Evaluation of Disputed Utterance Evidence

in the matter of

David Bain's Retrial

prepared for

Joe Karam / Duncan Cotteril Lawyers

by

Dr. Phil ROSE

January 2009

Executive Summary

An acoustic and auditory-phonetic analysis of the data in Bain's emergency call shows strong evidence in support of the hypothesis that the disputed utterance begins with Bain saying *I can't*, and against the hypothesis that it begins with *I shot*. Neither is it convincing to argue that the disputed utterance was not speech.

1.0 Background

On Wednesday 14th January 2009 I was emailed by Mr. Joe Karam, who wanted to know if I could help him in an urgent forensic speech science task related to the matter of David Bain, who was being retried for the murder, in 1994, of his family. Karam, who represents Bain, explained that the Crown allege the telephone call Bain made to emergency services contains a short passage amounting to an admission of guilt. I did a detailed cold transcription of the call, adducing some acoustic evidence, which is contained in my report *Transcription of Emergency Recording in the matter of David Bain's Retrial*, tendered to Karam and Duncan Cotteril solicitors on January 20th.

After perusing my transcription and accompanying notes, Karam told me what the Crown allege to be the incriminating content of the disputed utterance. He also furnished me with the opinion of several other parties, who had been asked by both Crown and Defence to listen to and transcribe the call, as to the content of the disputed utterance. He solicited my expert opinion on various matters related to these different hypotheses as to what was actually said. This is my report. It assumes familiarity with the contents of my first report.

2.0 Structure of report

I have structured the argument of the report in the following way. The first section summarises the findings of the various investigators asked to determine what was said. I then argue that their findings cannot be usefully evaluated, and that, instead of treating what they say they heard as evidence, the actual acoustics of the call should be treated as the evidence, by asking what is the probability of getting these acoustics under the competing hypotheses as to what Bain actually said. To evaluate the acoustics, the problem needs to be couched, not in terms of words, but in terms of the speech sounds that make up the words. Therefore it is first necessary to explain how the relevant speech sounds are produced and described, and this is done in section 3. This section ends with a hypothesis as to what Bain said, given the auditory-phonetic properties of the call. In the final section I present an acoustic analysis of these sounds, making use of known data from the rest of the call, which supports the auditory hypotheses. I conclude that you would be very much more likely to get the acoustics if what Bain actually said began with *I can't* rather than *I shot*.

3.0 Investigators' responses

This is a fairly typical forensic speech science case of *disputed utterance*: where there are conflicting opinions as to what was actually said, because the recording of the speech, for whatever reason, is unclear (French 1990, Rose 2002: 2). What constitutes the disputed part of the call is not controversial.

3.1 Claims

There are several hypotheses as to what Bain said.

1. Detective Ward in his statement says (p. 2) that he listened to the emergency call and heard the words *I shot the prick, I shot*. He says he transcribed the call, and this transcription is included in a separate document, endorsed by Mr Dempsey, the St. John's ambulance officer who took the call. Detective Ward's transcript of the call does indeed contain these words, separated by a "(pant)". Detective Ward notes that he listened many times to the call.

2. Mr Pearce, a sound engineer whom detective Ward asked to listen to the call for "extraneous noises", says in his statement (p. 2) that he became aware of "semi whispered words", and on subsequent listening clearly heard the words *I shot the*

prick I shot. He notes he and detective Ward listened for about four to five hours to the recording.

3. Mr Dempsey, the St. John's Ambulance officer who took the emergency call, and was asked by detective Ward to verify that it was his voice in the recording, says in his statement (pp. 2,3) that he listened to the call, was then given a transcript by detective Ward, "did not hear all the words written in the transcript", listened again, and clearly heard the words *I shot the prick I shot*.

4. Dr Innes, an expert in Conversation Analysis, includes a cold transcription of the call in her letter to Reed QC of 18th March '07. At line 25 (the disputed portion) she transcribes *(?p(x)ee)*. In the CA notation she uses this represents (p.2) "sounds which are relatively clear but the transcriber cannot be certain of what the word is". The sounds indicated are presumably a voiceless bilabial stop /p/ ("p"), separated by some unclear material ("x") from a high front unrounded vowel /i:/ ("ee")).

5. Dr Guillemin in his expert witness report of August 8th 2008 states (p. 4) that he was asked, among other things, to transcribe the call. In particular he was asked to pay attention to the disputed portion, and give his opinion as to whether Bain had said anything at this point. Dr Guillemin locates the disputed portion at "approximately 29 seconds from the start", but it is actually bit later than this, about 33 seconds from the onset of the waveform. However, the main thing is that his transcription contains nothing corresponding to discernable speech between Bain saying *Yes* (my ref: U13, 14). and the operator saying *what phone number you're calling from*? (my ref: line 23). He confirms this in his conclusion (p.7):

> Approximately 29 seconds into the recording I hear the sounds of David Bain gasping for breath. If he was also trying to say something through his gasps, this is not discernable to any degree of certainty.

6. Professor French and Mr Harrison, professionals from a major UK forensic audio and speech laboratory, were told of the Crown's allegation that Bain said *I shot the prick*, and were then asked, among other things, to transcribe the call. They state (pp. 6,7) that

> … the material could be heard as 'I shot the prick' or 'I shot that prick'. However … it also remains entirely possible that it is not speech. Rather, it could be no more than an audible out-breath that has, in the distress of the moment, been modified by a random and unfortunately-sequenced series of movements of the tongue and lips so as to create a series of sounds that could – albeit with a little effort – be heard as 'I shot the/that prick'.

Their transcription does not contain 'I shot the prick'; they only indicate where the disputed utterance lies in the conversation. In his witness statement, Mr Harrison further agrees with the proposition (p.10) that the disputed utterance is equally likely to be *I shot the prick* as non-speech. They report that they carried out an experiment to show that it was possible to assemble random non-speech noises from the recording, such that one could hear words from them.

7. Miss Cauley, a member of Prof. French's staff, was asked to do a blind transcription of the call, and transcribed (p.11) … *I can't touch it* for the disputed

utterance. Later she was told of the interpretation *I shot the prick* and said that she could hear how someone could hear that (p.13).

7. Dr Foulkes, a forensic phonetician and linguist from the UK, produced a cold transcription of the emergency call, in which he transcribed the disputed portion as (I can't breathe…). He was then told the Crown's assertion and was asked to re-examine the disputed portion, and comment on "whether the questioned section could comprise the words *I shot the/that prick.* He was also asked to respond to the report prepared by French and Harrison. He states (p. 3) :

- that the questioned section "could be heard as 'I shot the/that prick' ";
- that *I can't breathe* "is also a possible interpretation"; and
- that he agrees fully with French and Harrison that

(1) it is possible that it is not speech at all; and that
(2) it is not possible to resolve the issue.

He also states (p.5) … there is no objective phonetic or acoustic justification to support the interpretation 'I shot the prick'".
He further states that he attempted to conduct a detailed acoustic analysis to determine whether there were any grounds for preferring *I shot the prick* over *I can't breathe*, but was unable to reach a clear conclusion. He does not say what his actual analysis involved other than pairwise comparisons of the putative segments, and gives no acoustic quantification.

8. Dr Rose (me) was asked by Karam to transcribe the call, and insisted on doing it cold (see my report). I produced a transcription in quasi CA notation, with relevant portions also transcribed phonetically. I heard the disputed portion as *I can't help puking*, but I also indicated that this was unclear (this was formally indicated by its being transcribed in brackets as per CA convention). I presented acoustic evidence to support this interpretation.

3.2 Comments

The summary above notes several claims as to what Bain said, which polarise into incriminating (*I shot the prick*) and anodyne (*I can't breathe, I can't help puking, I can't touch*, not speech.) There are several important comments to make about this.

The first thing is that it is not at all surprising that the claims differ. As I explained in my initial report, given the unclear nature of the signal, it is a natural perceptual response for the brain of the listener to invoke whatever top-down processing it can to arrive at a percept. Therefore what is heard is not necessarily actually all in the signal. Most importantly, it is well known that this top-down processing is driven partly by expectation: *you hear what you expect to hear*. Thus, for example, I do not find it at all surprising that detective Ward hears something incriminating, and I do not doubt that *I shot the prick* is what he heard.

It follows from the above that you need, firstly, to distinguish clearly between *what the disputed utterance can be heard as* and *what, if anything, was the intended linguistic message – what Bain actually said.* It is clear from a perusal of the reports that this distinction was not always made. For example, Dr Foulkes was asked "whether the questioned section could comprise the words *I shot the/that prick*. This question refers to whether *what was said* was *I shot the prick.* He replied, correctly, that it "could be heard as" 'I shot the/that prick' and this is a different thing. Note also

that French and Harrison's report correctly uses the words *could be heard as*. Again, this is not the same as saying that that is what Bain actually said. However, at one point under cross, Harrison's answer does amount to a statement about what was actually said.

Now, most of these claims have in common the logical structure of *hypotheses as to what was said, given the evidence of what was heard*. So for example detective Ward's hypothesis is that Bain said *I shot the prick*, given that that is what he heard. The *hypothesis* is what Bain actually said; the *evidence* is what someone reports hearing him say.

There are several competing hypotheses as to what Bain said, and the trier of fact has to decide which one, if any, is most likely to be true. Now we come to the crux of the matter. I think it needs to be said that **you will not get anywhere trying to evaluate these claims as they stand, that is, trying to judge the merits of the different claims that Bain said something, given that something was heard.** This is because in order to evaluate whether one hypothesis is more true than another, given some evidence adduced in its support, one of the things you have to know is the strength of that evidence[1]. Logically, the strength of the evidence is the probability of getting the evidence under the competing hypotheses: what is the probability of getting the evidence assuming one hypothesis is true, relative to the probability of getting the same evidence assuming that the alternative is true? So we would need to estimate, for example:

What is the probability that detective Ward heard Bain say *I shot the prick*, given that he did in fact say *I shot the prick*, and what is the probability he heard *I shot the prick*, given that Bain did not say that, but something else, like *I can't breathe*. If he was more likely to hear *I shot the prick* given that that was what Bain said, then that would constitute support for the hypotheses that Bain said *I shot the prick*. Another way of saying this is that we would want to know how reliable Ward is in recognising speech degraded to the same extent as this case. But we would also have to do the same estimation for everybody else involved: what is the probability that Dr Foulkes thinks he hears *I can't breathe*, given that Bain said that, relative to the probability of Dr Foulkes reporting *I can't breathe* given that Bain said something else, like *I shot the prick*.

Now, it is in principle possible to run an experiment to estimate these probabilities, but it should be clear that this would be an absolute nightmare. For a start, one would have to take into account the expectation effect: I think it would increase the probability that detective Ward heard *I shot the prick*, given that Bain said something else; and it might be argued that it would also do that for anyone who had a prior expectation as to what was said because they had been told it beforehand. This would almost certainly reduce the strength of evidence, because it would mean that you would move towards being just as likely to get the evidence under both hypotheses. In fact it is a possible, but totally undesirable outcome that everybody ends up evaluated in the same way.

---

[1] The other thing is the odds in favour of one hypothesis over the other before the evidence is adduced. These are called the prior odds. I have not mentioned priors in this report, because I think that they will favour the defence, and even if one assumes flat priors, the probability of the defence hypothesis will prevail.

I suppose one might try to circumvent all this by (1) discounting as a matter of course any perceptual response to the disputed utterance that has been primed. And the perceptual response of anyone who has expectations concerning the case. (2) Giving more weight to professionals like French and Foulkes (and me), who have spent a lifetime listening to and analysing speech, and are painfully aware of what they can and can't say (an expert is, among other things, someone who knows what they don't know).

But there is a much better way. As I said in my first report, you need clearly to distinguish between three things: (1) what Bain actually said; (2) what someone *heard* him say; and (3) the acoustics of the disputed utterance. It is the latter – the acoustics – which is best taken as the evidence; NOT the words that were heard. Using the acoustics of the call as evidence has the major advantage that it is the *same evidence for all hypotheses*. That is, we would have for example to evaluate what the probability is of getting the acoustics of the disputed portion assuming Bain said *I shot the prick*, and assuming he said *I can't breathe*. The second major advantage is that the acoustics are quantifiable. This makes it easier to compare questioned and known data. If we want to estimate the probability that the disputed portion represents Bain saying x , and we have a known example of Bain saying x, then we can make some sort of statement as to the probability of getting the disputed acoustics assuming x is true. I will in due course show how this can be done. To do this, however, I will need to refer to and talk about different speech sounds, not words, and so I need to describe a little about what speech sounds are like and how they are properly described. This will then enable me to say what I think is happening in the emergency call, and give me a platform for the subsequent acoustic testing.

4.0 Speech Sounds

We are concerned here with only a few sounds, so it will be useful for the non-specialist reader at this point if I say a little about their articulatory nature and how they are described. Because it is not generally known by laymen, I wrote at length about this in my 2002 & 2003 books *Forensic Speaker Identification* and *The Technical Comparison of Forensic Voice Samples*. I summarise below the content of chapters 6 & 7 in the former, as far as it is relevant to the case in hand.

4.1 Consonants

Consonantal sounds like the first sound in *can't* or *shot* are described by reference to the place in the mouth at which they are made, and the way (and extent) in which the vocal tract is obstructed to make them. Thus the sound at the beginning of the word *can't* is described as a velar plosive. *Velar* means that it is made at the soft palate, or velum, at the back of the mouth (the place of articulation). *Plosive* means that the sound is made by totally shutting off the flow of air through the mouth; letting the air pressure build up behind the closure; and then releasing the occlusion so the air pops out. In a velar plosive, the top of the tongue body shuts off the air when it contacts the soft palate. The velar plosive at the beginning of the word *can't* is transcribed [k]. The square brackets mean their contents refer to an actual speech sound.

The *sh* sound at the beginning of the word *shot* is called a palato-alveolar fricative (the first term denotes the *place*; the second term the manner). Palato-alveolar means it is made at the front of the mouth, from just behind the teeth ridge, to the hard palate in mid mouth). *Fricative* means that instead of shutting off the air totally, as in a plosive, the air is forced thru a narrow channel, giving rise to turbulence. In the case of a palato-alveolar fricative, the narrow channel is formed between the crown of the tongue (the tip and the blade), and the roof of the mouth from just behind the alveolar ridge to the palate. The palato-alveolar fricative at the beginning of the word *shot* is transcribed [ʃ].

To make things easier to envisage, I reproduce in figure 4.1 parts of figures from Laver's (1994) phonetics textbook which shows tracings from mid-sagittal x-rays of the mouth and tongue during the articulation of a velar plosive and a palato-alveolar fricative. You can see the tongue contacting the back part of the roof of the mouth in the velar plosive, and the front part of the tongue forming a long narrow channel from just behind the alveolar ridge to the front of the hard palate in the palato-alveolar fricative.

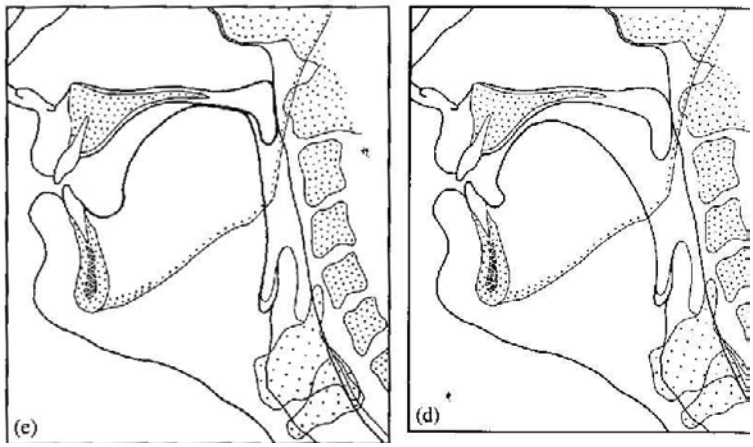

Figure 4.1 Tongue position in a velar stop (left) and a palato-alveolar fricative. After Laver (1994 p.377, 246). Reproduced with permission.

4.2 Vowels

Basically, vowel sounds like the *a* in *can't* or the *o* in *shot* are described by referring to the location in the mouth of the tongue body - whether it is high, mid or low; or front, central or back. It is also necessary to say what the lips are doing – whether they are rounded or not. So the vowel in the NZ word *can't* is described as being *low central unrounded*, meaning the tongue body is low in the mouth, neither in the front or the back, and the lips are not rounded. It is transcribed [a]. The vowel in the word *shot* is low back and rounded, and transcribed [ɒ].

4.3 Phonemes and allophones

In Language, speech sounds exist on two different levels. At one level there are basic, target sounds called *phonemes*. These are the sounds that a speaker of a given language aims to make when they say a word. They are usually transcribed in oblique slashes. So for example, the NZ English word *can't* is made up of four phonemes: /k/ /a:/ /n/ /t/. The first is a *velar plosive* phoneme. Change the phoneme and you are

liable to create a different word: for example substitute the /k/ by /ʃ/ - a palato-alveolar fricative - and you get *shan't* /ʃa:nt/.

Speech scientists distinguish the phoneme, which is an abstract contrastive unit of sound, from its realisation - the actual sound produced - which is called an *allophone*. An allophone is written in square brackets, and the relationship between phoneme and allophone is symbolised as follows: /x/ → [y]. This is to be read as *the phoneme x is realised by the allophone y*. For example, we write, using now the correct symbols, /k/ → [k]: "The (NZ English!) phoneme k is realised as the allophone k."

This may look redundant ("k is realised as k?!") but it is not, because the sounds exist on two different levels of structure: one of the ks is a phoneme, and one is its realisation. This becomes clearer when it is seen that the same phoneme can be realised by more than one speech sound (i.e. it can have more than one allophone). In this case it is necessary to stipulate the environment in which a particular allophone occurs, thus:

/x/ → [y] / a
→ [z] / b

This is read as *the phoneme x is realised by the allophone y in the environment a, and by the allophone z in the environment b*. Quite often the particular realisations of a phoneme can be seen to occur because of the influence of surrounding sounds. The speaker aims to make a particular speech sound, but his tongue and his lips are perturbed away from this target by sounds that have come before, or in anticipation of sounds that are yet to be made. A very simple example of a single phoneme being realised by two allophones in this way can be found with the /velar plosive/ in English.

Try this experiment. Say *car* several times and then try to isolate the first sound you said in it: *car car car car c c c c*. This is one allophone – [k] – of the velar plosive phoneme /k/. Now say *key* several times and try to isolate the sound you are making at the beginning of this word. You should be able to hear that you are making a very different sound at the beginning of the word *key* than the word *car*: two different ks if you will. You can hear the k in *key* has a higher pitch than the k in *car*. The one at the beginning of the word *key* is in fact made further forward in the mouth than the one at the beginning of the word *car*, that is why it has a higher pitch. The k sound in NZ *car* is a velar plosive [k], the k sound in *key* is called a fronted velar, transcribed as [k+] or sometimes [c]. Figure 4.2, again from Laver's (1994) mid-sagittal x-ray tracings, shows the tongue position in a plain and fronted velar plosive. You can seen how the tongue body moves forward quite a lot from the plain velar position on the left to the fronted velar position on the right. This results, in the fronted velar, in a shorter front cavity in the mouth, the air in which resonates at a higher frequency, and a longer back cavity, which has a lower resonant frequency.

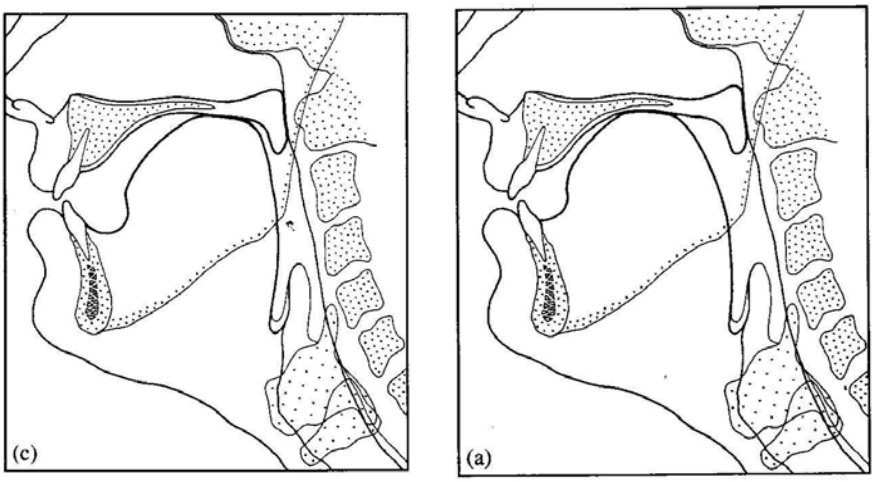

Figure 4.2. Tongue position in a plain velar plosive (left) and a fronted velar. After Laver (1994 p.377). Reproduced with permission.

Why do you get a fronted velar in the word *key*, and a plain velar in the word *car*? Because the body of the tongue in the *vowel* in the word *key* lies to the front of the mouth – it is a *front* vowel - and the body of the tongue in the vowel in the word *car* lies in the centre – it is a central vowel. When you say *key*, the tongue body aims for a central position to make a velar closure for the consonant, but is pulled forwards in anticipation of the fronter position of the vowel. A fronted velar is the result. Note that there is no change in phoneme – it is still a /velar plosive/, or /k/, but its realisation is different. This situation we can symbolise as at 4.1:

/ k /
→ [k+] / before a front vowel (like in *key*) (4.1)
→ [k] / before a central vowel (like in *car*)

This sort of thing – where one sound becomes more like another adjacent sound is called assimilation, and is quite common in languages.

A phoneme may have more than two allophones. For example, the voiceless velar plosive /k/ phoneme in most varieties of English actually has many more than the two just illustrated. This is where the relevance to Bain comes in.

5.0 Auditory Phonetic Analysis

Table 5.1 represents the speech sounds of the different claims phonemically.

Table 5.1 Different claims phonemically transcribed		
1	*I shot the prick*	/aɪ ʃɒt ðə prək/
2	*I shot that prick*	/aɪ ʃɒt ðɛt prək/

3	*I can't breathe*	/aɪ kaːnt briːð/
4	*I can't touch*	/aɪ kaːnt tatʃ/
5	*I can't help puking*	/aɪ kaːnt halp pjuːkəŋ/
6	NO SPEECH	n/a

These are not actually as different as they first appear. Ignoring for the moment the claims that speech is not involved at all (I will argue later that this can be discounted), they all agree, for example, that the disputed utterance begins with "I", and the difference between *prick* and *breathe* is not desperately great either, as both contain a bilabial plosive followed by an /r/ followed by a non low vowel. Even *pric**k*** and *pu**k**ing* share a /k/.

The main difference between these claims is of course between the Crown's *I shot* and the others' *I can't.* This means that the first thing that separates the Crown's hypothesis from the rest is the difference between a velar plosive phoneme /k/ in *can't* and a palato-alveolar fricative /ʃ/ in *shot*. It will obviously aid the Crown's case considerably if it can be shown that the sound after the "I" (which both agree on) is more likely to be a /ʃ/ rather than a /k/, and *mutatis mutandis*. Since phonemes are abstract things, for this to happen the Crown needs to show that the allophone in question – the actual sound occurring – is an allophone of /ʃ/ rather than of /k/. What I will show below is that it is in fact the other way round: the allophone occurring after /aɪ/ cannot be an allophone of /ʃ/, and is very likely to be an allophone of /k/.

In my first report, I observed this on the disputed utterance:

> *4.1 Comments on vocalisation* U.14. Since I suspect that this is the allegedly incriminating portion, I shall comment on my hypothesis in some detail. This portion remained unclear to me during my first few passes. Since I could not make out its linguistic message, I transcribed it phonetically as narrowly as I could and left it until later. My transcription is at 4.1
>
> [ç̬ə̯ ɐ̯**ç̬**əʔə̆ pʉ̯.ɪ̯] (the dot indicates a syllable boundary). (4.1)

Later, I said I thought the allophonic string [ɐ̯ç̬əʔə̆ pʉ̯.ɪ̯] realised /aɪ kaːnt halp pjuːkəŋ/, i.e. *I can't help puking*, explicitly aligning[2] the /k/ in *can't* with what I transcribed as [ç]. Now, [ç] represents a palatal fricative. (For many speakers of English, it occurs as the realisation of /h/ before /j/ in words like *Hugh* /hjuː/, [ç̬uː].) So my transcription at 4.1 was claiming that I heard a palatal fricative [ç] and my interpretation of this was as an allophone of a velar plosive phoneme /k/. So the first thing to note is that *I did not hear a* [ʃ], which would have been the realisation of /ʃ/, had Bain said *I shot*. The second thing to note is the sound I heard –

[2] I made a mistake in table 4.3 in assuming that the vowel phoneme in NZ English *help* was /I/. It is in fact /a/. It is nice to note this actually makes the acoustic evidence more probable, given the hypothesis.

a palatal fricative – cannot be an allophone of /ʃ/. If the sound is a palatal fricative (I present acoustic evidence in due course to strongly support this) *Bain cannot have said shot*.

The next thing to do is to establish the connection with the word *can't*. To do this I need to explain how a palatal fricative allophone might be expected to realise a velar plosive phoneme, the sound at the beginning of the word *can't*. This needs explaining because a palatal fricative [ç] and a velar plosive [k] are two rather different speech sounds. They differ both in *manner of articulation*: plosive *vs* fricative; and *place of articulation*: velar *vs* palatal (these terms were explained above). The difference in place has in fact already been explained in the previous section. It is normal in English for /velar plosives/ to be fronted to a palatal (or fronted velar) place in the vicinity of front, especially high and mid-front vowels (Laver 1994: 376-378). There is actually a nice demonstration of this in Bain's speech when he says *I came (home…)* (my ref: U4; control+click to hear). The /k/ in *came* is audibly fronted to [k+] before the following front vowel [e] of /ɛɪ/.

What could have caused this fronting? As already explained, normally one expects a following vowel to exert a fronting or backing influence on a preceding, especially dorsal, consonant. In this case, the nature of the vowel – how front or back – is not clear (though it will become so). The fronted nature of the consonant could also be the result of a perseverative co-oarticulation from the preceding front off-glide in the /ai/. This would also have the effect of pulling the body of the tongue forwards to a fronted velar, or palatal position.

There is, however, evidence from elsewhere in the call that the fronting is due to the vowel. After the operator asks *and your last name?* (my ref: U21), I noted in my transcription that, before answering *Bain*, Bain makes the sounds [ˇɪ cˈɐˑ], where c is a fronted velar [plosive]; possibly weakly ejected, or aspirated. I further noted that this sounds very much like Bain was saying, again, *I can't*.

This utterance is acoustically much clearer than the disputed one. Its vowel is phonated, for one thing. It is worth noting that there is a good deal of agreement on the part of the other transcribers – including detective Ward – that Bain said a velar plosive of some kind here, and most agree that it is followed by an *a* or *er* vowel. Innes writes ik- ahh; Dempsey/Ward write ah ga ah; French and Harrison write (Ig-er); Foulkes writes (G-) uh- ; Rose writes (I ca-). Guillemin writes Ah, and is the only one not to note a velar here. Now, although the consonant is clearly fronted, one does not expect the vowel in this word (/a:/) to be front, but central, and I suspect either that it is the case that Bain simply fronts his /k/ before /a:/, or that he has a fronter than normal allophone for /a:/, or that the fronting is a part of the stress of the moment. What this does show, absolutely clearly, is that he has a fronted velar before /a:/, and that it would not be a surprise, if the disputed word were *can't*, to find a fronted velar.

Fronting a plosive results in a plosive. The disputed sound in question is a fricative. So where does the fricative element come from? It is quite common in running speech for /plosives/ to fail to make complete closure and thus be realised as quasi fricatives, especially between relatively open vowels. This was shown by Elliott in her 2001

paper to be quite common for the velar plosive in Australian English *okay*. The main thing to understand here is that a fronted velar plosive, when fricativised, is effectively the same as a palatal fricative, that is, the sound which we actually observe. This is illustrated in figure 4.3, again from Laver (1994), which shows the x-ray tracings for a fronted velar plosive [k+] and a palatal fricative [ç]. You can appreciate how, if you fail to make contact in a fronted velar plosive, a palatal fricative may result.

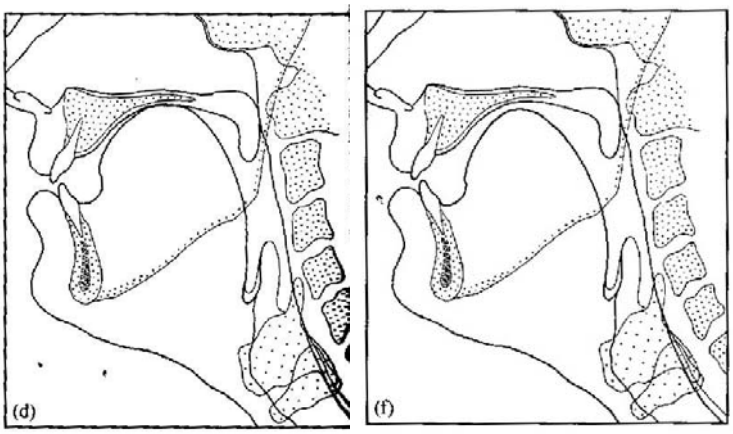

Figure 4.3. Tongue position in a fronted velar plosive (left) and a palatal fricative. After Laver (1994 pp.377, 246). Reproduced with permission.

To sum all this up, I have shown that it is perfectly plausible that the palatal fricative [ç] in the disputed utterance has arisen as the allophone of a velar stop /k/, by way of velar fronting. The speaker aims for a velar plosive to realise /k/, it is perturbed forwards to a fronted velar, but the speaker only achieves the occlusion typical of a fricative, which results in the palatal fricative allophone. I would maintain, further, that the responses of the transcribers who wrote *can't*, were reflecting the underlying velar.

Finally, it is important to note that it sounds as if Bain attempts to say *I can't* again right after the disputed portion (myref: U.15 control+click to hear). So there would appear to be three repeats of this in the call, two of which begin with a fricative and one with a plosive (control+click here to hear them all together). In my report I transcribed the second repeat with a palatal fricative too: ((( ) = [ɸ̞ç])), and noted "Most likely again he starts to say *I can't.*" The palatal fricative in this second putative *I can't* is clearly the origin of the *sh* in the second *I shot* heard by detective Ward and others, and for exactly the same reason as the first. So detective Ward and others have misheard here too.

That the disputed allophone is more likely to realise /k/ than /ʃ/, and that therefore the first part of the disputed utterance is more likely to represent *I can't* rather than *I shot*, is of course support for the defence, but it needs to be backed up by acoustic analysis. This we now turn to. I will focus on two questions: (1) how likely is it that the disputed utterance is not speech; and, more importantly, (2) how likely are the acoustics, assuming that Bain said *shot* rather than *can't*. The evaluation of the rest of the disputed utterance is immaterial, once a decision has been reached with respect to

this question. Although I could investigate the rest, I suspect that the acoustics of this part would only be slightly more probable assuming *breathe* or *puking* than *prick*, and therefore not be especially informative (although still in support of defence). I hope that this comment à la Fermat will suffice.

6.0 Acoustic Analysis

6.1 Aim

This acoustic analysis makes use of *formants* to compare known and disputed data. Very simply, but essentially, when you speak, the air in your vocal tract vibrates at certain frequencies, called formants. The frequencies are determined by the shape of the vocal tract and its overall length. The shape of the vocal tract is determined by the particular sound you are making: if you are saying an *a* vowel as in *can't*, you put your vocal tract in a different shape from when you are saying *o* in *shot*, and the formants consequently differ. The overall length of the tract is a reflection of your overall size: tall men generally have longer vocal tracts than short men, and therefore lower formants; women have shorter vocal tracts than men, and therefore higher formants. There are usually several formants involved in a sound, and these can sometimes be related to separate bits of the vocal tract. The ensemble of formants is called an *F-pattern*. Formants can be extracted by computer and measured. They are also known as vocal tract resonances, and I also use the term resonances here.

So the idea, and reasoning, is simple: if you can show that the F-pattern for a disputed sound is very similar to that of a speaker's known sound x, and different from that of another of their known sounds y, this means that you are more likely to get the disputed F-pattern assuming that the disputed sound was x rather than y. This would then constitute support for the hypothesis that the disputed sound is more likely to be x than y[3].

This acoustic analysis section starts by addressing the hypothesis that the disputed portion does not represent speech, and will argue on the basis of the acoustic evidence that the first part of the disputed utterance is in fact not non-speech but "I", (as is agreed by most investigators). Then I concentrate on the part of the disputed acoustics that is heard as *sh* in *shot* or *c* in *can't*, and show that the sound is far more likely to be the first sound in the word *can't*. Finally I compare the part of the disputed portion that corresponds to (*I*) *can't/ (I) shot* with a known (*I*) *ca-* and show their acoustics are very similar overall, and nothing like *shot*.

6.2 'Non-speech' hypothesis

As already noted above, French and Harrison state of the disputed portion that it

> … could be heard as 'I shot the prick', or 'I shot that prick'. However, whilst we cannot discount the possibility that the material amounts to speech […] it also remains entirely possible that it is not speech. Rather, it could be no more than an audible out-breath that has, in the distress of the moment, been modified by a random and unfortunately-sequenced series of movements of the tongue and lips so as to create a series of sounds that could – albeit with a little effort – be heard as 'I shot the/that prick'.

[3] Prior odds pending – see section 3, footnote 1

Their reasons for this conclusion are as follows. They note (p .6) that the disputed call is produced on an outbreath with *no voicing*, and *modification of the airflow in the oral tract*[4]. This means that Bain's vocal cords were not vibrating as one would normally expect them to be doing if he was speaking normally, but that you can hear that he was moving his tongue and lips. They then note that there are a number of parts in the call that sound similar to the disputed portion in being produced on an outbreath with the impression of vocal tract movement. They note that Bain clearly produces speech on an outbreath with no vocal cord activity when he whispers part of the telephone number. So they conclude that the disputed portion could be non-speech, which could be heard as speech, namely *I shot the prick*. As I read this, it is a claim that the disputed portion is not speech, but that it could be heard as such. I also note the degree of scepticism as to the possibility that Bain did in fact say *I shot the prick* conveyed by their rider *albeit with a little effort*.

In the video linkup notes of evidence, a slightly different view is expressed, where Harrison agrees with Mr Raftery's question (p.10):

> … you and Professor French have said that those words may be there but it is equally possible they may not be and they may be exhaled breath.

It is therefore not exactly clear what is being claimed, as *equally possible* and *entirely possible* are not the same thing. Moreover, unfortunately you can't really do much with the word *possible* when it comes to evaluating different hypotheses, and it is doubly difficult to interpret what *equally possible* means with respect to two mutually exclusive but non exhaustive hypotheses, like *it's not speech* and *he said I shot the prick*. It is a pity that Raftery's question did not use the word *probable*, since *equally probable* would have been interpretable. You cannot blame French and Harrison for this.

Possibility and probability structure two different semantic dimensions (Broeders 1999). Something being possible says nothing about its probability other than it cannot be 0%, which means the event is certain not to occur. Two things being *equally possible* then can be taken to mean that they are equally probable, but that that probability may have any value whatever, other than one. If both events are mutually exclusive (which they are) and exhaustive (there was no other interpretations possible than *I shot the prick* and *it's not speech*), then their joint probabilities of occurrence must sum to 100%, which means that the two hypotheses here – that it was not speech and that Bain said *I shot the prick* – would both have to have a probability of 50% each. But there is no guarantee that French and Harrison imply that the events are exhaustive (there could be some other hypotheses, as indeed there are), so the best interpretation I can put on their *equally possible* is that both hypotheses are equally probable, but that that probability may have any value whatever, other than one. I cannot interpret *entirely possible* at all. This is all epistemologically very weak (it doesn't tell you much), and is therefore not very useful from the point of view of working out which hypothesis is more likely to be true.

---

[4] This is not totally accurate: as I pointed out in my first report, the spectrogram of the disputed portion contains a short amount of periodicity, which shows that Bain's cords were in fact vibrating for a short time.

French and Harrison adduce as evidence for their claim (p. 7) that they were able to *edit[ing] together parts of out-breaths from different areas of the call* which *could easily be heard as whispered utterances*. In other words, they assembled bits of known non-speech from Bain's exhalations that one could hear as speech. I don't doubt this is true, especially as many of Bain's actual utterances end in exhalations, and one would expect them to therefore carry the acoustic imprint of the preceding vocal tract speech gesture. It also nicely shows that their claim is *possible*, as they say. It must also be emphasised that their main purpose in doing this was to demonstrate *the danger of misinterpreting the caller's audible exhalations as speech*. But, as I have just argued, possibility is not probability. *The fact that it is possible has no bearing on the probability that the disputed portion is not speech*. That is a totally different claim.

In addition their claim re randomness runs very close to being untestable, and therefore vacuous. As I understand it, they claim that the acoustics look like and sound like speech but in fact are nothing but random movements of Bain's vocal tract caused by emotional distress and moving by chance in the same way as speech, whilst not realising some underlying communicative intent. Under this scenario, there is no way one could distinguish between speech acoustics that realised the speaker's communicative intent without the accompanying phonation (like the whispered telephone number); and identical acoustics that are just "random". The *reductio ad absurdum* for this position would be the preposterous claim that the whispered telephone number is just random movements of the speaker's vocal tract. Unless some definition of randomness is given, and how to recognise it, it is difficult to make use of this claim, one way or the other.

It is also a highly improbable claim. In order to produce speech – even something so apparently simple like the word *I* – the different parts of the vocal tract have to accomplish an enormous amount of precisely choreographed and co-ordinated movements realising an equally complex set of neural command precursors. Getting, by chance, a vocalisation that was identical to speech but not speech would be exceedingly improbable. Moreover, as I pointed out above, it is possible that the initial part of the disputed utterance – the putative *I ca-* – recurs at least once, which would square the already vanishingly small probability of its acoustic pattern being random.

I think the best way out of this impasse is to recast the hypothesis to invoke the probabilities of the evidence, rather than the possibilities of the hypotheses. Then it can be evaluated. We need to ask: what is the probability of getting the observed acoustics, assuming they realise speech sounds; and what is the probability of getting them, assuming they realise, in French and Harrison's terms "… an audible out-breath that has, in the distress of the moment, been modified by a random and unfortunately-sequenced series of movements of the tongue and lips so as to create a series of sounds …" that sound like *I shot the prick*.

I demonstrate this with the first portion of the disputed call, which all parties except for those who think it is non-speech, agree in hearing as "I", but which we can simply call *the acoustics at the beginning of the disputed utterance*. The sound of "I" is usually transcribed as the diphthong /aɪ/. We need thus to compare two conditional probabilities:

(1) What is the probability of getting the acoustics at the beginning of the disputed utterance assuming Bain is saying /aɪ/?

(2) What is the probability of getting the same acoustics assuming they are the result of an " … audible out-breath […] modified by a random […] series of movements of the tongue and lips.

As luck would have it, there are several unambiguous tokens of /aɪ/ in Bain's call, which will allow us to estimate the first probability. Three tokens of /aɪ/ are stressed (said clearly) and occur in the word *five*: one in the address *sixty five* and two in the telephone number. The first occurrence in the phone number, as pointed out by several observers, is whispered in its entirety. To show what a typical F-pattern of /aɪ/ looks like, figure 6.1 shows the non-whispered /aɪ/ from the telephone number. Its time-varying F-pattern is very clear and as expected for this diphthong, with F1 increasing then decreasing; F2 increasing; and F3 more or less stable. Labial offset transitions are visible, but not onset. To show what formant frequencies look like I have superimposed them on the spectrogram: they are the dotted red lines, and you can see how they run through the middle of the darker portions of the spectrogram.

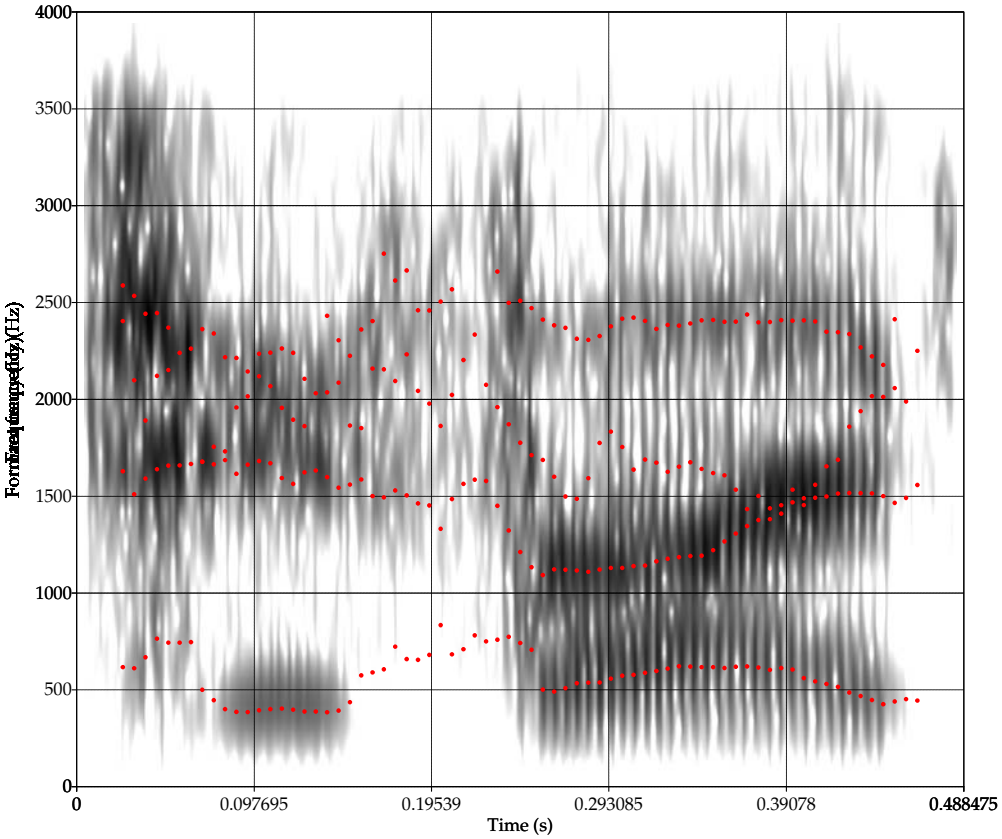

Figure 6.1 Wide-band spectrogram, with superimposed formants (*Burg*, 4 below 3 kHz) of Bain's non-whispered *five* in *two-five-two-seven*.

The wideband spectrogram of Bain's whispered *five* is shown in figure 6.2. You can pick out F2 and F3, but F1 is more difficult to see – it has been attenuated and shifted higher, as expected with the tracheal coupling from whisper.

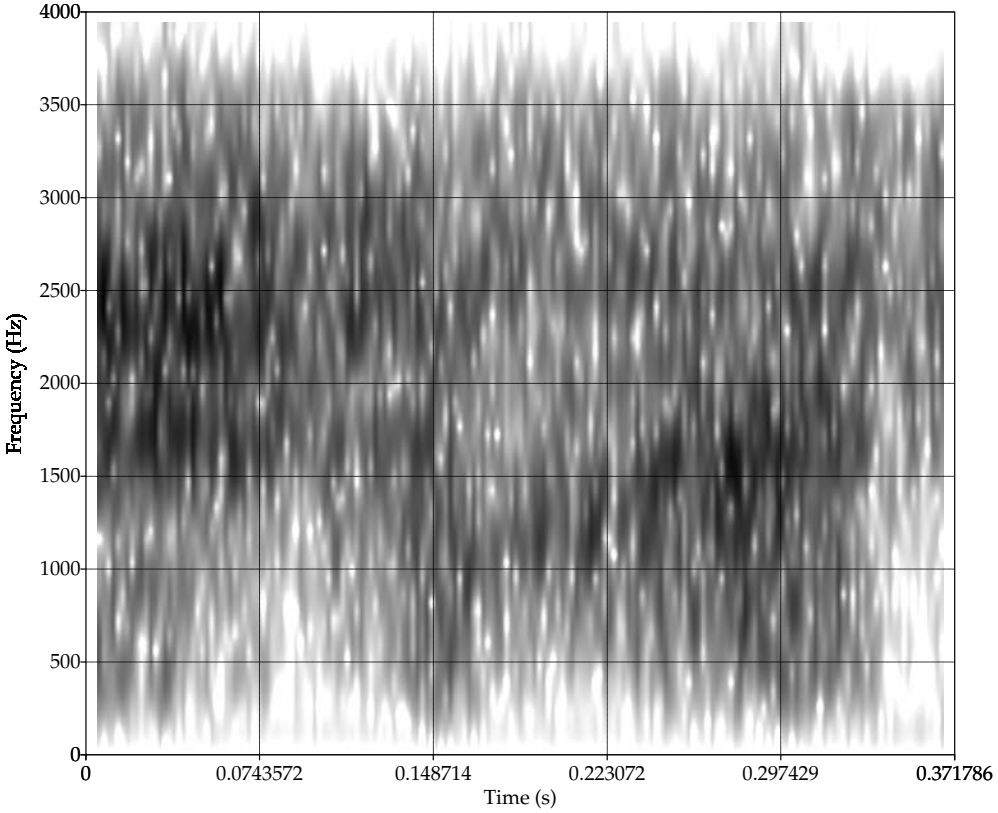

Figure 6.2 Wide-band spectrogram of Bain's whispered *five* in *two-five-two-seven*.

In addition, there are three other /aɪ/s in the call: two in *my .. my family* and one in *I came home* … . All of these are said rather more quickly than the first three, and have considerably shorter duration. Like one of the *five*s in the telephone number, they also have non-modal phonation type, presumably from the emotion: one is creaky; one is falsetto at the end. These six known /aɪ/ tokens can be used to assess the similarity between the questioned portion, which French and Harrison say may be non-speech, and the /aɪ/s which are noncontroversially speech.

Figure 6.3 shows a wideband spectrogram of the initial part of the disputed utterance, including the last part of the preceding inbreath. This is the first part of what French and Harrison claim could be random.

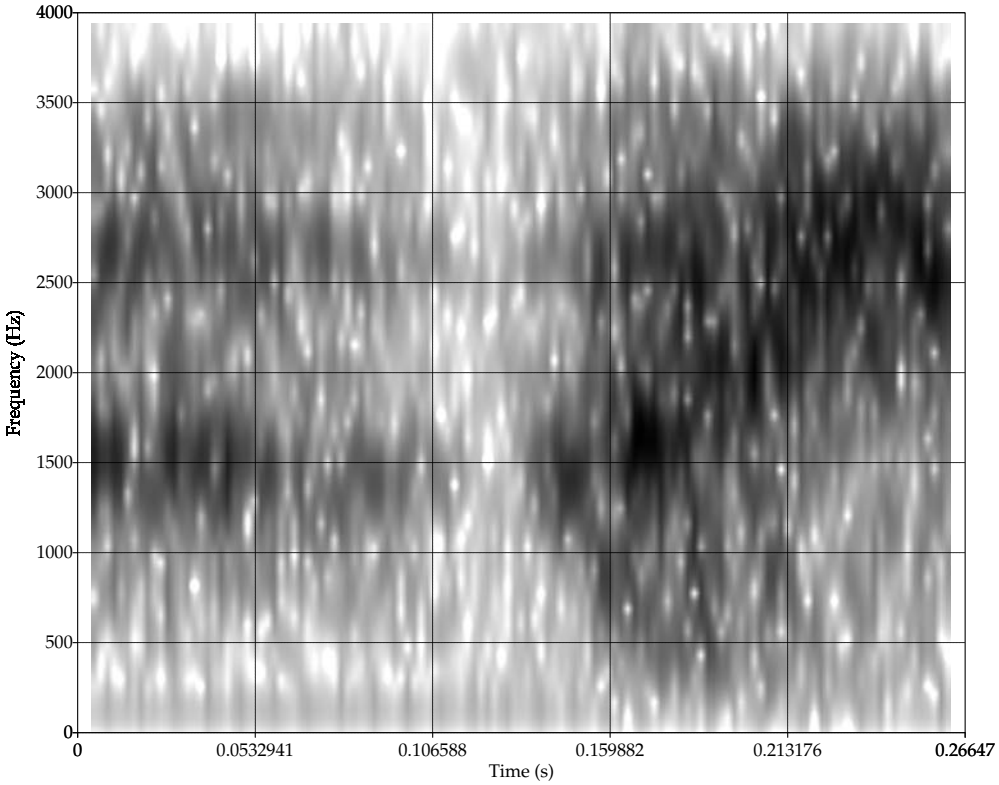

Figure 6.3. Wideband spectrogram of first part of disputed portion. On the left is an inbreath.

A very clear resonance can be seen sweeping up from about 1.4 kHz, a little below the frequency of the lowest strong pole of the inhalation, to about 2.7 kHz. There is also another strong resonance above this, which starts at just above 2.5 kHz and perhaps increases a little to 2.8 kHz. There is also some higher amplitude energy between about 500 Hz and 1.5 kHz.

I extracted with *Praat* the F-pattern of the six known /aɪ/s (*Burg*, 4 formants below 4 kHz). I then plotted F1 – F4 to get an idea of the distribution of F-pattern in Bain's /aɪ/. Because the tokens are of different durations, and because they clearly are more similar at offset than onset, I have plotted them aligned at offset. I then extracted the F-pattern of the questioned portion, and superimposed it on the known /aɪ/s, again aligned at offset. The result, which is not really a surprise, is shown in figure 6.4.

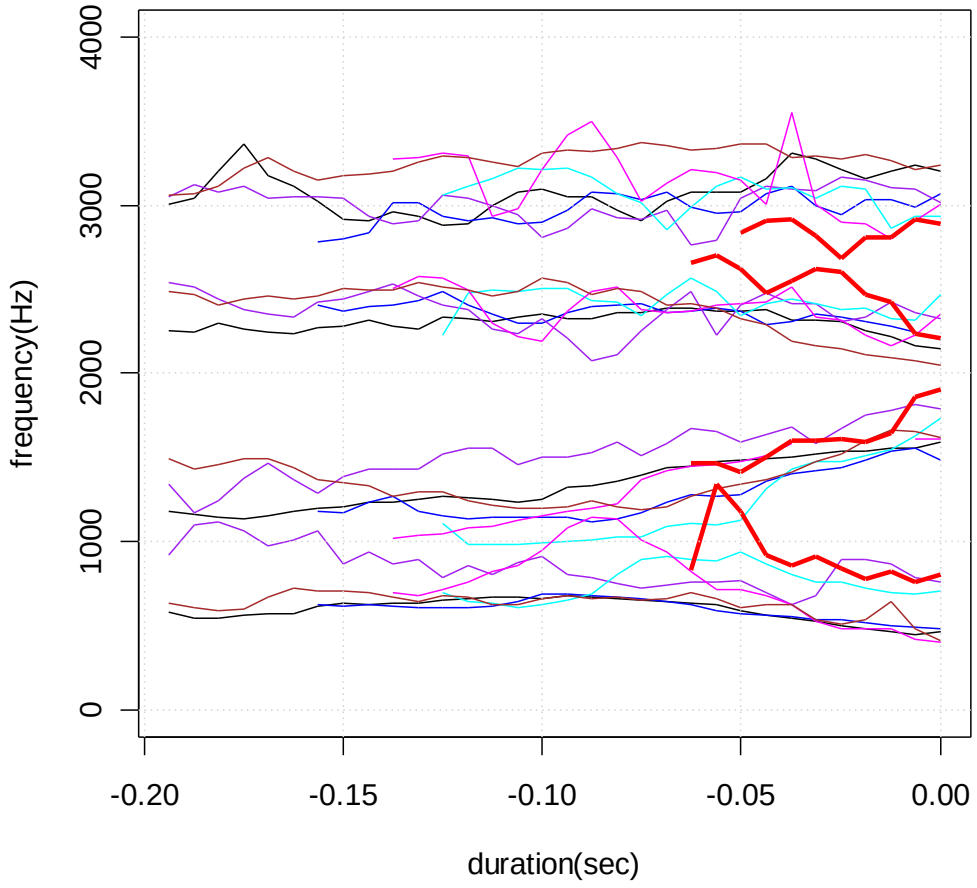

Figure 6.4 F-pattern of six of Bain's /aɪ/ tokens (thin lines) compared to F-pattern in disputed portion (thick red line). F-patterns aligned at offset.

Figure 6.4 shows, firstly, that the F-pattern in the disputed portion is shorter than in the known values, as expected from its auditory impression. This is what is expected from an /aɪ/ representing "I" in utterance-initial position. The agreement is quite good, especially in overall configuration (note e.g. the agreement in formant slope). Extremely good agreement in F2 can be seen, where the questioned trajectory lies more or less in the middle of the distribution of the known values. F3 agreement is also very good towards the end of its time course, but above the top of the known distribution earlier on. F4 agreement is poor, lying below the known distribution for about two thirds of its time course. F1 is within the known distribution for about half of its time course. Its apparent overall higher location relative to the known data is because the questioned data was said voiceless, and this is known to raise the frequency of F1 (as well as increase its bandwidth).

Eyeballing the distribution of the questioned and known /aɪ/ data, I would think that overall one would be about 60% likely to get it if it realised Bain's /aɪ/. This estimate would of course change depending on what subset of formants you looked at. You

would for example be extremely probable to get the questioned F2 if it was Bain's /aɪ/, but this has to be balanced against the unlikelihood of getting the questioned F4 if it was his /aɪ/. It is possible to evaluate it more accurately with a appropriate multivariate likelihood ratio, against a reference sample, as done in Rose (2006), and Rose Kinoshita & Alderman (2006), but this would be steamrollering the nut. The main thing here is that you would be very very highly unlikely to get this pattern if it was the result of a random set of gestures: the set of individual random non-speech gestures is presumably very big, so the probability of getting, by chance the right ones, *and* in the correct order, is very small. This means that the strength of evidence in favour of these acoustics being one of Bain's /aɪ/s rather than non-speech must be very big.

To make the point another way, one could also take a leaf out of French and Harrison's book and examine all Bain's known exhalations in the call (I have transcribed them in first report) for an F-pattern configuration that resembles the questioned F-pattern. I have done this, and I cannot find one. That is an indication that the probability of getting an /aɪ/-like F-pattern from exhaled non-speech is very small (infinitely small, if you just take this call).

I don't think it is worth belabouring the point any further, especially as French and Harrison concede that the questioned portion may also be speech. The only thing that needs to said therefore is that there is strong evidence in support of the hypothesis that the questioned portion *is* speech, and in fact /aɪ/ as most people heard. I will henceforth no longer consider the proposal that we are dealing with non-speech here.

6.3 Analysis of sh/c

Figure 6.5 shows a spectrogram of the disputed utterance that was heard as *I shot* by detective Ward and others, and as *I can't* by the rest. I have discussed the part corresponding to the *I* in the preceding section. The subsequent portion, lasting from about csec. 12 to csec. 18, corresponds to what was heard as *sh* or *c*, and can be seen to consist of narrowband high level aperiodic energy located between about 2.4 and 3.3 kHz.

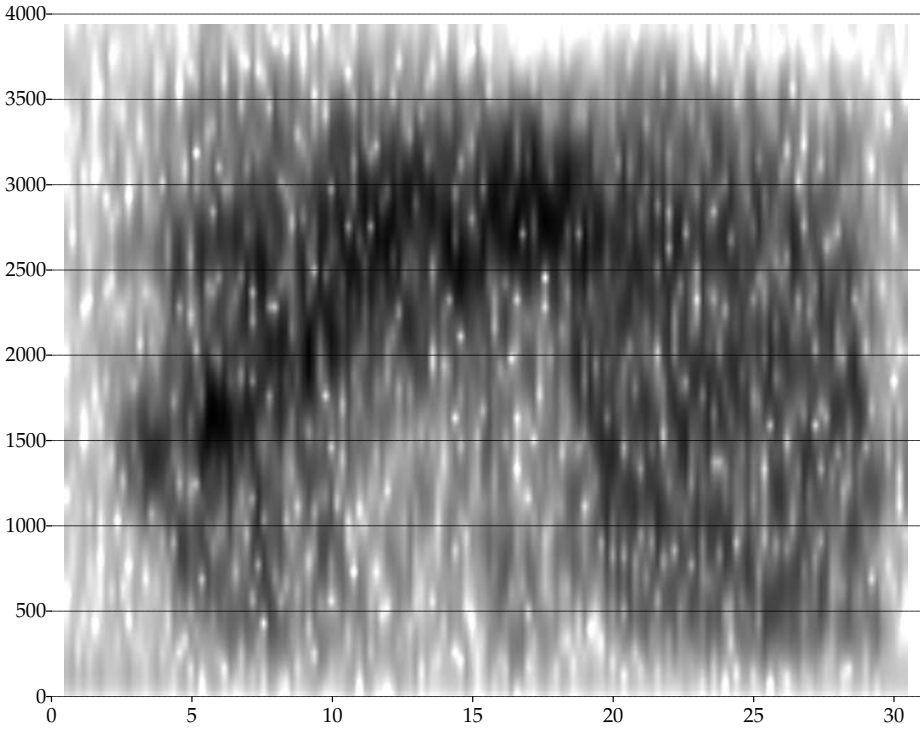

Figure 6.5 Wideband spectrogram of the portion of the disputed utterance corresponding to *I shot/ I can't*.

Figure 6.6 shows the same portion with superimposed formants (*Burg*, 4 below 4 kHz).

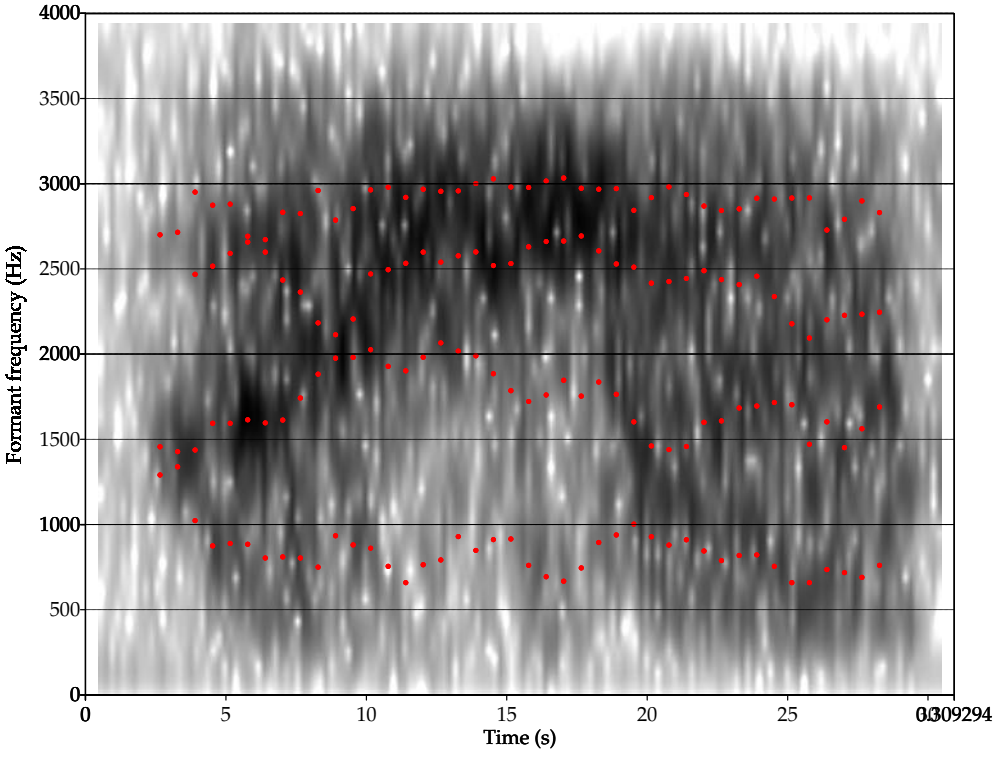

Figure 6.6 Wideband spectrogram, with superimposed formants, of the portion of the disputed utterance corresponding to *I shot/ I can't*.

It can be seen that the high amplitude portion consists of two poles separated by about 0.3 – 0.4 kHz. The higher pole is at about 3.0 kHz, the lower at about 2.6 kHz. The extracted pole meandering between about 2.0 and 1.7 kHz is continuous with adjacent F2 and is therefore, under the circumstances, a rather well extracted F2. (The origin of the pole between 1.0 and 0.5 kHz is not material to the discussion: it may be an F1 tuned up by the tracheal coupling.) The continuity of the upper poles with the F-pattern of the preceding vowel is difficult to see clearly, but the lower of the two strong poles appears continuous with F3. This also agrees with Bain's values for F3 in high front vowels, which are at about 2.5 kHz.

This segment has the auditory and acoustic properties of a voiceless fricative, like *sh*, and this is presumably why it was heard at the beginning of the word *shot* by a non-phonetician (I noted in my first report, before I was told that it had been so heard, that I suspected that some might hear it as *sh*).

But it is not a *sh*. First of all, phonetically, it does not even sound like a *sh*. As pointed out above, its auditory-phonetic characteristics are more those of a voiceless palatal fricative [ç] – a sound made further back in the mouth and in a very different way to the sibilant [ʃ]. Secondly, its acoustics are not what one would expect from a *sh*. You would, in fact, be far more likely to get the distribution of acoustic energy in the disputed segment in question if it was articulated a little further back in the mouth, in the position of a palatal fricative [ç].

The acoustic evidence supporting this, mostly in the acoustic phonetician's bible *Acoustic Phonetics* (Stevens 1998: 403ff.) is as follows. As well as F2 transitions into and out of adjacent vowels, palato-alveolars are characterised by several aspects of their spectrum. Firstly, there is the auditorily important characteristic lower bound at around 2.0 kHz, which contributes to its lower pitch and helps distinguish it from /s/. In his (1960) study of fricative spectra, Strevens states (p.41) that in [ʃ] the "Lowest frequency varies between 1600 and 2500 cps .. peaks of energy tend to occur not less than 1000 cycles apart and the aspect of amplitude cross sections shows a weighting towards the bottom of the pattern.". Figure 6.7 reproduces his figure of the extent of [ʃ] spectra he examined and measured. The lower bound at between about 1.6 and 2.5 kHz can be easily seen.

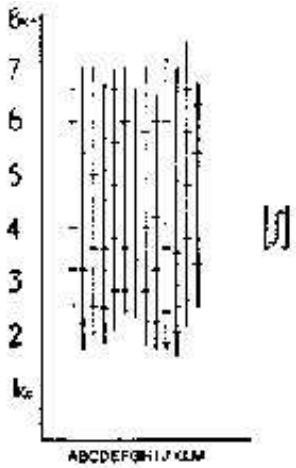

Figure 6.7. Spectral extent of [ʃ] in 12 males reproduced from Strevens (1960: figure 3e).

It can be seen in the figure of the disputed portion (figure 6.5) that the lower bound of its fricative part is clearly at the higher end of these spectra – at about 2.4 kHz – and is therefore atypically high for a palatal-alveolar. The narrowband nature of the disputed fricative is also atypical for a palato-alveolar, which, as can be seen in figure 6.7 is typically wideband, extending way above the disputed fricative's upper bound of 3.3 kHz[5]. From the point of view of its high lower bound and narrowband distribution, then, you would be unlikely to get the disputed fricative acoustics if they had come from a palato-alveolar *sh* sound.

In [ʃ] one expects from the acoustic theory of speech production to find two strong poles below 4 kHz, separated by about 1 kHz or a bit less. The highest pole, at about 3.5 kHz or slightly lower, is contributed by the quarter-wavelength resonance of the sublingual cavity tuned down by any labialisation. The next highest pole, at about 2.5 kHz, is contributed by the quarter-wavelength resonance of the palatal channel behind the post-alveolar constriction. Finally, a pole from the configuration behind the post-dental source, together with a weak back cavity resonance, may contribute to a spectral prominence around 1.9 kHz.

These characteristics can be exemplified in a beautiful example of a /ʃ/ from the emergency operator, when he says *very shortly*. Note too that this /ʃ/ is in a very similar phonological environment to *I shot*, so it shows you – a little!!! – what *I shot* might look like acoustically. As a nice example of the cussedness of things, there is no known example of Bain's [ʃ] in the call, otherwise I would of course have used that. Figure 6.8 shows a wideband spectrogram of the operator's /ʃ/ with superimposed formants. The quality is very good, presumably because the signal was able to be recorded directly and not over the telephone channel. In figure 6.8, one can clearly see most of the features as described by Stevens and Strevens for /ʃ/, in particular:

- the highest /ʃ/ resonance, continuous with vocalic F4, at about 3.3 kHz.

[5] One has to remember that the lack of wideband distribution in the disputed portion may at least in part be due to the telephone effect, but it is clear from inspection of Bain's acoustics that that does not cut-in until somewhat above the frequency range in question. For example in several examples of Bain's speech spectral components are validly extracted up to 3.4 – 3.4 kHz).

- the next highest resonance, continuous with vocalic F3, at about 2.5 kHz;
- the /ʃ/ F2 continuous with the vocalic F2 at about 2 kHz;
- a low lower bound at ca. 2.0 kHz.

It is not clear whether the extracted resonance at 2.4 kHz is spurious or reflects the configurational pole. The F2 is weaker, but still contributes to broadband high amplitude energy down to about 2.0 kHz. Note in particular the ca. 700 Hz separation of the two highest poles.

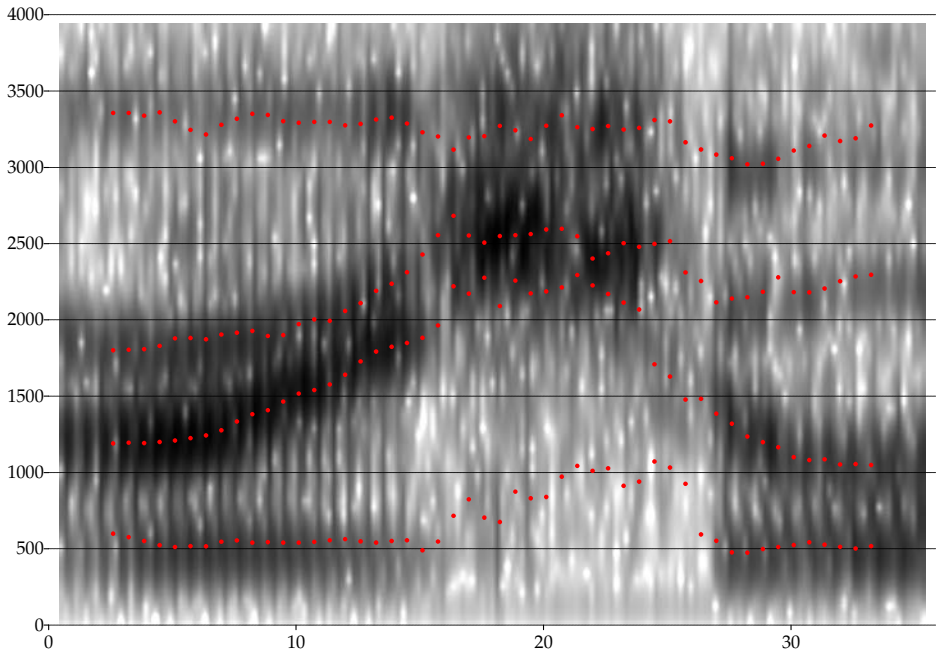

Figure 6.8. Wide-band spectrogram with superimposed formants (4 below 4kHz) of emergency operator's /ʃ/ in *very shortly*.

Now, if we turn once more to the disputed portion in figure 6.6, there is further evidence that its acoustics are atypical of a palato-alveolar but more typical of a palatal, which is not surprising, given that it sounds like a palatal and not a palato-alveolar.

Below about 4 kHz palatal fricatives are expected to have a narrowband distribution centered around 3 kHz. This can be nicely seen in Jassem's palatal fricative spectrum in Ladefoged and Maddieson (1996: figure 5.32, p. 177)[6]. The distribution is composed of two main resonances: one contributed from the cavity in front of the constriction, and one from the palatal constriction itself. A front cavity of between 3.5 - 4 cms length typical for a palatal would have a quarter-wavelength resonance of 2.2 – 2.5 kHz, which would be tuned up from its trumpet-like shape, its non-rounded opening and the non-finite impedance from the palatal channel. Johnson (1997: 120-121) points out that the peak resonance frequency in a palatal fricative matches the frequency of adjacent vocalic F3. The palatal constriction would contribute a half-wavelength resonance. Assuming, after Fant (1960: 73) an idealised palatal

[6] As Ladefoged and Maddieson note (p. 176), the other figure of a palatal spectrum on p. 176, which is diffuse rather than acute, seems to be confused with the alveolo-patalal spectrum.

constriction length of 6 cms, this would give a frequency of 2.9 kHz. A lower and weaker frequency, continuous with adjacent vocalic F2, would also be expected from the half-wavelength resonance of back cavity.

It is worthwhile noting that Fant (1960: 73) includes a case almost exactly the same as this in his discussion of the F-patterns of compound tube resonators and horns. The third example down in his figure 1.7-7, reproduced below, is an idealised resonator model for the production of a velar stop before [æ]. The excitation of this model would result, as shown, in an F-pattern very similar to that observed, with F3 being the λ/4 resonance of the front cavity, at (35,000/[4 * 4cm. =]) ca. 2.2 kHz, and F4 the λ/2 resonance at (35,000/[6*2] =) 2.9 kHz.

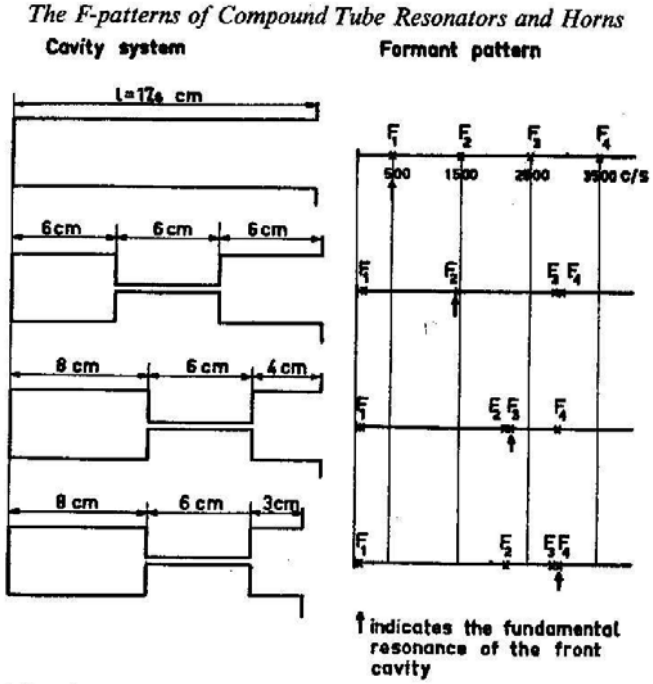

Fig. 1.4-7. Three-tube resonator models and corresponding F-patterns. The dimensions have been chosen so that the frequency of the fundamental resonance of the front cavity corresponds to F_1, F_2, F_3, and F_4 respectively. These idealized resonator systems show some essentials of the articulation of velar and palatal consonants.

Very similar characteristics to those just described as expected for palatal fricatives can be seen in the disputed fricative in figure 6.6. The highest pole at 2.9 kHz agrees exactly with Fant's predicted value and is likely to be the half-wavelength resonance of the palatal constriction. The pole at about 2.6 kHz, continuous with F3, would then be associated with a quarter-wavelength front cavity resonance. The weaker pole at ca. 1.8 kHz is continuous with F2 and is a back cavity resonance, assuming after Johnson (1997) an idealised back cavity length of about 10 cms. Thus once again these are the acoustics that you would expect to get more from a palatal fricative than a palato-alveolar.

It is important to emphasise here that absolute acoustic values cannot be used as indications of a particular sound. As was pointed out above, absolute values reflect not only the sound being made but also the anatomical dimensions of the individual making it. So, since one is not justified in assuming that the operator and Bain share

the same anatomical endowment, one cannot compare the absolute frequencies of their output as such. In any case, both Bain's disputed fricative and the operator's [ʃ] actually have in common a frequency at ca. 2.5 kHz! Rather than this indicating that they are making the same sound, it is more likely that Bain's 2.5 kHz value reflects his front *cavity* resonance in a palatal fricative, and the operator's 2.5 kHz value reflects the post-alveolar *constriction* in a palato-alveolar. It is rather the relative values that are important. For example, it is evident that the disputed fricative's two highest poles are separated by far less than the 1 kHz expected for a [ʃ]; and that they have a narrower bandwith and extend down further than in a typical palato-alveolar.

Because absolute values reflect not only the sound being made but also the anatomical dimensions of the individual making it, one important and testable thing that this acoustic analysis does commit us to, however, is this. If we assume the articulatory dimensions required to generate palatal fricative acoustics (i.e. a palatal constriction of ca 5.75 cms to generate a λ/2 resonance of 3.0 kHz ; a front cavity of ca. 3.4 cms to generate a λ/4 resonance of ca. 2.6 kHz; and a back cavity of about 9.5 - 9 cms to generate a λ/2 resonance of ca 1.8 – 1.9 kHz, this would predict an above-average vocal tract length for Bain of ca. 18.5 + cms (the notional Caucasian male tract length is assumed to be 17 – 17.5 cms.). For this analysis and its conclusions to hold, then, *Bain would have to be a taller than average man*. This was in fact confirmed by Karam on Sunday 21st, who suggested a height of about 6 foot 3 or 4 inches. This information of course now increases considerably the probability of getting these acoustics assuming Bain said a palatal rather than a palato-alveolar fricative.

6.4 Comparisons with known data

In this section I compare the F-pattern of the first part of the disputed utterance corresponding to *I ca-* with the F-pattern of the portion after the operator asks *and your last name?* (my ref: U21). It will be recalled that this was transcribed by nearly all investigators as a velar stop followed by a non-low vowel, mostly *a*. I'll refer to it as the *I ca* utterance.

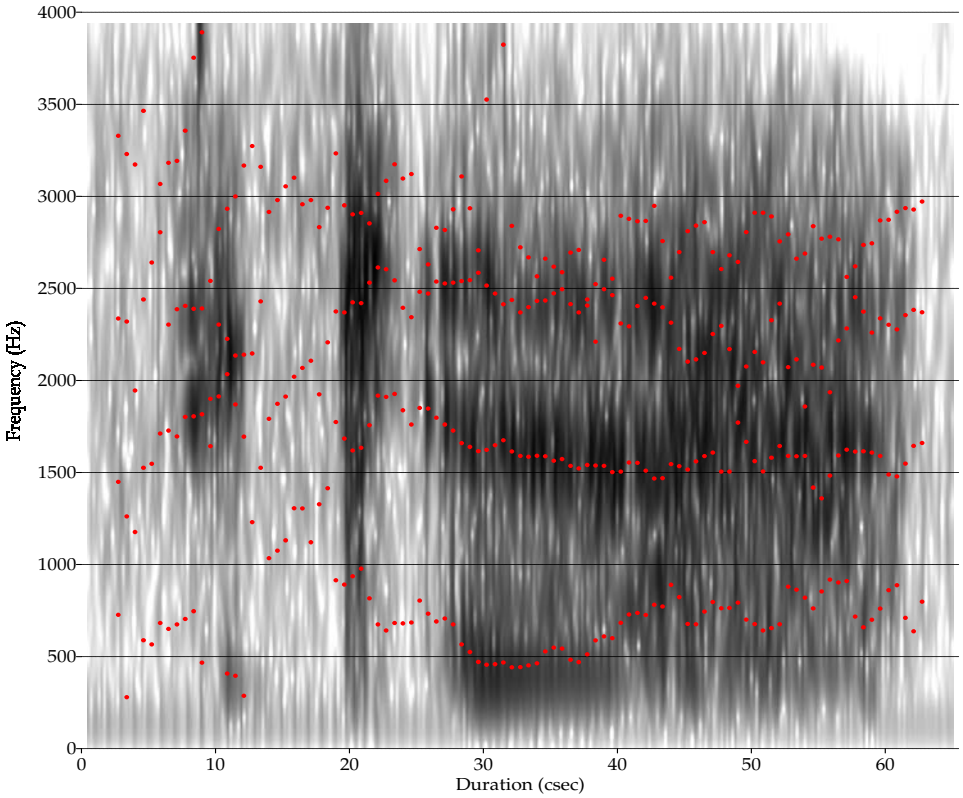

Figure 6.9 Wideband spectrogram, with superimposed formants (*Burg*, 4 below 4 kHz) of *I ca* utterance.

Figure 6.9 shows a spectrogram, with superimposed formants, of the *I ca* utterance I heard as the beginning of *I can't* and transribed as [ɪ̆ cˈɐˑ]. The spectrogram shows a short, barely phonated portion with F2 and F3 in expected positions for a high front vowel, and showing a clear velar pinch. This portion corresponds to what I transcribed as [ɪ̆], and I assume corresponds to the last part of an /aɪ/ diphthong realising "I" before a velar. A ca. 10 csec quiescent portion follows which I assume is the hold phase of the plosive. A ca. 3-4 csec. abrupt high amplitude release is followed by a low amplitude portion before laryngealised onset of phonation. This is what I heard as weak ejection. The release has an F2 at just below 2 kHz, an F3 at just over 2.5 kHz, and an F4 at just under 3 kHz, all typical of Bain's fronted velar release (these values can be seen to be very similar to his [c] in *I came home*), and this plosive I heard as a fronted velar. There is then 15 csec. or so of clear periodicity with clear F-pattern corresponding to a low (< high F1) frontish (< high F2) vowel. (I transcribed this as a slightly prolonged central vowel ɐ). Phonation ceases and cedes to a noise-excited F-pattern, at vowel target, of an exhalation before Bain can finish the word.

I don't think it is contentious that we have here an example of Bain starting to say /ka:/. It is also highly likely that the preceding sound is him saying "I". It is therefore useful to compare the F-pattern in this known sequence of sounds with the first part of the F-pattern in the disputed portion, which is claimed to represent either *I can't* or *I shot*. This is shown in figure 6.10. The F-patterns cannot be compared without special alignment, since the duration of their consonantal portion differs. So I have aligned

them in two different ways: in the first, in the top panel, the F-patterns are aligned at the point of onset of the consonant/offset of /aɪ/. In the bottom panel they are aligned at the offset of consonant/onset of the vowel. This allows the true relationship between the known and disputed F-patterns to be seen.

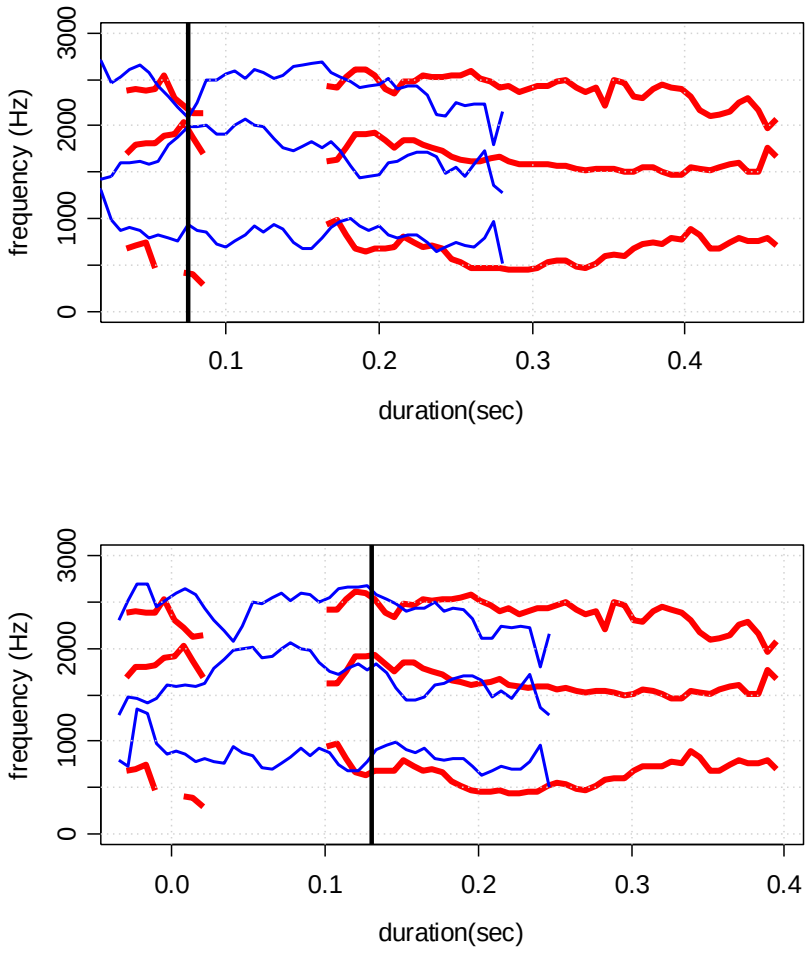

Figure 6.10. Comparison of F-pattern in the part of Bain's disputed utterance corresponding to *I shot/I can't* (thin blue lines) and in his *I ca* (thick red lines). Top panel shows F-pattern aligned at onset of consonant; bottom panel shows F-pattern aligned at offset of consonant. Vertical lines show point of alignment.

As can be seen, figure 6.10 shows very good agreement in F-pattern between the *I ca* utterance, and the portion of the disputed utterance corresponding to *I shot/I can't*. The velar pinching is very clear in both (top panel), and the frequencies at the onset of the vowel (bottom) also show good agreement. Indeed, the F-pattern during the *I ca* vowel and the portion after the fricative in the disputed utterance agree nicely too, even to the extent of a higher F1 in the voiceless disputed utterance. This shows that the probability of getting the first portion of the disputed acoustics assuming that Bain said *I ca-* is rather high: they are the acoustics you would expect to get if he had said this.

I have already demonstrated that the acoustics would be far more likely assuming Bain said /k/ rather than /ʃ/. As far as the vowel goes, if Bain had said the vowel /ɒ/

in *shot*, you would expect, according to one of basic principles of acoustic phonetics concerning the relationship between vowel quality and acoustics, to at the very least see an F2 sweeping down to a much lower value, around 1 kHz. What this would look like can be seen in the operator's *shortly* in figure 6.8. The vowel in *shortly* is of course not /ɒ/ but /oː/, but its F2, reflecting backness and rounding, should be comparable to /ɒ/.

There is no evidence for this low F2 in the disputed portion, which, as already demonstrated actually has an F-pattern like the *a* in *I ca*. One must conclude therefore that you would many times more likely to get the acoustics if the vowel was /aː/ rather than /ɒ/, which again is strong support for the hypothesis that he said *I can't* and not *I shot*.

6.5 Summary

The acoustic analysis section above has shown firstly that one can effectively discount the position arguing that the disputed portion was non-speech. Secondly, and more importantly, it has shown that you would be more likely to get the acoustics of the disputed *sh/c* segment if it was not *sh in shot*, but made further back in the mouth, in the vicinity of the palate. As explained, this kind of sound is a possible realisation of the initial k sound in *can't*. Finally it has been shown that you would be more likely to get the acoustics of the segment following the consonant if it had been *a* in *can't* than *o* in *shot*. In other words, these are all the acoustics that you would expect to get if Bain had said *I can't*, and they are definitely not the acoustics you would expect to get if he had said *I shot*.

7.0 Summary

This report was an attempt to clarify and evaluate some of the hypotheses made by previous investigators as to whether Bain, in the emergency call, said the incriminating material *I shot the/that prick*, or something anodyne, like *I can't breathe*. Following the orthographic transcription in my first report, I carried out an auditory-phonetic and acoustic analysis of parts of the call, on the basis of which the following was demonstrated:

.

- There is strong evidence to support the claim that at least the first part of the disputed portion is actually speech, and not random non-speech noises.
- There is evidence to support the claim the first part of the disputed portion realises /aɪ/. It is reasonable to suppose this signals the pronoun "I", as most investigators actually heard.
- There is strong evidence to support the claim that the second sound of the disputed portion realises /k/, and not /ʃ/.
- There is strong evidence to support the claim that the vowel after the /k/ is not /ɒ/, but /a/.

These points taken together provide very strong support indeed for the defence hypothesis that Bain actually said *I can't* and did not say, as the Crown claims, *I shot*.

In fact I find, like Dr Foulkes, no acoustic or auditory-phonetic evidence at all that he said *I shot*.

8.0 References

T. BROEDERS (1999) 'Some observations on the use of probability scales in forensic identification'. *Forensic Linguistics* 6/2: 228-241.

T. C. DEMPSEY(n.d) Statement.

T. C. DEMPSEY (2007) Transcription of David Bain Ambulance Call.

J. ELLIOTT (2001) 'Auditory and F-pattern variation in Australian *okay*: a forensic investigation'. *Acoustics Australia* 29/1: 37-41.

G. FANT (1960) *Acoustic Theory of Speech Production*. Mouton & Co.

J.P. FRENCH (1990) 'Analytic procedures for the determination of disputed utterances'. In Kniffka (ed.) *Texte zur Theorie und Praxis forensischer Linguistik*. Max Niemayer Verlag. pp. 201-13.

P. FOULKES (2008 April) Transcription of 111 call.

P. FOULKES (2008) Supplementary report – DRAFT.

P .FRENCH & P. HARRISON (2008) Report on Examinations of 111 Audio Tape.

B. GUILLEMIN (2008) Expert Witness Report.

B. INNES (2007) Letter & enclosures to Reed QC.

K. JOHNSON (1997) *Acoustic and Auditory Phonetics*. Blackwell.

J. LAVER (1994) *Principles of Phonetics*. CUP.

P. LADEFOGED & I. MADDIESON(1996) *Sounds of the World's Languages*. Blackwell.

M. J. PEARCE (n.d) Statement.

P. ROSE (2002) Forensic Speaker Identification. Taylor & Francis.

P.ROSE (2003) *The Comparison of Forensic Voice Samples*. Thomson Law Book Co.

P.ROSE (2006) 'The Intrinsic Forensic Discriminatory Power of Diphthongs'. In Warren & Watson (eds.) *Proc. 11th Australasian International Conf. Speech Science and Technology*.

P. ROSE, Y. KINOSHITA & T. ALDERMAN (2006) . 'Realistic Extrinsic Forensic Speaker Recognition with the Diphthong /ai/'**.** In Warren & Watson (eds.) *Proc. 11th Australasian International Conf. Speech Science and Technology*.

K. N. STEVENS (1998) *Acoustic Phonetics*. MIT Press.

P. STREVENS(1960) 'Spectra of Fricative Noise in Human Speech'. *Language and Speech* 3: 32-49.

VIDEO NOTES (2008) Notes of evidence taken before the Hon Justice Panckhurst.

D. J. WARD (n.d.) Statement.

Phil Rose
February 22nd 2009

Ph.D. (Cambridge)
M.A., B.A. Hons. *First Class* (Manchester)
Dip. I.P.A. *First Class* (London).

Reader in Phonetics & Chinese Linguistics
Australian National University.
British Academy Visiting Professor
Joseph Bell Centre for Forensic Statistics and Legal Reasoning, University of Edinburgh.

Chairman, Forensic Speech Science Committee
Australasian Speech Science and Technology Association.

Former Member of Council, *International Phonetics Association*.
Member, *International Association of Forensic Phonetics and Acoustics*.
Member, *International Phonetics Association*

email: **philip.rose@anu.edu.au**