

RUBÉN WAINSCHENKER

JORGE DOORN

MARCELA CASTRO

Departamento de Computación y Sistemas–Facultad de Ciencias Exactas

Universidad Nacional del Centro. Provincia de Buenos Aires

Tandil, Argentina

E-Mail: rfw@exa.unicen.edu.ar

E-Mail: jdoorn@exa.unicen.edu.ar

E-Mail: mcastro@exa.unicen.edu.ar

Estudio de la velocidad del habla

Resumen. La magnitud velocidad del habla no tiene una definición precisa, pese a que es abundantemente utilizada tanto en la actividades diarias como en ocupaciones tales como la locución, la dactilografía y la taquigrafía entre otras.

En el contexto del estudio del comportamiento y caracterización de los alófonos típicos de un idioma, como componentes fundamentales en la síntesis robusta del habla, dicho concepto es trascendental, no hallándose bibliografía precisa.

En este artículo se presenta una caracterización cuantitativa de las nociones intuitivas de la velocidad normal, lenta y rápida del habla. Este trabajo tiene una raíz fuertemente experimental ya que sus conclusiones son obtenidas de mediciones realizadas sobre 120 textos emitidos por diferentes locutores a distinta velocidad en un contexto completamente libre de condicionamientos y preconceptos.

Introducción. El habla es una de las principales manifestaciones de la inteligencia humana naturalmente utilizada para dar y recibir información. La voz es la realización acústica del complejo mecanismo que es el lenguaje humano. En los tiempos actuales, éste ha sido tomado como objeto propio de diversas disciplinas. Cada ciencia o arte lo aborda desde un punto de vista distinto y particular complicando la situación pues se produce una superposición de ámbitos. [1]

La habilidad de producir sonidos individuales del habla, no implica la habilidad de “hablar”. Desde el punto de vista de la síntesis de voz, la simple concatenación de los sonidos elementales de un lenguaje conduce a generar un sonido poco natural e inteligible. La producción de voz natural y continua es un desafío permanente en aplicaciones informáticas que necesitan sintetizar voz, ya que la mayoría de los enfoques conocidos sólo producen un sonido artificial. [2] Un aspecto muy poco tenido en cuenta pero que indudablemente será de gran importancia en el futuro, es la incorporación de parámetros en el modelo de síntesis que controle la velocidad del habla producida. La síntesis de voz imita la naturaleza por lo que su primera obligación es conocerla en la mayor profundidad posible. El presente trabajo pretende producir una contribución en este sentido.

La fonética del Castellano hablado en las regiones del sur y centro de Argentina (incluido Uruguay) tiene una marcada diferencia en relación con el Castellano tradicional. A partir de este aspecto es posible que se detecten inconsistencias y ambigüedades lo que lleva a la necesidad de realizar un estudio del habla regional más profundo. [3]

Velocidad del habla. El concepto “velocidad del habla” juega un papel trascendente dentro del contexto del estudio del comportamiento y caracterización de los alófonos típicos del Castellano Rioplatense, cuando se busca desarrollar una síntesis robusta de la voz [2,4]. Navarro Thomas [5] define como velocidad normal a aquella que se obtiene cuando se emiten 205 ppm mientras que Loprete [1] establece esta misma noción entre 120 y 150 ppm. Los contextos de ambas afirmaciones difieren, ya que la primera corresponde al Castellano Hispano hablado a mediados del siglo XX y la segunda al Castellano Rioplatense de fines del mismo siglo.

Otras fuentes [6] parecen confirmar al menos el límite inferior del habla normal por ejemplo en los cursos de taquigrafía consideran apto un estudiante que se acerque a 120 ppm. El problema de los límites imprecisos persiste, ya las dificultades de este sistema de medición son evidentes cuando se consideran la calidad de la unidad de medición, pues distintos textos tienen diferente cantidad de alófonos por palabra. Y tanto en lo que indica el sentido común como en las experiencias que se relatan en las siguientes secciones, queda claro que la unidad alófono por unidad de tiempo es mucho más invariante con la naturaleza del texto, especialmente en textos cortos.

Para evaluar lo apropiado de una técnica de medición es conveniente recordar los principios básicos de la teoría de mediciones, ampliamente difundida en otras disciplinas. Cinco categorías de mediciones son habitualmente aceptadas [7,8]:

1. Cualitativa
2. Ordinal
3. Por intervalos
4. Proporcional
5. Cardinal

La calidad de la información que se obtiene crece de la medición cualitativa a la medición cardinal. La primera mejora respecto de la medición cualitativa establecida por la medición ordinal consiste en garantizar un orden parcial entre los objetos medidos. La segunda mejora consiste en asegurar un orden total mientras que, la medición proporcional se caracteriza por tener un punto de referencia absoluto o cero de la escala. Finalmente, la medición cardinal requiere una unidad de medida precisa, repetitiva y confiable.

Las mediciones realizadas en todo el trabajo experimental reportado en este artículo fueron hechas en alófonos por segundo. La unidad alófono no cumple con la propiedad requerida en las mediciones cardinales, pero se aproxima mucho más a ella que otras que las mediciones basadas en palabras por minuto. Las mediciones cualitativas no pueden despreciarse, especialmente en el área de velocidad del habla, ya que la extensión de su uso es muy grande pero es conveniente que las mismas se precisen con una medición cuantitativa comparativa.

En locuciones prolongadas un factor distorsivo es introducido por las pausas entre palabras, especialmente en las locuciones lentas, ya que en las locuciones normales y rápidas estos lapsos se contraen y desaparecen.

Se presenta entonces, en este artículo, un experimento de mediciones de la velocidad del habla basada en el conteo de la cantidad de alófonos pronunciados en un lapso medido. Para ello se tomó un muestreo de 30 personas que leyeron 120 textos a distintas velocidades. Para cuantificar el concepto intuitivo de habla rápida, normal o lenta se hizo escuchar estos textos a 31 personas que los rotularon con criterio propio en cinco clases distintas. A partir de éstos datos se establecieron los valores medios de velocidad normal en aproximadamente 13 alófonos por segundo lo que en valores conocidos corresponde a 150 ppm. Un resultado inédito es la ponderación de los conceptos de velocidad muy lenta, lenta, rápida y muy rápida del habla. Para contrastar las mediciones realizadas con datos de la bibliografía se utilizó la cantidad media de alófonos por palabra estimada en 5,1 alófonos por palabra.

Proceso Experimental. En el desarrollo de la experiencia se tomaron grabaciones de diferentes textos provenientes de diversas fuentes: lecturas de artículos, conversaciones, locuciones de televisión, transmisiones radiales de partidos de fútbol, noticieros, reportajes, monólogos. [10]

	<i>Alf pron. Durac. Vel</i>				<i>Alf pron. Durac. Vel</i>				<i>Alf pron. Durac. Vel</i>		
	[a]	[s]	[a/s]		[a]	[s]	[a/s]		[a]	[s]	[a/s]
T1	260	22.00	11.8	T14	413	28.00	14.8	T27	714	28.76	24.8
T2	298	30.00	9.9	T15	626	30.00	20.9	T28	445	31.36	14.2
T3	395	30.00	13.2	T16	903	31.70	28.5	T29	701	32.60	21.5
T4	570	30.00	19.00	T17	130	36.62	3.6	T30	488	33.00	14.8
T5	443	30.00	14.8	T18	427	29.70	14.4	T31	567	33.00	17.2
T6	397	30.00	13.2	T19	791	28.80	27.5	T32	356	31.53	11.3
T7	577	30.00	19.2	T20	61	31.00	2.0	T33	67	30.00	2.2
T8	335	20.00	16.7	T21	166	30.50	5.4	T34	543	40.00	13.6
T9	220	20.00	11.0	T22	364	30.00	12.1	T35	235	20.00	11.8
T10	218	21.00	10.4	T23	574	34.33	16.7	T36	310	25.00	12.4
T11	305	16.00	19.1	T24	826	35.34	23.4	T37	486	30.00	16.2
T12	95	28.00	3.4	T25	130	33.35	3.9	T38	216	30.00	7.2
T13	216	29.00	7.5	T26	381	31.08	12.3	T39	227	30.00	7.6
				T40	708	30.00	23.6				

Tabla 1. Características medidas en los diferentes textos.

De los 120 textos disponibles se seleccionaron 40 de manera de lograr una muestra lo más uniforme posible y con gran variedad de velocidades de habla. Se procesaron luego las grabaciones extrayendo la cantidad de alófonos con las que se calculó la velocidad de locución. En la Tabla 1 se muestran los datos principales de los textos utilizados.

Velocidad						
	[a/s]	Muy Lento	Lento	Normal	Rápido	Muy Rápido
Texto20	2.0	31				
Texto33	2.2	31				
Texto12	3.4	31				
Texto17	3.6	31				
Texto25	3.9	31				
Texto21	5.4	9	22			
Texto38	7.2	4	24	3		
Texto13	7.5	4	24	3		
Texto39	7.6		22	9		
Texto2	9.9		21	10		
Texto10	10.4		7	24		
Texto9	11.0		4	27		
Texto32	11.3		5	26		
Texto35	11.8		5	26		
Texto1	11.8		5	26		
Texto22	12.1		4	27		
Texto26	12.3		5	26		
Texto36	12.4			31		
Texto3	13.2			28	3	
Texto6	13.2			28	3	
Texto34	13.6			28	3	
Texto28	14.2			25	6	
Texto18	14.4			25	6	
Texto14	14.8			25	6	
Texto5	14.8			20	11	
Texto30	14.8			23	8	
Texto37	16.2			8	23	
Texto23	16.7			8	23	
Texto8	16.8			5	26	
Texto31	17.2			3	23	5
Texto4	19.0				26	5
Texto11	19.1			7	24	
Texto7	19.2				21	10
Texto15	20.9				21	10
Texto29	21.5				20	11
Texto24	23.4					31
Texto40	23.6					31
Texto27	24.8					31
Texto19	27.5					31
Texto16	28.5					31

Tabla 2. Comparación de la velocidad del habla determinada en forma cualitativa y cuantitativa
Los mismos 40 textos fueron medidos cualitativamente utilizando un grupo independiente de oyentes que, con criterio propio, les debían adjudicar una y sólo una de las siguientes medidas: muy lento,

lento, normal, rápido y muy rápido. Estos oyentes no tuvieron acceso en ningún momento a ningún dato de velocidad de los textos estudiados ni a las opiniones de los otros oyentes, y cada uno escuchó los 40 textos en un orden diferente al de los demás.

Resultados. Los valores obtenidos como resultado de la experiencia se presentan en la Tabla 2. En cada línea se muestra la identificación del texto, su velocidad medida en cantidad de alófonos por segundo (a/s) y en las siguientes cinco columnas la cantidad de oyentes que clasificaron la rapidez de locución con el mismo rótulo. Se ordenó la Tabla en forma creciente con las velocidades de habla medidas para resaltar el ajuste con la clasificación realizada por los oyentes. La Tabla 3 contiene los valores medios y las desviaciones standard para cada columna, expresadas en alófonos por segundo. Se establece entonces que al hablar a (3.3 ± 1.3) a/s, (8.5 ± 2.1) a/s (12.9 ± 2.0) a/s (17.7 ± 2.2) a/s (24.4 ± 3.0) a/s se habla: muy lentamente¹, lentamente, normalmente, rápidamente y muy rápidamente.

Clasificación	Valor medio [Alófonos / Segundo]	Desviación [Alófonos / segundo]
Hablar muy lentamente	3.3	1.3
Hablar lentamente	8.5	2.1
Hablar normalmente	12.9	2.0
Hablar rápidamente	17.7	2.2
Hablar muy rápidamente	24.4	3.0

Tabla 3. Valores medios y desviaciones de las velocidades de locución

Como se muestra en la Tabla 4, cada promedio con su desviación standard permiten definir un intervalo de confianza para los cuales existe una probabilidad del 68 % de que una locución arbitraria pertenezca o no a la categoría cualitativa correspondiente. Otra información a tener en cuenta es, cuál sería la velocidad límite tal que la probabilidad de pertenecer a una categoría es igual a la anterior. En otras palabras, para qué velocidad teórica del habla se tendría que el 50 % de los encuestados la calificarían como rápida, y el otro 50 % como normal. Estos valores servirían para establecer los límites entre las categorías y se visualizan en la Figura 1 como los puntos de intersección entre curvas contiguas. Estos límites están también incorporados en la Tabla 4.

	Muy Lento	Lento	Normal	Rápido	Muy Rápido
Promedio	3.3	8.5	12.9	17.7	24.4
Desviación	1.3	2.1	2.0	2.2	3.0
Límites	5.3		10.7	15.2	20.5

Tabla 4 Límites entre los diferentes grupos cualitativos.

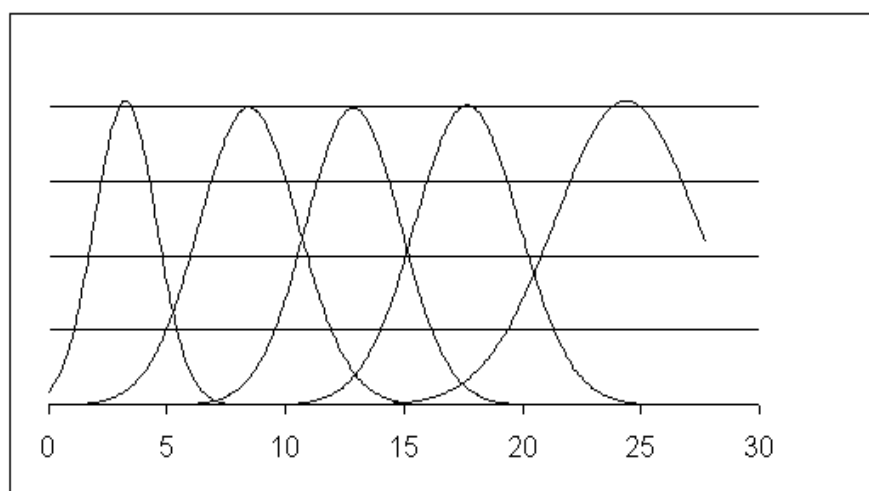


Figura 1 Distribuciones teóricas de probabilidades para cada grupo cualitativo

La Figura 1 muestra las curvas de densidad de probabilidad correspondiente a las cinco categorías presentadas en unidades arbitrarias para facilitar su interpretación.

¹Se entiende que para los extremos no tiene sentido una de las desviaciones ya que no se estableció una categoría contigua adicional.

A partir de la Tabla 4 se puede considerar que todo texto emitido a una velocidad inferior a 5,3 a/s entra en la categoría de muy lenta en tanto que si está entre 5,3 a/s y 10,7 a/s se la puede considerar como lenta; entre 10,7 a/s y 15,2 a/s se trata de una locución normal; será rápida si se emiten 15,2 a/s y 20,5 a/s y será muy rápida si supera los 20,5 a/s. Hablando en forma poco precisa los límites entre las categorías utilizadas son 5, 10, 15, y 20 a/s.

Comparación con información bibliográfica. Como ya se indicó más arriba, los datos disponibles están expresados en cantidad de palabras por minuto; de tal manera que fue necesario establecer un valor representativo de cantidad de alófonos por palabra. Para obtenerlo se recurrió a una serie de 30 textos elegidos al azar de los que se contó la cantidad de palabras presentes así como la cantidad de alófonos. Se llegó al valor de 5,1 alófonos por palabra como número promedio. La confiabilidad de este dato es excelente en relación con la precisión de los datos de referencia disponibles.

Traduciendo los valores de la Tabla 4 a palabras por minuto para el caso del habla normal se puede expresar que la misma varía de 128 a 180 palabras por minuto, valor que concuerda razonablemente con C. A. Loprete [1] quien calcula que una velocidad media de una conversación gira en torno a las 120–150 ppm. De manera análoga, en su curso de taquigrafía, M. Vasallo y C. Fusca de Elías proponen un régimen para alcanzar velocidades aceptables para tomar nota de conversaciones, comenzando por una velocidad de 20 ppm y aumentando hasta alcanzar las 102 ppm con lo que se cubriría todo el rango desde muy lento, lento y completando el curso cerca del umbral superior de una locución lenta. [6]

No se disponen de datos acerca de locuciones diferentes de la normal, pero la razonable equidistancia entre las medias de las cinco categorías utilizadas hace pensar que los resultados obtenidos son confiables también en estos casos.

Conclusiones. Los resultados obtenidos muestran que:

- Se puede medir con suficiente precisión y repetitividad la velocidad del habla utilizando como unidad la cantidad de alófonos emitidos por segundo (a/s)
- La práctica diaria es lo suficientemente estable para categorizar la rapidez del habla al menos en las cinco categorías mencionadas
- Se pueden establecer razonables relaciones entre la velocidad del habla y las categorías establecidas, las que se presentan en la Tabla 4 y la Figura 1.

Se desconoce la influencia cognitiva de la información contextual sobre la percepción intuitiva de la velocidad del habla y como afectó este fenómeno en el experimento realizado. Por ejemplo, es del conocimiento común que un locutor de fútbol relata hablando en forma muy rápida, quizás esta información haya afectado alguna opinión recogida en el experimento.

Bibliografía.

- [1] Loprete Carlos. "El lenguaje Oral, fundamentos, formas y técnicas " Ed Plus Ultra pp 37 (1984).
- [2] Casacuberta Francisco, Vidal Enrique. "Reconocimiento Automático del Habla". Marcombo pp 63–68 (1987).
- [3] Barbuzza Rosana, Doorn Jorge. "Uso de acentuación y relajación en Castellano Rioplatense para la Síntesis de Voz". Anales del Evento Científico Internacional COMPUMAT'97, Universidad de Cienfuegos, Cuba (1997).
- [4] Sproat Richard. "Multilingual Text-to-Speech Synthesis: The Bell Labs Approach". Lower Academic Publishers. pp 115–123. (1998).
- [5] Navarro Tomas. "Manual de Pronunciación Española". Madrid, Consejo Superior de Investigaciones Científicas, (1950).
- [6] Vasallo Miriam, Fusca de Elias Carmen. "Estenografía Vigente/2" Ed. Kapeluz pp 166–169 (1992).
- [7] Klaassen Klaas B. "Electronic Measurement and Instrumentation". Cambridge University Press. pp 1–15 (1996).
- [8] Briand Lionel, Morasca Sandro, Basili Victor. "Property-Based Software Engineering Measurement" IEEE Transactions on Software Engineering, 22,1, pp 68–85 (1996).
- [9] Zuse, H. "A Framework of Software Measurement. Berlin, Walter de Gruyter (1998).
- [10] Ghia Christian, Gianolli Claudio. "Influencia de la velocidad del habla en los alófonos y sus variaciones en el lenguaje rioplatense". Tesis de graduación en Ingeniería de Sistemas. UNCPBA, Argentina. (2000).