Telephony*, Harlow: Standard Telecommunication Laboratories.

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

Nolan, F. J. (1990) 'The limitations of auditory-phonetic speaker identification', in H. Kniffka (ed.), *Texte zu Theorie und Praxis Forensischer Linguistik*, Tübingen: Max Niemeyer Verlag, 457–79.

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

Ormerod, D. (2002) 'Sounding out expert voice identification', *Criminal Law Review*, 771–90.

Rose, P. J. (2003) 'The technical comparison of forensic voice samples', in I. Freckelton and H. Selby (eds), *Expert Evidence*, Sydney: Thomson Lawbook Company, Ch. 99.

Rose, P. J., Osanai, T. and Kinoshita, Y. (2003) 'Strength of forensic speaker identification evidence: multispeaker formant- and cepstrum-based segmental discrimination with a Bayesian likelihood ratio as threshold', *The International Journal of Speech, Language and the Law*, 10(1): 179–202.

Rose, P. J. and Simmons, A. (1996) 'F-pattern variability in disguise and over the telephone – comparisons for forensic speaker identification', *Proceedings of the Speech Science and Technology 96 Conference*, Flinders University, Adelaide, 121–6.

Sanders, E., Van den Heuvel, H. and Choukri, K. (1999) 'Building speech databases for cellular networks', *Proceedings of the Sixth European Conference on Speech Communication and Technology (Eurospeech 99)*, Budapest. <http://www.isca-speech.org/archive/eurospeech_1999/e99_2275.html>

Summers, W. V., Johnson, K, Pisoni, D. B. and Bernacki, R. H. (1989) 'An addendum to "Effects of noise on speech production: acoustic and perceptual analyses"', *Journal of the Acoustical Society of America*, 86: 1717–21.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I. and Stokes, M. A. (1988) 'Effects of noise on speech production: acoustic and perceptual analyses', *Journal of the Acoustical Society of America*, 84: 917–28.

Wells, J. C. (1982) *Accents of English* (3 vols.), Cambridge: Cambridge University Press.