

La cualidad de voz en fonética judicial*

Juana Gil & Eugenia San Segundo
Laboratorio de Fonética. Consejo Superior de Investigaciones Científicas

Resumen

La cualidad de voz es un parámetro que reviste gran importancia para la investigación y la práctica de la fonética judicial, pese a no haber sido aún suficientemente estudiado, especialmente por lo respecta al panorama bibliográfico español. Ni su definición ni su caracterización en términos perceptivos o acústicos están completamente claras. Por esa razón, en este trabajo se repasa el desarrollo del concepto, los análisis a que ha dado lugar y las interrogantes que todavía suscita entre los expertos. Finalmente, se avanzan algunas recomendaciones sobre el empleo de la cualidad de voz como parámetro de comparación de voz, dirigidas a aquellos profesionales que deban realizar informes periciales con fines forenses.

Palabras clave: Fonética judicial, cualidad de voz, ajustes fonoarticulatorios, LTAS, identificación del locutor.

0. Introducción

No creemos estar realizando una generalización incorrecta al afirmar que todos los hablantes poseen un conocimiento intuitivo de lo que, en fonética, se denomina *cualidad de voz*, y que en términos no especializados podría parafrasearse como “el modo en que las personas suenan”, el efecto perceptivo de sus voces, mediante el cual podemos diferenciarlas entre sí, y merced al cual, muy frecuentemente, las identificamos.

Son cientos los términos impresionistas con los que, en todas las lenguas, los hablantes describimos las voces que escuchamos. Si eligiéramos al azar un conjunto de algunas obras literarias, comprobaríamos cuán elevada es la posibilidad de encontrar en ellas caracterizaciones de voces a través del empleo de atributos de muy diverso origen: calificaciones de origen estético (una voz bonita, anodina), emocional (una voz cordial, angustiada), actitudinal (una voz relajada, juvenil), discursivo (una voz entrecortada), visual o táctil (una voz aterciopelada, áspera), fisiológico (una voz nasal, vibrante), etc., todos ellos son atributos que se repiten y se combinan entre sí una y otra vez con el fin de transmitir al lector una idea lo más aproximada posible de la cualidad de voz de la que se quiere dotar al personaje en cuestión¹. Lo interesante es que el lector reproduce mentalmente esa cualidad sin mayores dificultades.

Hablamos, por tanto, de una noción implícitamente compartida por todos los seres humanos, y que, no obstante, es extremadamente difícil de abordar desde un punto de vista científico por las razones que a continuación se explicarán. Su propia naturaleza,

* Este trabajo se ha realizado en el marco del proyecto de investigación *CIVIL: Cualidad Individual de la Voz e Identificación del Locutor*, desarrollado en el Laboratorio de Fonética del CSIC y financiado por el Ministerio de Innovación y Ciencia (Ref.: FFI2010-21690-CO2-01), y asimismo gracias a una beca FPU-Programa Nacional de Formación de Profesorado Universitario (BOE 11.07.2009) concedida a Eugenia San Segundo.

con todo, la convierte en una variable con grandes posibilidades de rendimiento en muchos ámbitos, en particular en el que aquí nos interesa, la fonética judicial, en la medida en que, de modo complejo pero indiscutible, contribuye al reconocimiento por parte de los oyentes de las voces familiares, y a la caracterización y, en su caso, identificación de las no familiares. En las páginas que siguen, trataremos de presentar las características fundamentales del concepto, los problemas que plantea su tratamiento y las soluciones que ya se han avanzado para algunos de ellos.

1. La noción de cualidad de voz

Como es sabido, los recursos fonéticos que el hablante utiliza para plasmar su intención comunicativa se han dividido tradicionalmente en torno a dos niveles² fundamentales: el de carácter segmental y el de tipo suprasegmental. Pese a reconocer los problemas de definición que conlleva esta división (como la exclusión del tono léxico del ámbito suprasegmental), esta clasificación ha sido adoptada por muchísimos autores porque representa una división fundamental del *spoken medium*. No obstante, en la realidad acústica, estos dos niveles no pueden mantenerse separados ni pueden llegar a constituir producciones aisladas e independientes. La representación fonética se genera justamente a partir de la integración de los dos.

Sin embargo, fue Abercrombie (1967) quien señaló que la representación fonética en realidad tiene tres *inputs* o tres “ámbitos” diferenciables. Los dos primeros ya mencionados (el segmental y el suprasegmental), y un tercero que comprendería los “rasgos de cualidad de voz”:

The term voice quality refers to those characteristics which are present more or less *all the time* that a person is talking: it is a quasi-permanent quality running through all the sound that issues from his mouth. (p. 91) [cursiva nuestra].

De acuerdo con Abercrombie, tales rasgos de cualidad de voz a su vez abarcan *a*) aquellos que están fuera del control del hablante (por el hecho de estar determinados por algún aspecto de su configuración vocal) y *b*) aquellos otros sujetos al control voluntario del locutor. Esta clasificación se corresponde *grosso modo* con la distinción entre rasgos “intrínsecos” y “extrínsecos” realizada posteriormente por Laver (1976, p. 57) aunque este último autor especifica que los “extrínsecos” están bajo el control del hablante “de forma consciente *o no*”.

Nolan (1983), en un trabajo fundamental al que luego volveremos, destaca dos aspectos importantes en relación con esta noción. Por un lado, comenta que “por desgracia, el término ‘cualidad de voz’ se utiliza con muchos significados diferentes” (p. 33) –y remite a Crystal (1969, p. 99 y ss.) para una discusión pormenorizada con respecto a este asunto³–, y, por otro lado, destaca que el término también se utiliza, sin ninguna implicación de “casi permanencia”, para referirse específicamente al modo de vibración de las cuerdas vocales. Es decir que la denominación “cualidad de voz” es ambivalente, puesto que puede referirse a las características permanentes de la voz de una persona o circunscribirse al tipo de fonación (*phonation type*) con el que se pronuncia un enunciado concreto. Este apunte de Nolan se ha visto corroborado con el paso de los años y actualmente el empleo de la expresión “cualidad de voz” sigue siendo ambiguo, en cuanto que puede referirse más restringidamente a las diversas modalidades de fonación que puntualmente puedan adoptarse, o, de manera más global, a la cualidad perceptiva general

que “tiñe” todas las producciones vocales de un hablante. En el presente trabajo se emplea siempre en esta segunda acepción.

En el momento actual los fonetistas ya saben qué factores intervienen en la impresión psicoacústica que asociamos con la cualidad de voz de una persona; otra cosa bien distinta es precisar la contribución que cada uno de esos elementos aporta al resultado global, porque, en primer lugar, esta percepción final es multidimensional y claramente holística o, si se prefiere, gestáltica, es decir, el todo que percibimos es más que la suma de las partes; y porque, en segundo lugar, la importancia perceptiva de los distintos elementos que conforman la cualidad de voz puede variar en función del momento, de la situación o de la condición del oyente (Kreiman y Sidtis, 2011). Aun así, como decimos, puede afirmarse que el timbre de una voz depende básicamente de la acción conjunta de tres grandes grupos de condicionantes:

- a) Condicionantes biológicos, que escapan –hasta un cierto punto⁴– al control del hablante y que son de naturaleza extralingüística (“intrínsecos”, en terminología de Laver, 1976). El sexo del hablante, su edad, su complexión física, su estado de salud o su historial médico determinan en buena medida la cualidad vocal de la persona, de modo que algunos aspectos del sonido de su voz están condicionados por el desarrollo y las propiedades de su anatomía y por la fisiología de su tracto vocal, o bien por las patologías permanentes (por ejemplo, paladar escindido), semipermanentes (un estado hormonal concreto) o transitorias (por ejemplo, un catarro) que pueda sufrir. Muchas veces, en estos últimos casos, la voz se convierte en pieza clave para el diagnóstico médico. Dos ejemplos adicionales de rasgos dependientes de la fisiología del hablante serían los límites superior e inferior del rango en el que puede inscribirse su frecuencia fundamental⁵.
- b) Condicionantes sociales, entre los que se incluyen la lengua, el dialecto o el sociolecto con los que se vincula el hablante. Al adquirir un idioma o una variante geolectal dada, o bien al identificarse con un determinado grupo social o familiar, el locutor aprende, interiorizándolos por imitación, una serie de ajustes articulatorios (en inglés, *articulatory settings*) propios de tal colectividad, que inciden en su cualidad de voz y marcan así su pertenencia al grupo. Se han estudiado mucho, por ejemplo, los ajustes articulatorios típicos de ciertos dialectos del inglés: el de Liverpool (Coadou, 2006, 2007; Knowles, 1978), el de Glasglow (Stuart-Smith, 1999) o el de Norwich (Trudgill, 1974). Sin embargo, apenas se ha iniciado la investigación con respecto al español. Tampoco para nuestra lengua existen estudios sobre la cualidad de voz de ciertos estratos sociales o profesionales, como pudiera ser, por citar ejemplos, la de los sacerdotes, la de los profesionales de los medios de comunicación⁶ o la de las personas pertenecientes a las clases sociales más desfavorecidas.
- c) Condicionantes psicológicos o paralingüísticos, como los rasgos de personalidad, los estados de ánimo permanentes o las emociones transitorias que entrañan cambios en la voz (cf. Laver, 1968). Los experimentos realizados en este sentido para confirmar cuáles son los correlatos vocales de las distintas emociones han brindado resultados bastante coherentes. En la Tabla 1 se adapta un cuadro resumen elaborado por Kreiman y Sidtis (2011) a partir de los estudios de Juslin y Laukka (2001), Murray y Arnott (1993) y Scherer (1986, 2003). Nótese, sin embargo, que todos los cambios recogidos en la tabla se refieren a rasgos

relacionados con la frecuencia, la intensidad o la envolvente espectral. Las asociaciones que atañen a los distintos modos de fonación o a los ajustes supralaríngeos se han analizado mucho menos, y así solo pueden encontrarse observaciones todavía muy dispersas en la bibliografía, basadas en estereotipos, como el que identifica la voz *breathy* con la intimidad o la sensualidad, el que vincula la voz ronca o *harsh* a la agresividad o la autoridad, y el que liga la voz resultado de la “labialización” con la transmisión de afecto.

	Tristeza	Miedo	Enfado	Alegría	Aburrimiento
Frecuencia media	Ligeramente más baja	Mucho más alta	Mucho más alta	Más alta	Más baja o sin cambios
Gama tonal	Más monótona	Más amplia o más estrecha	Mucho más amplia	Mucho más amplia	Más monótona o normal
Contornos entonativos	Inflexiones negativas	Cambios abruptos / inflexiones positivas	Cambios abruptos / inflexiones positivas	Inflexiones positivas suaves	Movimiento escaso
Intensidad	Menor	Normal	Más intensidad	Más intensidad	Menor
Tempo	Ligeramente más lento	Mucho más rápido	Ligeramente más rápido	Más rápido o más lento	Más lento
Espectro	Menor energía en altas frecuencias	Más o menos energía en altas frecuencias	Más energía en altas frecuencias	Más energía en altas frecuencias	Menor energía en altas frecuencias

Tabla 1. Adaptación, elaborada por las autoras, de la **Tabla 9.1** que Kreiman y Sidtis (2011, p. 321) realizan a partir de los estudios de Juslin y Laukka (2001), Murray y Arnott (1993) y Scherer (1986, 2003).

2. Descripción perceptiva de la calidad de voz

Se ha dicho más arriba que no es difícil encontrar descripciones impresionistas y subjetivas de la calidad de voz. Mucho más complicado resulta objetivar esas percepciones y medir la calidad de voz de manera fiable. A estos efectos, en el campo de la fonética se suele emplear el protocolo perceptivo ideado por Laver (1979) –y ampliamente comentado en, entre otras muchas obras mencionables, Laver y Hanson (1981), Ball, Esling y Dickson (2000); Mackenzie Beck (2005) –, conocido como *Vocal Profile Analysis* (VPA de ahora en adelante), que básicamente analiza la voz del hablante disociándola en componentes, entendidos como ajustes musculares de largo alcance que conllevan consecuencias perceptivas y que, juntos, conforman el resultado final, esto es, su calidad a medio-largo plazo. Por ejemplo, la voz de Marilyn Monroe, especialmente en algunas de sus grabaciones más famosas, se caracterizaba por poseer una calidad que, en términos impresionistas, se podría calificar de “sensual”, y en términos más objetivos, conforme al protocolo del que hablamos, se describiría como el resultado de, principalmente, un ajuste *breathy* (con escape de aire) del modo de fonación, permanente o cuasi-permanente en todas las emisiones de la actriz.

El protocolo completo en el cual se precisan los distintos *settings*⁷ que determinan la cualidad de voz aparece reproducido en la Tabla 2. Según puede comprobarse en ella, los ajustes se agrupan en dos grandes secciones: los referidos a las cavidades supralaríngeas y los que conciernen a la cavidad laríngea. En realidad, como explican Laver *et al.* (1981, 270), la interrelación entre los ajustes laríngeos y supralaríngeos es muy fuerte, tanto por lo que respecta a los sistemas de actividad muscular que subyacen en ambos sistemas como por lo que hace a la percepción: los efectos perceptivos de unos y otros pueden verse recíprocamente potenciados o enmascarados. Por tanto, la división sobre la que se articula el protocolo ha de entenderse simplemente como una separación a efectos prácticos que se corresponde bien con la clásica distinción entre rasgos referidos a la fuente y rasgos referidos al filtro.

Enfrentados con una voz cualquiera, los usuarios de este tipo de protocolo descriptivo deciden, en primer lugar, si es “neutra” o “no neutra” en relación con cada uno de los ajustes que se enumeran en la columna de la izquierda. El punto de referencia es la cualidad de voz que resultaría de una disposición general en la cual todo el tracto vocal presentara una tensión moderada; los pliegues vocales vibraran de modo natural, por estar sometidos a un grado de aducción y una tensión longitudinal equilibradas; las cavidades supraglóticas no estuvieran ni expandidas ni constreñidas; y la longitud de todo el tracto también alcanzara un valor intermedio, sin verse aumentada ni disminuida.

Si la voz analizada resulta ser “no neutra” con respecto al modelo de referencia, entonces el siguiente paso consiste en precisar si se enmarca en el rango de lo que puede considerarse una voz “normal” o más bien dentro de la esfera de lo patológico⁸ (*abnormal*, en la tabla). En ambos casos, el analista debe concretar el grado de desviación que la voz presenta con respecto al modelo: la escala del 1 al 3 se considera normal; la escala del 4 al 6 es ya patológica. Algunos ajustes no permiten, por su propia naturaleza, esta gradación; por ejemplo, el ajuste de fonación *falsetto* solo admite la especificación presente / ausente.

Así pues, de acuerdo con el protocolo VPA, la descripción impresionista de una voz como, por ejemplo, “de ultratumba” podría parafrasearse en términos más técnicos afirmando que está marcada con respecto a los ajustes relativos a la *posición del cuerpo de la lengua* (baja) y a la *altura de la laringe* (baja). Además, presentaría *constricción faríngea*, casi con seguridad modos de fonación diferentes del modal (*creaky* o *breathy*), tal vez un cierto grado de *labialización* y, ya desde el punto de vista prosódico, un *rango tonal estrecho* y situado en las bajas frecuencias, y un *tempo* lento.

Vocal Profile Analysis Protocol

Judge: Tape: Sex:

Date of Analysis: Speaker: Age:

I. VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SETTING	SECOND PASS					
	Neutral	Non-neutral			Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial				Lip Rounding/Protrusion						
				Lip Spreading						
				Labiodentalisation						
				Extensive Range						
				Minimised Range						
2. Mandibular				Close Jaw						
				Open Jaw						
				Protruded Jaw						
				Extensive Range						
				Minimised Range						
3. Lingual Tip/Blade				Advanced						
				Retracted						
4. Lingual Body				Fronted Body						
				Backed Body						
				Raised Body						
				Lowered Body						
				Extensive Range						
5. Velopharyngeal				Nasal						
				Audible Nasal Escape						
				Denasal						
6. Pharyngeal				Pharyngeal Constriction						
7. Supralaryngeal Tension				Tense						
				Lax						
B. Laryngeal Features										
8. Laryngeal Tension				Tense						
				Low						
9. Larynx Position				Raised						
				Lowered						
10. Phonation Type				Harshness						
				Whisper(y)						
				Breathiness						
				Creak(y)						
				Falsetto						
			Modal Voice							

Tabla 2. Protocolo del Voice Profile Analysis (Laver *et al.*, 1981 (1991), p. 268)

Este no es el único tipo de protocolo que se ha empleado al intentar objetivar y medir perceptivamente la cualidad de voz, aunque, como se ha dicho, es el más empleado entre los fonetistas. Por el contrario, los foniatras, los logopedas o los otorrinos, que lógicamente deben precisar las propiedades de las voces tratadas de modo consensuado entre todos ellos, se han venido sirviendo preferentemente del protocolo, también basado en la percepción, denominado *GRBAS* (siglas en inglés de *Grado de disfonía; Ronquera o aspereza; Breathiness* o escape de aire; *Astenia* o hipofunción vocal; y *Strain* o tensión), desarrollado por la Sociedad Japonesa de Logopedia y Foniatría (cf. Hirano, 1981) y que es un tipo de análisis factorial que pretende evaluar la voz evitando las redundancias y los solapamientos en las escalas descriptivas subyacentes (escala de afonía, de sibilancia o de soplos, etc.) al reducirlas al pequeño conjunto de factores mencionado⁹. Estos se derivan, pues, de los juicios realizados sobre las escalas inicialmente seleccionadas, y ese es, precisamente, su mayor inconveniente porque, como es lógico, la validez de los factores depende en última instancia de la validez de las escalas a las que se vinculan, que nunca se ha establecido de manera clara (cf. Kreiman *et al.*, 2007).

Otros protocolos empleados en particular para la evaluación clínica de la voz, y, por ello, mencionados en la bibliografía especializada en este campo, se asemejan al *GRBAS*, aunque difieren en el número de parámetros con que operan. Por ejemplo, el denominado *Buffalo III Voice Profile* (Wilson, 1987) maneja 12 diferentes: frecuencia laríngea, tono, intensidad, resonancia nasal, resonancia oral, aporte de aire, hiper- o hipotensión muscular, velocidad, ansiedad vocal, inteligibilidad, esfuerzo de los músculos vocales y valoración global. El *Stockholm Voice Evaluation Approach (SVEA)*, desarrollado en Suecia (cf. Hammarberg, 2000), se articula en torno a 13 parámetros, y el *Consensus Auditory Perceptual Evaluation (CAPE-V)*, creado en 2003 en los Estados Unidos a raíz de un congreso previo de la *American Speech-Language-Hearing Association*, contiene seis parámetros perceptivos primarios (ronquera, escape de aire, tensión, tono, intensidad y grado de patología general) que son evaluados utilizando escalas analógicas visuales de 100 mm y especificados como presentes sistemática o intermitentemente en las producciones analizadas: vocales aisladas, oraciones y un fragmento de habla continua (cf. por ejemplo, para más detalles, Kempster *et al.*, 2009).

Esta clase de descripciones perceptivas, componenciales o dimensionales, de la cualidad de voz, a pesar de ser muy útiles, sobre todo en fonética clínica, y de haber recibido también amplio apoyo entre los expertos del ámbito propiamente lingüístico o incluso judicial (cf., por ejemplo, sobre el *VPA*, Camargo y Madureira (2008), French (1994), Jessen (2008), Nolan (1983) y la bibliografía en estas obras recogida), presenta algunos puntos discutibles. Son sistemas, en primer lugar, que dependen en gran medida de la habilidad inherente de sus usuarios para percibir y discernir de un modo fiable los detalles de una voz (Kreiman *et al.*, 1992). En segundo lugar, como afirma Nolan (2005), en el caso concreto del *VPA*, el marco ideado por Laver y sus colaboradores no está diseñado para capturar los aspectos menos lingüísticos del habla de las personas, el “detalle fonético” relevante. Es decir, puesto que los tractos vocales son isomórficos, pero no absolutamente idénticos, las pequeñas diferencias de tamaño o forma que presentan harán que dos personas suenen diferentes aunque “elijan” las mismas opciones articulatorias. Esto se explica porque el modelo está basado en el hablante y en sus capacidades articulatorias o fonatorias, sin tomar en cuenta en ningún momento la perspectiva del procesamiento por parte del oyente: no resulta evidente el peso de cada uno de los ajustes que conforman el protocolo descriptivo en la impresión psicoacústica que finalmente el oyente se crea, ni cómo todos ellos se relacionan entre sí (cf. Kreiman

et al., 2007). En la medida en que la percepción de la cualidad de voz es el resultado de la interacción entre el hablante y el oyente, puede además darse el caso de que los mismos mecanismos articulatorios o fisiológicos subyacentes originen sensaciones / percepciones distintas en oyentes inscritos en situaciones asimismo diversas, o que estos primen perceptivamente diferentes aspectos de la voz dependiendo de las circunstancias. Kreiman y Sidtis (2011) sintetizan esta crítica en pocas palabras: “Featural systems (...) model quality as if it inheres in voices, without also accounting for such listener-dependent factors as attention, experience, and response bias” (p. 19).

En este sentido, el sistema de descripción perceptiva de la voz basado en el escalado multidimensional (*EMD*) es algo diferente. Este es un procedimiento para representar las percepciones que los oyentes tienen de los estímulos –las voces– en un diagrama visual que se conoce como “mapa perceptivo”. Los oyentes que actúan como jueces comparan pares de voces, sin hacer uso, como en los protocolos anteriores, de escalas para medir atributos concretos (que ya no son necesarias), y juzgan directamente su grado de semejanza: cuanto más similares sean las voces, más cerca estarán situadas en el espacio que conforma el mapa perceptivo. Comprobando después cómo se han distribuido los distintos estímulos en tal espacio y el modo en que se correlaciona esa distribución con las características acústicas que dichos estímulos poseen, pueden deducirse las dimensiones (frecuencia fundamental, intensidad, frecuencias formánticas, etc.) sobre las que el oyente se apoya preferentemente para juzgar el grado de proximidad entre las voces, y, en última instancia, pues, el grado en que esas propiedades particulares definen y determinan la cualidad de voz general.

Este tipo de estudios también presenta problemas porque los resultados que ofrecen no son siempre coincidentes (Baumann y Belin, 2008), ya sea porque las características de los estímulos con los que se ha trabajado en los experimentos son distintas (palabras aisladas frente a oraciones, por ejemplo), ya sea porque los oyentes parecen aplicar diferentes estrategias perceptivas dependiendo de los casos, por lo que las conclusiones de los diversos experimentos son lógicamente divergentes (cf. Gelfer, 1993).

Así pues, es evidente, por lo hasta ahora expuesto, que la descripción perceptiva de la cualidad de voz no es un asunto fácil y que siempre se enfrenta a la objeción que alude a la falta de fiabilidad o coherencia en los juicios de los evaluadores, sean estos los profesionales que conciben y utilizan protocolos escalares, sean los oyentes participantes en los estudios que valoran las voces a partir de muy variables dimensiones perceptivas. Es por todo ello por lo que algunos autores han tratado de encontrar correlatos más fiables de las distintas cualidades de voz en el análisis acústico, según se explica en el siguiente apartado.

3. Correlatos acústicos de la cualidad de voz

It will be worth developing and improving this work [*the work of specifying the acoustic correlates of an auditory phonetic framework for classifying voice qualities*] since, from the point of view of speaker identification it provides an approach to the problem of classifying voices alternative, and complementary, to the more usual one of picking readily measurable acoustic features and investigating, in a relatively unguided way, how these features vary among a population of speakers. (Nolan, 1983, p. 108).

Efectivamente, según afirma Nolan en la cita anterior, conviene, y mucho, especificar los correlatos acústicos de los marcos fonéticos de tipo auditivo que se

emplean para clasificar las diversas cualidades de voz. De ese modo, a partir de la información que el análisis perceptivo proporciona, la exploración acústica está más claramente dirigida y, al tiempo, avala las impresiones subjetivas de los analistas. De la revisión de la bibliografía a este respecto –que, por cierto, no es muy abundante en términos relativos– se pueden extraer las correspondencias que se resumen, de manera muy sucinta, a continuación (pueden consultarse exposiciones mucho más detalladas en Laver (1994), en Camargo (2002) y en Keller (2004), por ejemplo):

1. Correlatos acústicos de los ajustes longitudinales (esto es, aquellos que alargan o acortan el tracto oral: en general, el alargamiento tiende a bajar las frecuencias formánticas, y el acortamiento, a elevarlas)

Laringe elevada: aumento de la frecuencia fundamental y aumento de la frecuencia de todos los formantes, especialmente el primero.

Laringe baja: disminución de la frecuencia fundamental (puede llegar a producirse un modo de fonación *breathy*) y disminución de la frecuencia de todos los formantes, especialmente del primero.

Protrusión labial: disminuye la frecuencia de todos los formantes, especialmente la de los altos. Ya Lindblom y Sundberg (1971) señalaron este efecto, en particular por lo que respecta al F2 y al F3.

Labi dentalización: este ajuste, que suele acompañar un tipo de habla risueña, con sonrisas, acorta levemente la extensión del tracto vocal, y produce un descenso de la frecuencia de todos los formantes.

2. Correlatos acústicos de los ajustes latitudinales (esto es, aquellos que expanden o contraen el tracto vocal horizontalmente):

Ajustes labiales: disminuyen o aumentan el espacio entre los labios; sus consecuencias acústicas son muy variables y no están claramente determinadas.

Desplazamiento lingual antero-posterior: cuanto más adelante en la cavidad oral esté situada la lengua, mayor será la distancia entre el primer y el segundo formante; cuanto más atrás se desplace, más cercanos aparecerán estos dos formantes.

Faringe constreñida: consecuencias acústicas no determinadas, aunque parece producirse en muchos casos una elevación del F1 y un descenso del F2.

Faringe expandida: este ajuste provoca la disminución de la frecuencia del primer formante.

Mandíbula abierta: a mayor grado de abertura, mayor elevación del F1.

Mandíbula cerrada: a mayor grado de cierre, mayor descenso del F1 y del F2.

3. Correlatos acústicos de los ajustes velofaríngeos (esto es, aquellos que determinan el grado de nasalidad que caracteriza a largo plazo a la secuencia hablada):

Nasalidad excesiva: la hipernasalidad lleva asociada la aparición de uno o más formantes nasales (en altas y bajas frecuencias) junto con más antirresonancias y una disminución generalizada de energía.

Denasalización: se caracteriza por ligeras alteraciones en el F1 y el F2, no así en el F3 ni el F4.

4. Correlatos acústicos de los ajustes laríngeos (esto es, aquellos que dan lugar a las distintas modalidades de fonación):

Falsetto: se corresponde con unos valores de la frecuencia fundamental dos o tres veces más altos que los que presenta la voz modal, así como con la existencia de pocos armónicos y con un “tilt” (o declinación) espectral muy pronunciado.

Murmullo: este modo de fonación se caracteriza por presentar una gran cantidad de ruido, en todas las frecuencias, pero especialmente en las altas.

Creak: responde a valores muy bajos de la frecuencia fundamental, con unos pulsos glotales muy espaciados entre sí.

Voz ronca o áspera: la frecuencia fundamental presenta aperiodicidad y, asimismo, microoscilaciones (*jitter*).

Breathiness: acústicamente, la voz *breathy* o soplada entraña la presencia de una cantidad moderada de ruido y una rápida disminución de la energía en las altas frecuencias.

5. Correlatos acústicos de la tensión general del tracto vocal: una mayor amplitud general en las voces tensas que en las relajadas, con mayor energía en los armónicos superiores a los 1000 Hz, y menor amortiguación en los inferiores, debido a la mayor impedancia o resistencia de las paredes rígidas del tracto a la absorción de la energía acústica (cf. Gil, 2007 y la bibliografía allí citada). Por tanto, existe un descenso menos pronunciado, más suave, del *tilt* espectral.

Obsérvese que las correspondencias recogidas arribas se refieren a los parámetros proporcionados por un análisis perceptivo de tipo *VPA*. Un estudio muy interesante con el que se ha tratado de establecer la correlación entre los parámetros subjetivos del otro protocolo perceptivo más empleado, el *GRBAS*, y 19 parámetros acústicos objetivos (tales como el *jitter* absoluto y porcentual; la variación de la frecuencia fundamental, el *shimmer* en decibelios, la ratio armonicidad-ruido, etc.) es el de Bhuta *et al.* (2004). Las conclusiones de este trabajo apuntan, por ejemplo, a que tanto una cualidad *breathy* (B, de acuerdo con la clasificación *GRBAS*) como una asténica (A) se corresponden bien con la ratio mayor entre la energía de los armónicos de menor frecuencia con respecto a los de frecuencia más alta; y, asimismo, la proporción del ruido con respecto a la armonicidad (NHR) resulta ser el índice que mejor se correlaciona con la cualidad de voz “ronca” o “áspera” (R). Todavía son necesarios muchos más estudios en esta línea.

No puede concluirse la revisión de los correlatos acústicos de la cualidad de voz sin hacer mención al procedimiento de análisis acústico con el que más a menudo se la ha vinculado, el denominado *LTAS*, al que nos referiremos en el siguiente subapartado.

4. El empleo del *LTAS*

Puesto que los componentes de la cualidad de voz son tendencias fonatorias o articulatorias a las que los hablantes se ajustan a largo plazo, el análisis acústico idóneo pareciera ser aquel que estudie las características del habla también con un alcance

amplio, es decir, desde una perspectiva dinámica y no puramente estática o segmental. Por esta razón, el denominado *Long Term Average Spectrum*, o, en español, Espectro Medio a Largo Plazo (de ahora en adelante LTAS, como es conocido por sus siglas en inglés) ha sido un tipo de análisis que se ha venido empleando en el ámbito de los estudios de la calidad de voz desde hace tiempo (cf., por ejemplo, Löfqvist, 1986; Löfqvist y Mandersson, 1987).

El LTAS es el espectro promediado de la amplitud o intensidad a lo largo de la cadena hablada en una gama de frecuencias seleccionada. Se obtiene haciendo simplemente la media de un gran número de análisis espectrales “instantáneos” de los fragmentos de una muestra de habla que necesariamente ha de ser extensa (“Long-term average spectra require around 15 sec of continuous speech to settle down, perhaps longer if the content of different samples is not the same”, se lee en Nolan, 1983, p. 143)¹⁰. La idea que subyace en su aplicación para el estudio de la calidad de voz es que los diferentes ajustes articulatorios o fonatorios que producen las diversas cualidades de voz se han de ver reflejados en las correspondientes divergencias entre los perfiles espectrales medios. En la Figura 1, pueden verse superpuestos los LTAS obtenidos, mediante el programa PRAAT, a partir del habla de las dos autoras del presente trabajo, quienes leyeron el mismo texto durante 17 segundos.

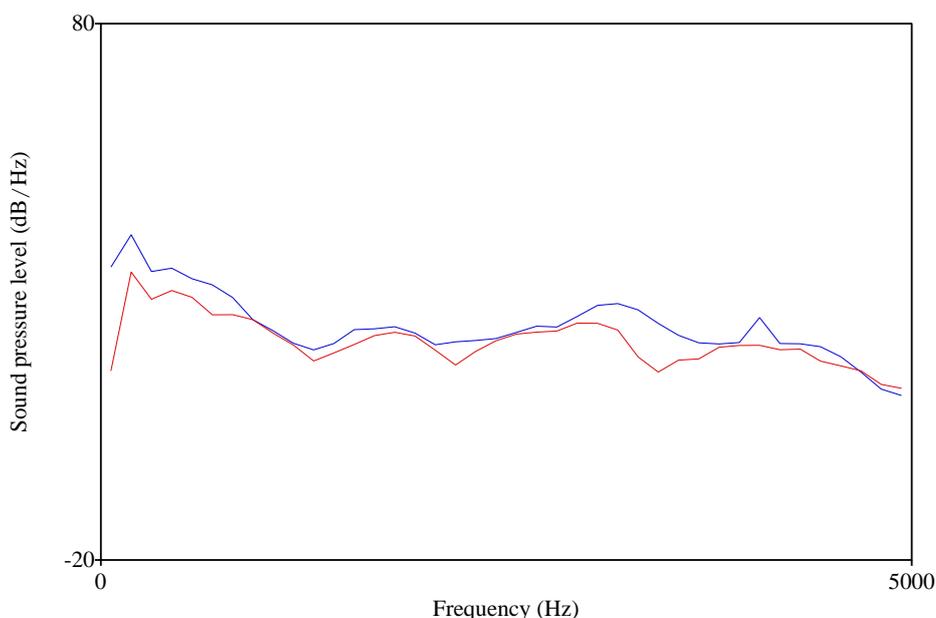


Figura 1. Representación esquemática del LTAS de dos voces femeninas leyendo el mismo texto.

Es fácil comprobar que la voz de una de las dos hablantes (en azul) presenta, en general, más intensidad (eje vertical) a lo largo de toda la lectura, pero especialmente en las bajas y en las altas frecuencias (eje horizontal), que la voz de la otra locutora (en rojo).

Las primeras referencias al empleo del LTAS con fines judiciales se remontan a principios de los años 70, con los trabajos clásicos de Zalewski, Majewski y Hollien (1974), Doherty (1976), Doherty y Hollien (1978) y Wendler *et al.* (1980, 1986). En el primero de esta serie de artículos, los autores ya defienden que el análisis acústico de la señal de voz a través del LTAS es un método fiable para la identificación del locutor –al menos, en condiciones de laboratorio (p. 24)– y proponen algunas técnicas estadísticas

independientes aplicables cuando el análisis de voz se lleve a cabo mediante un LTAS. Se describe la técnica conocida como “correlación cruzada” y se compara con el enfoque estadístico basado en distancias euclidianas, para concluir que ambos métodos permiten obtener valores elevados de identificación de hablantes; además, se sugiere la posibilidad de usar ambas técnicas conjuntamente. Poco después, en Doherty (1976) y en Doherty y Hollien (1978), se compararon tres tipos de medidas: las espectrales –las que se basan en el espectro a largo plazo–, las temporales y las relacionadas con la f_0 . Tras comprobar la capacidad identificativa de cada uno de estos tres tipos de parámetros, mediante un análisis discriminante que permite clasificar a los hablantes, se observó que el espectro a largo plazo era el mejor indicador para la identificación de un locutor, incluso cuando se utilizaba de forma aislada. No obstante, los resultados también mostraron que los niveles de identificación mejoran cuando se combinan los tres tipos de medidas antes enumeradas¹¹.

Con todo, es Pittam (1987b) quien, unos años después, vincula claramente el LTAS con los análisis de la cualidad de voz, considerándolo un marcador no solo del estrato social, sino también de la personalidad del hablante, y juzgándolo como el parámetro más adecuado, en el ámbito acústico, para dar cuenta de la naturaleza no localizada, sino “de largo alcance”, que la cualidad de voz posee (cf., asimismo, Pittam, 1987a y Pittam y Millar, 1988). El autor realiza una pormenorizada revisión histórica del empleo del LTAS por los diversos especialistas y llama la atención sobre un aspecto fundamental en fonética judicial, como es la influencia del lapso temporal transcurrido entre las muestras de habla que se deben comparar: “If voice quality is to be used in speaker recognition tasks, the posible effect of such a time variable needs to be clarified” (p. 5). El paso del tiempo puede, en efecto, producir un cierto grado de variación intralocutor, que afectará a su LTAS. De hecho, Löfqvist y Mandersson (1987) encontraron muchas divergencias entre los espectros a largo plazo de los mismos hablantes grabados en distintos momentos del día¹².

A investigar también la influencia de la variación en el LTAS dedicaron sus trabajos Harmegnies y Landercy (1988), Molina de Figuereido y Bernales (1999) y Noderberg y Sundberg (2004), por ejemplo. Los primeros diferencian entre hablantes “coherentes” (aquellos cuyos LTAS, extraídos de frases distintas, son muy similares) y hablantes “no coherentes” (aquellos que exhiben grandes diferencias en los LTAS extraídos de frases no coincidentes). Desde el punto de vista del reconocimiento del locutor, estos resultados pueden tener consecuencias importantes, como los autores defienden:

[...] in speaker recognition tasks based upon the LTS, the error probability associated with the recognition of a given speaker is likely to depend on the speaker's LTS coherence. Furthermore, if one accepts that the LTS coherence in fact reflects the voice coherence, this remark extends to all kinds of speaker recognition tasks. (Harmegnies y Landercy, 1988, p. 85).

En cuanto a Molina de Figuereido y Bernales (1999) el tipo de variación del LTAS que les preocupa es la relacionada con la velocidad de elocución del habla, por lo que pueda perjudicar a su identificación. Así, en el experimento que realizaron, para cada hablante (de un total de 29) se extrajeron dos LTAS, a partir de dos lecturas de un mismo texto: en velocidad normal y en velocidad rápida. Con el fin de reducir la información derivada del espectro a sus características más globales, se calcularon los cuatro primeros momentos espectrales, tratando cada LTAS como una “distribución virtual de probabilidades” (cf. Forrest *et al.*, 1988).

Los dos primeros momentos son la media y la varianza, o sea, expresan la concentración media de energía en el ELT [LTAS] y la variabilidad a lo largo del ELT. El tercero y cuarto momentos están asociados respectivamente al grado de asimetría y al grado de achatamiento del ELT. (Molina de Figueroa y Bernal, 1999, p.3).

Los resultados de la comparación entre hablantes muestran, en primer lugar, que en algunas zonas del LTAS se observa mayor variabilidad interlocutor, en concreto en las regiones del tercer y cuarto formante (2500-3500 Hz). No obstante, los autores señalan que este método (esto es, la representación de los LTAS por medio de los cuatro primeros momentos estadísticos) conlleva ciertos errores, por lo que esa reducción de información del espectro no debería aplicarse con fines forenses. En segundo lugar, un análisis de *cluster* realizado en el estudio muestra que los rangos frecuenciales que identifican mejor a un hablante son, por un lado, el rango de 500 a 1000 Hz, y por otro, el de 1000-2000 Hz. La información espectral por encima de los 4000 Hz parece poco relevante. En tercer lugar, se comprueba que todos estos resultados no se verían afectados por la velocidad de elocución de los hablantes, que, en este caso concreto, eran hablantes nativos de español chileno.

Finalmente, Nordenberg y Sundberg (2004) centraron su atención en la posible variación generada por las modificaciones en la intensidad del habla. Estos autores analizaron las producciones de 15 hombres y 16 mujeres leyendo el mismo texto 7 veces con distintas intensidades vocales en cada caso, y concluyeron que, en el rango frecuencial que se extiende hasta los 4000 Hz, el espectro mostraba una correlación lineal con el nivel de intensidad general empleado por cada sujeto (si bien las altas frecuencias siempre se veían algo más incrementadas que las bajas, y, desde luego, más allá de 4000 Hz, la variación individual era muy grande). Los contornos del LTAS del hablante con cualquier intensidad de habla, pues, podrían predecirse, en esa gama de frecuencias inferior a 4000 Hz, con una precisión de 2-3 dB arriba o abajo.

Son muchísimos más los títulos que podrían citarse en los que distintos autores han examinado las ventajas y desventajas de este parámetro (cf. los recogidos en Gil, 2012). El hecho de que el LTAS no solo se haya empleado en los estudios relacionados con la identificación del hablante, sino también en el análisis de las voces patológicas, en el de las diferencias vocales asociadas a la edad, el sexo o el idioma, en el de la transmisión de la emoción en el habla, etc., ha sido sin duda la causa de que se haya generado tan gran cantidad de bibliografía. A lo largo de los años, pues, se han ido señalando las ventajas de este parámetro, pero también sus inconvenientes. Además del derivado de la variación intralocutor al que ya hemos hecho mención, los otros tres problemas más importantes pueden resumirse en los siguientes puntos:

1. Las diferencias que los LTAS presentan no siempre parecen corresponderse bien con las realmente existentes entre hablantes o entre las muestras de un mismo hablante. Kitzing (1986) comprobó que a veces algunas voces perceptivamente muy distintas se correspondían con LTAS que solo diferían en un 3%-6%. El propio Nolan (1983), aunque reconoce la validez del espectro para estudiar los ajustes fonatorios, cuestiona su utilidad para discriminar entre los articulatorios.
2. El LTAS no toma en consideración valores temporales que son muy importantes para la definición de una cualidad de voz dada (Keller, 2004). Por ejemplo, en la percepción del modo de fonación *creak*, lo que resulta determinante no es el valor que alcance la frecuencia fundamental (que puede oscilar desde 50 a 90 Hz), sino el espacio temporal que exista entre los pulsos glotales, que es una información

que el LTAS no brinda, como tampoco ofrece la relativa al *jitter* o la duración y evolución temporal de las transiciones entre vocales y consonantes.

3. El problema más evidente, sin embargo, es el referido a la influencia en el LTAS de la información acústica puramente segmental, que no se deslinda de la relativa a la voz. Por tanto, resulta complicado, por ejemplo, comparar muestras de habla con contenidos fonéticos distintos. Algunos autores, como Klasmeyer (2000), citado por Keller (2004), y este mismo autor, proponen realizar siempre los LTAS sobre los núcleos vocálicos, de manera que se neutralice al máximo la interferencia del componente segmental (v. también Mendoza *et al.*, 1996, que siguen un procedimiento similar a este respecto).

En definitiva, el LTAS es un parámetro más, aprovechable para cualquier estudio acústico de la cualidad de voz. A la postre, no obstante, es el fonetista experto el que, a la vista de todo lo aquí expuesto, habrá de valorar en cada caso concreto no solo si conviene emplearlo, sino también cómo plantear su uso y cómo calibrar los datos que ofrezca.

5. La cualidad de voz en fonética aplicada: su tratamiento en fonética judicial

El concepto de cualidad de voz se relaciona, en diferente grado, con varias de las aplicaciones de las que es susceptible la fonética. Ya se mencionó anteriormente que la descripción de la voz es un paso inexcusable en el tratamiento de algunas patologías del habla; asimismo, en el campo de la enseñanza de lenguas, la noción de cualidad de voz propia de un idioma, derivada en buena medida de su base de articulación característica, va despertando progresivamente el interés de los especialistas. Por último, en la esfera de la fonética judicial, que es la que aquí nos concierne, constituye en realidad un reto con el que se han enfrentado, antes o después, los más importantes expertos, a los que aludiremos a continuación.

Uno de los que más aportaciones ha realizado en relación con esta noción ha sido Francis Nolan, en la Universidad de Cambridge. La primera referencia a la cualidad de voz la encontramos en el segundo capítulo de su libro más conocido (*The Phonetic Bases of Speaker Recognition*, 1983), al que hicimos ya mención más arriba y en el que el concepto de *voice quality* va a ocupar un lugar destacado. Nolan (1983) presenta un modelo de la comunicación del habla, derivado básicamente de las ideas de Abercrombie (1967), que pretende proporcionar un marco teórico completo de las fuentes de variación inter- e intralocutor, o, al menos, un marco más completo y fiable, en sus propias palabras, que el que estaba a disposición de los expertos en reconocimiento de hablantes hasta ese momento. En el cuarto capítulo de Nolan (1983) es donde más extensamente se desarrolla la idea de que explorar los correlatos acústicos de la cualidad de voz a largo plazo (Nolan habla de *long term voice qualities*, en plural) es importante para el reconocimiento de hablantes. Téngase en cuenta que, en el momento en que se publica su obra, el área de la “cualidad a largo plazo” apenas había sido planteada como objeto de estudio en fonética. Es a partir de los años 80 cuando algunos científicos comienzan la tarea de desarrollar para ella un marco descriptivo semejante al ya existente desde hacía tiempo para la cualidad segmental del habla. Nolan, pues, es pionero al considerar que esta es un área especialmente relevante en fonética judicial porque permite demostrar que los problemas a los que se enfrenta esta disciplina se originan fundamentalmente por la plasticidad del aparato vocal. En otras palabras, los primeros experimentos llevados a cabo en este ámbito sirven para llamar la atención sobre la amplia variedad de ajustes de largo alcance

que puede configurar un único tracto vocal, y sobre la importancia de crear un marco sistemático dentro del cual sea posible investigar las consecuencias acústicas de dicha variación.

Harry Hollien, por su parte, en su bien conocida obra *The acoustics of crime* (1990) trata la cualidad de voz en el capítulo dedicado a la parte acústica. Puede sorprender que no sea en el subapartado dedicado a la fonética, sino en el consagrado a la acústica, en el que encontramos la única referencia al concepto de cualidad de voz, sobre todo si consideramos lo poco que se conoce sobre sus correlatos acústicos (cf. *supra*, § 3). En cualquier caso, el espacio dedicado a explicar qué se entiende por cualidad de voz es más bien breve. Especialmente destacable, teniendo en cuenta lo ya explicado aquí, es el hecho de que Hollien (1990) señale que el término “cualidad de voz” se considera a menudo un “cajón de sastre”, ya que incluye un gran número de aspectos y dimensiones; o también una palabra comodín (en inglés la llama *wastebasket*, vocablo que define como “nombre que se usa cuando ningún otro parece apropiado”).

Mucho más explícito es Hermann J. Künzel, otro referente indiscutible en el campo de la fonética judicial. Este autor incluye decididamente la cualidad de voz entre las características que, junto con rasgos como la entonación, las pausas o el *tempo* (entre otros), permiten caracterizar a un hablante. En su libro *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung* (1994), Künzel afirma claramente que en la cualidad de voz se manifiestan propiedades específicas del hablante, y menciona en primer plano, como ejemplo interesante, el ajuste laríngeo conocido como *creaky*, rasgo de cualidad de voz que desempeñó un papel importante en un caso forense real al que el autor alude: se trata del secuestro y posterior asesinato del ministro de trabajo alemán Hanns Martin Schleyer (cf. p. 85). En el capítulo cuatro de su libro, además, donde aborda el tema de la elaboración de un peritaje fonético para la identificación de hablantes, Künzel recoge en una tabla los rasgos más relevantes que pueden caracterizar a un locutor. El autor distingue dos apartados fundamentales que debe contener un informe pericial de este tipo:

- 1) Aspectos relacionados directamente con la voz.
- 2) Aspectos relacionados con el habla y la manera de hablar (dialecto, idiolecto¹³, sociolecto, velocidad de locución y tipo de respiración).

Pues bien, en el primer grupo Künzel incluye ya la *cualidad de voz*, en tercer lugar, solo después de la frecuencia fundamental media y de su desviación estándar.

De igual modo, otro gran especialista en fonética forense, Peter French, concede gran importancia a la cualidad de voz como parámetro valioso para la comparación de voces e identificación del hablante. En su obra *An overview of forensic phonetics with particular reference to speaker identification* (1994) se ofrece fundamentalmente una introducción a la disciplina (cf. igualmente la más reciente revisión del campo en French y Stevens, en prensa) en la que el autor explica someramente en qué consiste el enfoque combinado perceptivo-acústico, clásico en el área. French distingue tres grandes subapartados para explicar este enfoque combinado. Por un lado, se mencionan los análisis fonético-perceptivos; en segundo lugar se alude a los análisis fonético-acústicos; y, finalmente, se hace referencia a los análisis de la frecuencia fundamental. La mención a la “cualidad de voz” la encontramos ya en la primera sección. En la Tabla 3 recogemos esta clasificación, con los distintos parámetros que incluye cada apartado.

Enfoque combinado perceptivo-acústico		
Análisis fonético-perceptivo	Análisis fonético-acústico	Análisis de la frecuencia fundamental
Parámetros segmentales	Análisis espectrográfico	F0 media
Tono, ritmo y fluidez	Rasgos de frecuencias y de energía espectral	Puntos de transición hacia modo de fonación “creak”
Cualidad de voz	Rasgos temporales	Entonación Estimaciones del “jitter”

Tabla 3. Componentes del enfoque combinado perceptivo-acústico, según French (1994)

Sorprende en esta clasificación que la cualidad de voz se inserte en un apartado diferente a los parámetros relacionados con la frecuencia fundamental, dado que son aspectos inextricablemente vinculados (de hecho, Laver consideraba que el análisis de la frecuencia y sus variaciones, de la intensidad y sus variaciones, y de la estabilidad de la voz constituía una segunda parte, complementaria, del VPA, cf. Laver *et al.*, 1981). Es especialmente notable el tratamiento separado del concepto de cualidad de voz con respecto a parámetros como el *jitter*, y también con respecto al modo de fonación *creak*.

En lo que concierne a la noción de cualidad de voz, French la define como “the combination of resonances upon which the speech signal is ‘overlaid’” (p. 175) y la considera una dimensión más susceptible de análisis perceptivo que de análisis acústico. Se menciona la obra de Laver (1980) como paradigma de la evaluación auditiva de la voz (cf. *supra*, § 2) y, además, se hace referencia a un método para comparar muestras de habla conocido como *composite recording*. Esta técnica consiste en alternar fragmentos de habla extraídos de la grabación dubitada con fragmentos de la grabación indubitada. Aunque este tipo de grabaciones híbridas son de utilidad principalmente para el analista, también pueden reproducirse ante el juez, bien para mostrar el método de trabajo seguido por el experto en la comparación de voces, bien para ilustrar las diferencias o semejanzas encontradas en ellas. French (1994) concluye el apartado mencionando que existen opositores al método que acaba de describir, aunque, a su juicio, se da una clara analogía entre este método que se sirve del tipo de grabación llamada híbrida y el empleado por los expertos grafólogos cuando muestran en un juicio las firmas dubitadas e indubitadas de forma yuxtapuesta, con el fin de demostrar en qué aspectos se asemejan o en cuáles se diferencian.

La estela marcada por todas estas obras se ha visto continuada por los trabajos del investigador australiano Philip Rose, particularmente en su obra más conocida, (*Forensic Speaker Identification*, 2002). Las semejanzas con el estudio de Nolan (1983) son especialmente notables en el sentido de que, en esta obra, Rose pretende describir, no solo el estado de la cuestión de la comparación de hablantes, en el marco de la disciplina más general de la fonética judicial, sino también, y muy detalladamente, las bases fonéticas en las que se sustentan los análisis forenses. Casi al final del libro, encontramos un apartado dedicado exclusivamente a la definición del término “voz” (cf. cap.10. *What is a voice?*), donde se pone de manifiesto la idea de Rose sobre qué es la cualidad de voz. En primer lugar, se establece una clara distinción entre cualidad de voz (*voice quality*) y cualidad fonética (*phonetic quality*), como ya había hecho Laver en muchas de sus obras

(cf., por ejemplo, Laver, 1991, 1994, entre otros trabajos citables). Mientras que con el último término Rose se refiere a los aspectos del sonido de una voz que conllevan información lingüística, más en concreto, fonológica, con el primer concepto se recubren las características de la persona que habla, “lo que uno puede oír cuando se elimina la “cualidad fonética”; por ejemplo, cuando se escucha a alguien hablando detrás de una puerta pero lo que dice no es inteligible” (p. 279, traducción nuestra).

En otras palabras, Rose incide de nuevo en el hecho de que la identidad de un hablante puede señalarse tanto mediante rasgos lingüísticos (como, por ejemplo, la elección de alófonos) como a través de rasgos no lingüísticos. Resulta muy clarificador el ejemplo que el autor expone a este respecto, relativo al ajuste articulatorio de velarización. Se trata de la <a> en la palabra inglesa *cart*. En inglés australiano, se pronunciaría [a:], pero, como indica Rose, también podría articularse, sin transgredir la norma lingüística, con mayor velarización, como [ɑ:]. Ello quiere decir que el grado de velarización de esta última vocal podría ser, o bien el resultado de un ajuste articulatorio casi permanente, o bien simplemente una realización alofónica permitida por la lengua. Así pues, si algunos rasgos pueden ser exponentes o bien de la “cualidad fonética” o bien de la “cualidad de voz”, ¿cómo saber cuándo estamos ante un caso o ante el otro? Esto depende básicamente de la duración de la aplicación del rasgo: si este es un exponente lingüístico, su duración será breve, y el alcance de su aplicación, limitado; por el contrario, los rasgos que implican ajustes articulatorios definitorios de la cualidad de voz son casi permanentes. En el caso de *cart*, una pronunciación [kʰɑ:t] indica que la velarización de la vocal es un rasgo fonético, claramente localizado. Por el contrario, una pronunciación [kʰɑ:t], que conlleva la pronunciación también más posterior del sonido inicial de la palabra, indicaría que la velarización constituye un ajuste articulatorio deliberado de “cualidad de voz”.

Finalmente, Rose (2002) incluye en su obra un apartado dedicado a la relevancia forense tanto de la cualidad fonética como de la cualidad de voz. Tomando ejemplos de Rose y Duncan (1995), afirma que, en la discriminación profana de voces (esto es, la llevada a cabo por no expertos), es más probable que la segunda predomine sobre la primera. Así, si dos muestras de habla, de la misma lengua, son muy similares en cuanto a la cualidad de voz, los oyentes profanos suelen formarse la opinión de que pertenecen a la misma persona independientemente de si contienen diferencias fonéticas, siempre y cuando estas no sean muy grandes. Para ejemplificar esta idea, supone la existencia de dos muestras de voz, constituida cada una de ellas únicamente por la palabra *hello*, pronunciada en un caso como [hɛ'leʊ] y en otro como [hɛ'leʊ]. Si estos dos *hello* fueran emitidos por distintas personas con una cualidad de voz semejante (por ejemplo, con una frecuencia fundamental casi idéntica, con un tipo de fonación similar, etc.), la mayoría de los oyentes profanos juzgarían que se trata del mismo locutor.

Antes de concluir este apartado cabe mencionar el estudio de Michel Jessen (2008), que dedica un apartado especial a la cualidad de voz en su primera sección, sobre comparación de voces. Jessen explica cómo los hablantes pueden diferenciarse por su empleo de una cualidad de voz neutra, es decir “modal”, o bien por sus desviaciones notables de dicho modelo neutro, al adoptar una voz soplada (*breathy voice*), o una voz rota (*creaky voice*), o una voz rasgada (*rough/harsh voice*), o una voz tensa (*pressed/tense voice*), o bien una combinación de algunas de ellas. Asimismo, este autor puntualiza que los locutores pueden diferir en el grado en que se presente ese rasgo de cualidad de voz, por ejemplo: presencia de una voz *marcadamente* susurrada o *ligeramente* susurrada. Obsérvese que en este punto Jessen se refiere solo a tipos de voz que vienen determinados

por la acción laríngea. Sin embargo, la cualidad de voz, como ya se ha dicho (cf., igualmente, Nolan 2005, 2007), también se puede entender en su sentido más amplio, abarcando características de origen supralaríngeo, como sucede en el caso de la voz nasal, y esto constituye igualmente un rasgo interesante para la identificación que los peritos han de tomar en consideración. De hecho, Jessen se refiere a las escalas de descripción perceptiva de la cualidad de voz (empleadas, por cierto, con fines forenses por la policía alemana, cf. Köster y Köster, 2004) que subsumen tanto el componente laríngeo como el supralaríngeo (cf. *supra*, §3), y precisa que se ha comprobado que el entrenamiento de los expertos forenses en el uso de tales escalas puede mejorar tanto el acuerdo entre expertos como la precisión en sus informes por lo que concierne concretamente a la percepción de la cualidad de voz (Köster *et al.*, 2007).

El apartado dedicado a este tema en Jessen (2008) termina con una referencia a un estudio de Nolan (2005), para remarcar ciertas limitaciones del empleo de la cualidad de voz en fonética judicial, debidas, por ejemplo, al canal telefónico a través del cual se realiza la mayoría de las grabaciones –si no todas– que deben comparar los expertos. En el estudio de Nolan (2005) mencionado por Jessen nos detendremos en el próximo apartado, puesto que constituye, no ya una referencia a la cualidad de voz en una obra general sobre fonética judicial, como las comentadas en esta sección, sino un estudio dedicado por completo a examinar las posibilidades de aplicación de esta noción a la práctica judicial real.

6. La cualidad de voz en la práctica forense: ventajas e inconvenientes

Ante todo lo expuesto en apartados anteriores, un experto en fonética judicial puede formularse de inmediato la pregunta de si es conveniente o no, si es útil o no, considerar la cualidad de voz como un parámetro de análisis más, susceptible de ser evaluado para la realización de una pericia, bien se trate esta de la caracterización de una única voz, bien sea una comparación forense de un cierto número de voces.

No todos los peritajes forenses que se realizan actualmente inciden en este concepto, a pesar de que la mayoría de los expertos lo considera fundamental para decidir sobre la identidad del hablante. Nolan (2005) hace referencia a sus propios informes periciales, y reconoce que la única tarea, en el marco de la fonética judicial, donde él mismo ha sido más explícito en la aplicación del análisis perceptivo de la cualidad de voz es en la preparación de ruedas de reconocimiento, en concreto en la fase previa de selección de “distractores” (esto es, de voces parecidas a las del sospechoso):

The only area of forensic phonetic work where I have been more explicit in applying voice quality analysis is in the preparation of two ‘voice lineups’ – identification parades to test whether it was the suspect whom a witness heard at the time of a crime. This requires the selection of (usually around eight) ‘foil’ voices, among which the suspect’s sample is placed. The foil samples have to be chosen so that the suspect’s sample does not stand out, and one aspect of this is voice quality. [...] My working notes on the potential foil samples included comments such as ‘heavy nasalization’, ‘tenser voice than [the suspect]’, and ‘light voice’ (not a term within Laver’s framework, but which I would have been using as a cover term for a voice with relatively high pitch and formant frequencies, and possibly phonation tending toward breather rather than tenser settings). I also rated the distance in terms of voice quality (and separately, accent) between each foil and the suspect on a numerical scale; my voice quality ratings were later shown to be in good agreement with unspecified similarity judgments by a group of subjects (Nolan & Grabe, 1996). (pp. 393-394).

Tanto en este mismo trabajo, Nolan (2005), como posteriormente en Nolan (2007), se mencionan algunas causas que podrían explicar por qué se soslaya la calidad de voz (con ejemplos tomados del ámbito británico, que son extrapolables a nuestro país). Una primera explicación podría ser la falta de conocimiento de la existencia del marco *VPA* diseñado por Laver, si bien esta parece ser una razón poco plausible, que Nolan rápidamente descarta, pues, aunque los expertos no aludan explícitamente al trabajo de Laver, en la mayoría de las grandes obras de referencia de fonética judicial se cita el trabajo de este investigador y, por consiguiente, su figura y su obra son suficientemente conocidas (cf. p. 396). Otro posible motivo aducido por Nolan es la falta de formación de los expertos en el análisis de la calidad de voz. Esta sí es, indudablemente, una causa evidente: en términos relativos, hay muy pocas personas de formación lingüístico-fonética que estén entrenadas en la aplicación del modelo *VPA* de análisis componencial de la voz (que, recuérdese, entre los fonetistas es el que se suele emplear con mayor frecuencia). Esto es un hecho. No obstante, existen otros importantes factores adicionales en juego, que Nolan cita y que, ciertamente, condicionan la mayor parte de los informes.

Uno decisivo, mencionado por el mismo Nolan, es la falta de tiempo a la que se enfrentan los expertos cuando tienen que llevar a cabo un peritaje. Realizar un examen completo de una muestra de voz bajo el marco descriptivo *VPA* implica varias escuchas sucesivas y detalladas, atendiendo aproximadamente a 40 ajustes posibles; y no hay que olvidar que, en un contexto judicial, por lo general no es una única voz la que el experto debe analizar sino que, al menos, existen dos, y, muy frecuentemente, bastantes más de dos.

Asimismo, otro factor igualmente potente es la calidad de las muestras. En el ámbito forense, el fonetista experto suele trabajar con muestras de voz obtenidas a partir de interceptaciones telefónicas. Puesto que el canal telefónico es un filtro de paso banda que corta las frecuencias inferiores a los 300 Hz y superiores a los 3500 Hz, intentar adoptar en este tipo de grabaciones el marco ideado por Laver para el análisis de la calidad de voz es un reto complicado. Además, las muestras de voz interceptada no solo poseen esa calidad que podríamos denominar “de tipo telefónico”, sino que, en muchas ocasiones, conllevan ruidos externos procedentes de la calle o, incluso, el solapamiento de distintas voces, si la llamada se ha realizado desde un espacio público. Esto lleva a Nolan (2005) a subrayar que será de poco valor probatorio cualquier informe que establezca una similitud entre dos muestras de voz si únicamente una de ellas es una grabación telefónica y la otra no. Filtrar la muestra no telefónica puede aportar cierto valor a la comparación, pero se corre el riesgo de provocar artificialmente que las voces comparadas suenen más similares de lo que tal vez los hablantes suenen en la realidad.

Finalmente, otra de las razones por la que los fonetistas podrían ser reacios a utilizar análisis perceptivos de calidad de voz en sus informes podría deberse al factor “variación intralocutor”, del que ya hablamos en el §1. Es conocido que la calidad de voz de un hablante varía según el estilo de habla y la función pragmática. Podemos encontrar, pues, una grabación dubitada en la que el hablante realice una llamada obscena utilizando una voz susurrada, mientras que en la grabación indubitada, obtenida en dependencias judiciales, la voz es completamente diferente. En estos casos, es tarea del fonetista forense decidir si esa diferencia en la calidad de voz entre las dos muestras se puede explicar por factores lingüísticos (como lo es el distinto contexto) o no.

Siendo todo esto así, ¿cuáles son las ventajas de realizar el análisis –perceptivo y, en el grado en que se pueda, acústico– de la cualidad de voz en un peritaje con fines judiciales?

Puede aducirse un argumento fundamental a favor de hacerlo. El análisis de la cualidad de la voz debe constituir una parte importante en la elaboración del “pasaporte vocal”¹⁴ del hablante, porque, como bien señala Nolan (2005, p. 389), la hipótesis de que cada individuo tiene un ‘idiolecto’ definido no está probada, y, aunque así fuera, es una idea poco práctica dado el escaso número de muestras de habla con las que se trabaja en el ámbito judicial. Es decir que, frente al estudio aislado de vocales, de consonantes o incluso de patrones entonativos (que reducirían el “pasaporte vocal” a una simple caracterización de la comunidad de habla en la que se inscribe el sujeto), el enfoque realmente pertinente en fonética judicial es el que se centra en las propiedades del habla *específicas* de un locutor; y entre ellas se encuentra la cualidad de voz. En efecto, la individualidad que reside en la voz del hablante procede, como aquí se ha dicho, de dos fuentes, su anatomía y sus hábitos articulatorios aprendidos, que se corresponden precisamente con la distinción que Laver (1976) estableció entre aspectos “intrínsecos” y “extrínsecos” de la cualidad de voz. Ambos aspectos quedan recogidos en el marco del VPA, que suministra, pues, la infraestructura conceptual necesaria para reflexionar acerca de cómo difieren los hablantes (cf. Nolan, 2005, p.409). En cierta medida, el estudio de la cualidad de voz es el estudio del factor común dinámico que subyace en todas las producciones de un locutor, lo que el propio Nolan denomina la “firma” de los hablantes¹⁵.

El análisis de la cualidad de voz habrá de realizarse, sin embargo, con precaución. Puesto que la objeción más importante para la aplicación forense del marco VPA, además de las ya señaladas al comienzo de este apartado y en apartados anteriores, se deriva del hecho de que, como se ha dicho, los trectos vocales de los hablantes no son completamente equivalentes, al describir las muestras de habla forense en términos de “ajustes”, simplemente se deberá tener siempre en cuenta que la etiqueta que describe un componente determinado de la cualidad de voz está basada predominantemente en el resultado auditivo (cf. *supra* § 2) y, en consecuencia, no podrá afirmarse sin género de dudas si el origen de ese componente viene determinando orgánicamente o es fruto de la elección del hablante, ni tampoco si es ese componente el principal responsable del resultado perceptivo general. Otra precaución que se ha de tener presente, y que ya señaló Nolan (2005, p. 400) adelantándose al curso que posteriormente seguiría la investigación, es la de que no basta con establecer si dos muestras son similares o diferentes, sino que, en el caso de que se asemejen, es importante establecer si lo hacen de forma tal que se diferencien del resto de la población relevante. Esto significa que siempre es aconsejable aplicar en la pericia un enfoque bayesiano, que expresa las probabilidades en forma de relaciones de verosimilitud (*likelihood ratio*)¹⁶.

Finalmente, es importantísimo tratar de afianzar las apreciaciones auditivas, en la medida de lo posible, mediante la pormenorización de las características acústicas a que los ajustes percibidos dan lugar. Por eso, precisamente, el enfoque fonético no automático más recomendable para llevar a cabo los peritajes judiciales es el híbrido, al tiempo perceptivo y acústico: la comparación de voces en el proceso de identificación del locutor es una tarea tan difícil que requiere de todas las herramientas de análisis que puedan arrojar algo de luz y contribuyan a fundamentar al máximo las conclusiones.

NOTAS

¹ Algunas descripciones se apartan claramente de los estereotipos: “Her voice is full of money”, he said suddenly. “That was it. I’d never understood before. It was full of money...” (F. Scott Fitzgerald, *The Great Gatsby*, Scribner and Sons, Nueva York, 1925, cap.7).

² Nolan (1983) utiliza la palabra “strand”, adoptando la terminología de Abercrombie (1967, p. 89 y ss.).

³ Crystal realmente prefiere restringir el uso de “cualidad de voz” para denominar los rasgos de tipo intrínseco.

⁴ La matización es importante porque, aun dentro de las limitaciones que la anatomía impone a cada hablante, la capacidad de alterar la voz es grande, y es frecuente encontrar personas con una habilidad especial para ello.

⁵ Para la discusión en profundidad de las correspondencias entre la cualidad de voz y las características físicas del hablante, puede consultarse Kreiman y Sidtis (2011), en particular el capítulo 4, en el que se hace una revisión crítica exhaustiva de los experimentos realizados hasta ahora para fundamentar las intuiciones que todos tenemos acerca de tales correspondencias.

⁶ Sobre la cualidad de voz profesional, puede consultarse en la red la tesis doctoral de Carolina Pérez, “Ajustes laríngeos responsables de la cualidad de voz en estilos de habla profesionales”, defendida en 2011. Cf., igualmente, Leino (1994).

⁷ Honikman (1964) es la obra clásica en la que se define por primera vez de modo detallado lo que es un “ajuste” (*setting*), si bien el concepto y el término se asocian en este trabajo únicamente con los ajustes de tipo articulatorio localizados en las cavidades supraglóticas. Otros autores posteriores amplían el ámbito recubierto por la noción para abarcar también todas las variaciones posibles de la actividad laríngea.

⁸ Conviene recordar que este protocolo se emplea con frecuencia en la esfera clínica, para describir las patologías vocales más frecuentes, en particular las que afectan a la laringe.

⁹ Cada uno de estos factores admite básicamente cuatro grados, aunque algunos especialistas proponen especificaciones intermedias: Normal (0), Leve (1), Moderado (2) e Intenso (3).

¹⁰ No todos los autores coinciden al precisar la duración de la muestra que el análisis del LTAS requiere. Li *et al.* (1969), por ejemplo, mencionan 30-40 segundos; mientras que Majewski *et al.* (1977) recomiendan disponer de al menos 1 minuto de habla.

¹¹ Esto es así siempre y cuando el habla no haya sido voluntaria o involuntariamente distorsionada por el hablante. En los experimentos de Doherty y Hollien (1978), por ejemplo, a fin de estudiar la influencia de la distorsión de la voz, se crearon tres condiciones de grabación: lectura normal, lectura bajo estrés, y lectura con disimulo de la voz. En el segundo caso, el estrés se generó mediante la aplicación aleatoria de un impulso eléctrico (no dañino, pero intenso) al hablante durante la lectura. Para la condición de disimulo, se pidió a los locutores que modificaran su voz de la forma que prefirieran, excepto mediante el uso de un acento extranjero o mediante el susurro. Los resultados de esta investigación mostraron que ninguno de los vectores a los que nos hemos referido en el texto, ni de forma aislada ni en combinación, era capaz de diferenciar a los hablantes cuando su voz presentaba algún tipo de distorsión.

¹² Sus grabaciones se realizaron antes y después del transcurso de una jornada laboral de los sujetos en la que se habían visto obligados a hablar mucho, lo cual puede también influir en el LTAS (véase, por ejemplo, Jónsdóttir *et al.*, 2003).

¹³ Este autor incluye dentro de lo que él denomina “idiolecto” fenómenos como las aspiraciones, las nasalizaciones o las características acústicas de las pausas y las dudas.

¹⁴ En los informes que se realizan en el Laboratorio de Fonética del CSIC, al que están adscritas las dos autoras del presente trabajo, sí se incluye habitualmente un apartado de análisis de la cualidad de voz de los sujetos estudiados, conforme al protocolo VPA.

¹⁵ Por “pasaporte vocal” entendemos la caracterización global del habla de un individuo a partir de las muestras recogidas: su pronunciación, sus rasgos dialectales y lingüísticos, el análisis del discurso empleado, etc.

¹⁶ En su obra posterior (Nolan, 2007), este autor va a dedicar bastante más espacio a desarrollar esta idea. Basándose en su propia experiencia práctica, considera que lo más cerca que podremos estar de encontrar esa firma única será a través del estudio dinámico del habla, por ejemplo, del movimiento de los formantes (trayectorias formánticas), y cita algunos estudios del inglés británico que está llevando a cabo en el momento en que escribe junto a una colega (McDougall y Nolan, 2007), así como otras colaboraciones científicas (Nolan y Grigoras, 2005) que tienen como objetivo desarrollar una herramienta para el estudio cuantitativo de la cualidad de voz supralaríngea: un método que utiliza la predicción lineal (*Linear Prediction Analysis*) para estimar las frecuencias formánticas en determinados momentos a lo largo de un discurso y, posteriormente, dibuja la distribución estadística de dichas frecuencias formánticas.

¹⁷ Para una exposición detallada del denominado “nuevo paradigma” aplicado en la comparación de voz con fines judiciales, que está basado en la ratio de verosimilitud (LR, *likelihood ratio*), puede consultarse, entre otros muchos títulos mencionables, Morrison (2009).

Referencias bibliográficas

- Abercrombie, D. (1967): *Elements of General Phonetics*. Edinburgh University Press. Edimburgo.
- Ball, M. J., Esling, J. y Dickson, C. (2000): *The transcription of voice quality*. En R. D. Kent y M. J. Ball (Eds.) *Voice Quality Measurement*. Singular. San Diego, pp. 49-58.
- Baumann, O. y Belin, P. (2008): “Perceptual scaling of voice identity: common dimensions for different vowels and speakers” en *Psychological Research*, Elsevier. Accesible en <http://www.psy.gla.ac.uk/docs/download.php?type=PUBLS&id=1339>
- Bhuta, T., Patrick, L. y Garnett, J. (2004): “Perceptual evaluation of voice quality and its correlation with acoustic measurements” en *Journal of Voice* 18 (3), pp. 299-304.
- Camargo, Z. (2002): *Análise da qualidade vocal de um grupo de indivíduos disfônicos: uma abordagem interpretativa e integrada de dados de natureza acústica, perceptiva e eletroglotográfica*. Tesis Doctoral, Pontificia Universidade Católica de São Paulo, Brasil.
- Camargo, Z. y Madureira, S. (2008): “Voice quality analysis from a phonetic perspective: Voice profile analysis scheme profile for Brazilian Portuguese (BP-VPAS)” en *Proceedings of the Fourth Conference on Speech Prosody (Speech Prosody 2008)*. Campinas. Brasil, pp. 57-60.
- Coadou, M. (2006): “Voice quality and variation: a pilot study of the Liverpool accent” en *Speech Prosody 2006*. Dresde. Comunicación 123.
- Coadou, M. (2007): “Voice quality and variation in English” en *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, pp. 2077-2080.
- Crystal, D. (1969): *Prosodic Systems and Intonation in English*. Londres. Cambridge University Press.
- Doherty, E. (1976): “Evaluation of selected acoustic parameters for use in speaker identification” en *Journal of the Acoustical Society of America*, 58, S1, p.107.
- Doherty, E. y Hollien, H. (1978): “Multiple-factor speaker identification of normal and distorted speech” en *Journal of Phonetics*, 6, pp.1-8.

- Forrest, K., Weismer, G., Milenkovic, P. y Dougall, R. (1988): "Statistical analysis of word-initial voiceless obstruents: preliminary data" en *Journal of the Acoustical Society of America* 84, pp. 115-123.
- French, P. (1994): "An overview of forensic phonetics with particular reference to speaker identification" en *Forensic Linguistics: The International Journal of Speech, Language and the Law* 1 (2), pp. 197-206.
- French, P. y Stevens, L. (en prensa): *Forensic Speech Science*. En M. Jones y R. Knight (Eds.) *The Bloomsbury Companion to Phonetics*. Continuum. Londres.
- Gelfer, M. P. (1993): "A multidimensional scaling study of voice quality in females" en *Phonetica* 50, pp. 15-27.
- Gil, J. (2007): *Fonética para profesores de español: De la teoría a la práctica*. Arco/Libros. Madrid.
- Gil, J. (2012): *Bibliografía Temática de Fonética Judicial*. Accesible en http://www.estudiosfonicos.cchs.csic.es/fonetica/civil/Gil,J. 2011_Bibliografia_Tematica_de_Fonetica_Judicial.pdf
- Hammarberg, B. (2000): "Voice research and clinical needs" en *Folia Phoniatica et Logopedica* 52, pp. 93-102.
- Harmegnies, B. y Landercey, A. (1988): "Intra-speaker variability of the long term speech spectrum" en *Speech Communication* 7, pp. 81-86.
- Hirano, M. (1981): *Clinical Examination of Voice*. Springer. Berlín.
- Hollien, H. (1990): *The Acoustics of Crime: The New Science of Forensic Phonetics*. Plenum Press. Nueva York y Londres.
- Honikman, B. (1964): *Articulatory settings*. En D. Abercrombie, D. G. Fry, P.A.D. MacCarthy, N.C. Scott y J. L. Trim (Eds.) *In Honour of Daniel Jones*. Longman. Londres, pp. 73-84.
- Jessen, M. (2008): "Forensic Phonetics" en *Language and Linguistic Compass*, 2 (4), pp. 671-711.
- Jónsdóttir, V., Laukkanen, A. M., y Siikki, I. (2003): "Changes in teachers' voice quality during a working day with and without electric sound amplification" en *Folia phoniatica et logopaedica*, 55(5), 267-280.
- Juslin, P. N. y Laukka, P. (2001): "Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion" en *Emotion* 1, pp. 381-412.
- Keller, E. (2004): "The analysis of voice quality in speech processing" en *Tutorial Research Workshop "Nonlinear Speech Processing: Algorithms and Analysis"*, 12-17 Septiembre, Vietri, Italia. Accesible en citeseerx.ist.psu.edu
- Kempster, G.B., Gerratt, B. R., Verdolini, K., Barkmeier-Kraemer, J. y Hillman, R. E. (2009): "Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol" en *American Journal of Speech Language Pathology* 18, pp. 124-132.
- Kitzing, P. (1986): "LTAS-criteria pertinent to the measurement of voice quality" en *Journal of Phonetics* 14, pp. 477-482.
- Klasmeyer, G. (2000): "An automatic description tool for time-contours and long-term average voice features in large emotional speech databases" en *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion: Speech Emotion-2000*, Newcastle, pp. 66-71.
- Knowles, G. (1978): *Scouse, the Urban Dialect of Liverpool*. Tesis Doctoral de la Universidad de Leeds.
- Köster, O. y Köster, J-P. (2004): "The auditory-perceptual evaluation of voice quality in forensic speaker recognition" en *The Phonetician* 89, pp. 9-37.

- Köster, O., Jessen, M., Khairi, F., y Eckert, H. (2007): "Auditory-perceptual identification of voice quality by expert and non-expert listeners" en *Proceedings 16th International Congress of Phonetic Sciences*, Saarbrücken, pp. 1845-1848.
- Kreiman, J., Gerratt, B. R., Precoda, K. y Berke, G. S. (1992): "Individual differences in voice quality perception" en *Journal of Speech and Hearing Research* 35, pp. 512-520.
- Kreiman, J., Gerratt, B. R. e Ito, M. (2007): "When and why listeners disagree in voice quality assessment tasks" en *Journal of the Acoustical Society of America* 122, pp. 2354-2364.
- Kreiman, J. y Sidtis, D. (2011): *Foundations of Voice Studies*. Wiley-Blackwell. Oxford.
- Künzel, H. (1994): *Sprecherkennung: Grundzüge forensischer Sprachverarbeitung*. Kriminalistik Verlag. Heidelberg.
- Laver, J. (1968): "Voice quality and indexical information" en *British Journal of Disorders of Communication* 3, pp. 43-54. Reproducido en Laver, J. (1991): *The Gift of Speech*. Edinburgh University Press. Edimburgo, pp. 147-161.
- Laver, J. (1974): "Labels for voices" en *Journal of the International Phonetic Association* 4, pp. 62-75. Reproducido en Laver, J. (1991): *The Gift of Speech*. Edinburgh University Press. Edimburgo, pp. 171-183.
- Laver, J. (1976): "The semiotic nature of phonetic data" en *York Papers in Linguistics* 6, pp. 55-62. Reproducido en Laver, J. (1991): *The Gift of Speech*. Edinburgh University Press. Edimburgo, pp. 162-170.
- Laver, J. (1979): "The description of voice quality in general phonetic theory" en *Edinburgh University Department of Linguistics Work in Progress* 12, pp. 30-52. Reproducido en Laver, J. (1991): *The Gift of Speech*. Edinburgh University Press. Edimburgo, pp. 184-208.
- Laver, J. (1980): *The Phonetic Description of Voice Quality*. Cambridge University Press. Cambridge.
- Laver, J. (1991): *The Gift of Speech*. Edinburgh University Press, Edimburgo.
- Laver, J. (1994): *Principles of Phonetics*. Cambridge University Press. Cambridge.
- Laver, J. y Hanson, R. J. (1981): "Describing the normal voice". En J. K. Darby (Ed.) *Speech Evaluation in Psychiatry*. Grune and Stratton. Nueva York, pp. 57-78. Reproducido en Laver, J. (1991): *The Gift of Speech*. Edinburgh University Press. Edimburgo, pp. 209-234.
- Laver, J., Wirz, S., Mackenzie, J. y Hiller, S. M. (1981): "A perceptual protocol for the analysis of vocal profiles" en *Edinburgh University Department of Linguistics Work in Progress* 14, pp. 139-155. Reproducido en Laver, J. (1991): *The Gift of Speech*. Edinburgh University Press. Edimburgo, pp. 265-280.
- Leino, T. (1994): *Long-term average spectrum study on speaking voice quality in male actors*. En A. Friberg, J. Iwarsson, E. Jansson y J. Sundberg (Eds.) *SMAC 93 (Proceedings of the Stockholm Music Acoustics Conference, 1993)*. Publicación n°. 79, Royal Swedish Academy of Music. Estocolmo, pp. 206-210.
- Li, K. P., Hughes, G. W. y House, A. S. (1969): "Correlation characteristics and dimensionality of speech spectra" en *Journal of Acoustical Society of America* 46, pp. 1019-1025.
- Lindblom, B. y Sundberg, J. (1971): "Acoustical consequences of lip, tongue, jaw, and larynx movement" en *Journal of the Acoustical Society of America* 50, pp. 1166-79.
- Löfqvist, A. (1986): "The long-time-average spectrum as a tool in voice research" en *Journal of Phonetics* 14, pp. 471-5.

- Löfqvist, A. y Mandersson, B. (1987): "Long-time average spectrum of speech and voice analysis" en *Folia Phoniatica* 3, pp. 221-9.
- Mackenzie-Beck, J. (2005): *Perceptual analysis of voice quality: The place of vocal profile analysis*. En W. J. Hardcastle y J. Mackenzie-Beck (Eds.) *A Figure of Speech: A Festschrift for John Laver*. Lawrence Erlbaum Associates. New Jersey, pp. 285-322.
- Majewski, W., Zalewski, J. y Hollien, H. (1979): *Some remarks on different speaker identification techniques*. En H. Hollien y P. Hollien (Eds.) *Current Issues in the Phonetic Sciences*. John Benjamins. Amsterdam, pp. 829-835.
- McDougall, K. y Nolan, F. (2007): "Discrimination of speakers using the formant dynamics of /u:/ in British English" en *Proceedings 16th International Congress of Phonetic Sciences*. Saarbrücken, pp. 1825-1828.
- Mendoza, E., Valencia, N., Muñoz, J. y Trujillo, H. (1996): "Differences in voice quality between men and women: use of the long-term average spectrum (LTAS)" en *Journal of Voice* 10, pp. 59-66.
- Molina de Figuereido, R. y Bernales, M. (1999): "Reconocimiento de hablantes basado en el espectro de largo tiempo" en *Actas del VI Simposio Internacional de Comunicación Social*. Santiago de Cuba, pp. 1372-1378.
- Morrison, G. S. (2009): "Forensic voice comparison and the paradigm shift" en *Science & Justice* 49, pp. 298-308.
- Murray, I. R. y Arnott, J.L. (1993): "Toward the simulation of emotion in synthetic speech: A review of literature on human vocal emotion" en *Journal of the Acoustical Society of America* 93, pp. 1097-1108.
- Nolan F. (1983): *The Phonetic Bases of Speaker Recognition*. Cambridge University Press. Cambridge.
- Nolan, F. (1991): "Forensic phonetics" en *Journal of Linguistics* 27, pp. 483-493.
- Nolan, F. (1997): *Speaker recognition and forensic phonetics*. En W.J. Hardcastle and J. Laver (Eds.) *A handbook of phonetic sciences*. Blackwell. Oxford, pp. 744-767.
- Nolan, F. (2005): *Forensic speaker identification and the phonetic description of voice quality*. En W.J. Hardcastle y J. Mackenzie-Beck (Eds.) *A Figure of Speech. A Festschrift for John Laver*. Lawrence Erlbaum Associates. New Jersey, pp. 385-411.
- Nolan, F. (2007): "Voice quality and forensic speaker identification" en *GOVOR* 24 (2), pp. 111-128.
- Nolan, F. y Grigoras, C. (2005): "A case for formant analysis in forensic speaker identification" en *International Journal of Speech, Language and the Law* 12 (2), pp. 143-173.
- Nordenberg, M. y Sundberg, J. (2004): "Effect on LTAS of vocal loudness variation" en *Logopedics, Phoniatrics, Vocology* 29, pp. 183-191. Accesible en <http://www.speech.kth.se/qpsr/tmh/2003/03-45-093-100.pdf>.
- Pittam, J. (1987a): "Discrimination of five voice qualities and prediction to perceptual ratings" en *Phonetica* 44, pp. 38-49.
- Pittam, J. (1987b): "The long term spectral measurement of voice quality as a social and personality marker: a review" en *Language and Speech*, 30, pp. 1-12.
- Pittam, J. y Millar, J. B. (1988): "The long-term spectrum of voice" en *2nd Proceedings of the Australian International Conference on Speech Science and Technology*. Sydney, pp. 308-13.
- Rose, P. (2002): *Forensic Speaker Identification*. Taylor & Francis. Londres.
- Rose, P. y Duncan, S. (1995): "Naive auditory identification and discrimination of similar voices by familiar listeners" en *Forensic Linguistics* 2 (1), pp. 1-17.

- Scherer, K. (1986): "Vocal affect expression: A review and a model for future research" en *Psychological Bulletin* 99, pp. 143-165.
- Scherer, K. (2003): "Vocal communication of emotion: A review of research paradigms" en *Speech Communication* 40, pp. 227-256.
- Stuart-Smith, J. (1999): *Glasgow: Accent and voice quality*. En P. Foulkes y G. Doherty (Eds.) *Urban Voices: Accent Studies in the British Isles*. Edward Arnold. Londres, pp. 201-220.
- Trudgill, P. (1974): *The Sociolinguistic Differentiation of English in Norwich*. Cambridge University Press. Cambridge.
- Wendler, J., Doherty, E.T. y Hollien, H. (1980): "Voice classification by means of long-term speech spectra" en *Folia Phoniatria* 32, pp. 51-60.
- Wendler, J., Rauhut, A. y Krüger, H. (1986): "Classification of voice qualities" en *Journal of Phonetics* 14, pp. 483-8.
- Wilson, D. K. (1987): *Voice Problems in Children*. Williams & Wilkins. Baltimore.
- Zalewski, J., Majewski, W. y Hollien, H. (1975): "Cross-correlation between Long-Term Speech Spectra as a criterion for speaker identification" en *Acoustics* 34, pp. 20-24.

PLEASE CITE AS: Gil, J. & San Segundo, E. (2014). La calidad de voz en fonética judicial. In Elena Garayzábal & Mercedes Reigosa (eds.), *Lingüística forense. La lingüística en el ámbito legal y policial* (pp. 154-199). Madrid: Euphonía Ediciones.