Acta Otorrinolaringol Esp. 2009;**60(1)**:43-8

ELSEVIER
DOYMA

Acta Otorrinolaringológica Española

www.elsevier.es/otorrino

ORIGINAL ARTICLE

Designing a new tool for hearing exploration

Antonio Rodríguez Valiente,[a,*] Carolina Pérez Sanz,[b] Carmen Górriz,[a] Adoración Juárez,[c] Marc Monfort,[d] José R. García Berrocal,[a] Juana Gil Fernández,[e] and Rafael Ramírez Camacho[a]

[a]*Servicio de Otorrinolaringología, Hospital Universitario Puerta de Hierro, Madrid, Spain*
[b]*Gabinete de Logopedia y Foniatría Frasis-Voz y Lenguaje, Madrid, Spain*
[c]*Colegio Tres Olivos, Madrid, Spain*
[d]*Centro Entender y Hablar, Madrid, Spain*
[e]*Laboratorio de Fonética, Consejo Superior de Investigaciones Científicas, Madrid, Spain*

Received March 26, 2008; accepted September 10, 2008

KEYWORDS
Speech audiometry;
Spectrogram;
Acoustic analysis

Abstract
Introduction: Many presbycusic patients have difficulty in understanding certain words. This could be justified because certain sounds in Spanish are more difficult to perceive, particularly the sounds with energy in the high frequencies. We propose to use a sentence as a tool to check this theory.
Materials and method: All the Spanish sounds were analyzed, measuring the degree of acoustic energy in all the frequencies. The conclusions drawn from the comparison of the results allowed the design of the tool that is proposed here.
Results: We established a gradient of perception difficulty, occlusive consonants being the least perceptible, followed by fricative, and finally all those segments with harmony and a clear formant structure. The Spanish sentence "Ana vio ese coche rojizo fino" is proposed as the tool for this study. This sentence comprises certain peculiarities that make it particularly useful for this purpose. It will allow us to check whether understanding deteriorates as we move from beginning to end, helping evaluate the importance of high frequencies for intelligibility.
Conclusions: A positive result could help in the design of amplification systems to improve speech intelligibility. In addition, the exploratory tool could allow neuro-acoustic exploration, useful in the central auditory pathology studies.

*Corresponding author.
E-mail address: arova777@hotmail.com (A. Rodríguez Valiente).

0001-6519/$ - see front matter

PALABRAS CLAVE
Audiometría verbal;
Espectrograma;
Análisis acústico

Diseño de una nueva herramienta para la exploración auditiva

Resumen
Introducción: Muchos pacientes presbiacúsicos presentan dificultad para entender ciertas palabras, hecho que podría estar justificado porque determinados sonidos en español presentan una dificultad mayor de perceptibilidad, en concreto los de energía en altas frecuencias. Se propone utilizar una oración como herramienta para comprobar la hipótesis de partida.
Material y método: Se analizaron todos los sonidos del español, midiendo el grado de energía acústica que presenta cada una de las frecuencias. Las conclusiones extraídas de la comparación de sus resultados permitieron el diseño de la herramienta que aquí se propone.
Resultados: Se establece una gradación de dificultad perceptiva, por lo que se puede decir que las oclusivas aparecen como los sonidos menos perceptibles, seguido de las fricativas y, finalmente, todos los segmentos que presentan armonicidad y una estructura formántica definida. Como herramienta para la práctica clínica, se propone la frase española "Ana vio ese coche rojizo fino", frase que cumple una serie de particularidades que la hacen especialmente útil para tal fin. Esta oración permitirá comprobar si la comprensión se va deteriorando a medida que se avanza desde la primera a la última de sus secciones, y así poder suponer que la capacidad para detectar auditivamente la presencia de energía reforzada en las altas frecuencias es indispensable para la inteligibilidad.
Conclusiones: Un resultado positivo podría tener como consecuencia el diseño de sistemas de amplificación que mejoren la inteligibilidad de la palabra. Además, el hecho de tener una herramienta exploratoria podría permitir la exploración neuroacústica, de utilidad en el estudio de la enfermedad auditiva.

Introduction

Most presbycusic patients have difficulty in perceiving and understanding certain words consisting of specific sounds. This finding makes it advisable to design an applied research tool valid to verify the assumptions made, namely that not all the sounds in the Spanish language have a similar degree of intelligibility, and that this difference in perceptibility between them is based on the location of their components in the available acoustic space. More specifically, it is desired to prove that the existence of very high frequency energy in certain phonatory elements, with the hearing range constraints entailed by this, may be a decisive factor in their level of audibility.

These observations in daily clinical practice have given rise to this research proposing the use of a short sentence, supported by its segmental and suprasegmental composition, in order to meet the demand of resources for the suggested research.

History of the problem

Intelligibility decreases progressively throughout life, to the extent that the presbycusis shows a decrease in differentiation, which increases over time. Various amplifying solutions have been designed to solve this problem. However, a certain number of patients still complain of poor intelligibility in adverse situations (noisy environments, several people talking at the same time, etc). Since most prostheses cover a frequency range which does not exceed 4000-6000 Hz,[1] the frequencies most affected by presbyacusis would not be covered.

Proposed tool

The Spanish language, in particular its Castilian variant, is a language that favours the intelligibility of statements since its phonological system comprises a small number of phonemes presenting relatively few allophones and its vocalic subsystem is also limited.[2] Furthermore, the proportion of monosyllabic words is relatively small. All this favours the understanding of messages by the hearer, as evidenced by the better performance of Spanish-speaking adults and children after cochlear implant, compared with the performance of speakers of other languages. Nevertheless, the intelligibility of emissions depends not only on the language in which they are delivered, but also on a number of factors, which include the quality of diction of the transmitter, the characteristics of the voice, the functional capacity of the receptor's hearing, and the environment in which the act of communication takes place. These variables may positively or negatively affect the transmission of information.

In order to attempt to verify the hypothesis, which deals in particular with the influence of the acoustic constitution of sounds on their audibility, it was decided to design a sentence in which each of its parts concentrated one specific type of sound, with regard to frequency distribution.

As is well known, tests have already been created to distinguish disorders affecting the fluent perception of oral sentences.[3-5] Thus, lists of disyllabic and trisyllabic Spanish words are used in verbal audiometry to assess the various grades and types of hearing loss. Their advantage is that the subject has to identify the components of the words emitted and repeat them in an environment in which the relationships between each word and the next are lost, so it is impossible to deduce them by their meaning or context. In this sense, it is worth noting that the decision to work with a statement larger than single words is due to the desire to create a stimulus as close as possible to real communication, in other words, more complex.

Material and method

Firstly, we proceeded to record all the sounds of the Spanish language pronounced by a male and a female speaker, both consonants and vowels, isolated in the case of vowels or in a phonetically normal position in the case of consonants, that is, in the *a_a* context, in order to consider its inherent characteristics without the influence of co-articulation with other elements. It was also necessary to do it this way so that the sounds had approximately the same duration, tone and stress, thus making their intensity or acoustic energy comparable. The recordings were made at the Phonetics Laboratory of the Superior Science Research Council using a Kay Elemetrics Co. CSL 4500 device with a sampling frequency of 44 100 Hz using an AKG C444L headworn microphone 10 cm from the speaker's labial commissure and with a lateral displacement of 45°, as stipulated in the procedural guidelines for acoustic analysis.[6-8]

Next, the recordings were edited, and the registered sounds delimited so that none of them, particularly in the unvoiced occlusives and fricatives, contained fragments corresponding to the transitions with adjacent vowel sounds. Once this task was completed, the edited sounds were analyzed using the freely-available PRAAT software, the Spectrum/Spectrum Tier analysis tool of which shows the degree of acoustic energy, measured in dB, presented by each of the frequency components in each segment. Table 1, by way of example, reproduces the data obtained in connection with the unvoiced fricatives.

Finally, the comparison of these data, that is, the comparison of the frequencies for each element and its corresponding intensity, or sone, both in the male and the female voices, allowed us to draw certain conclusions that were helpful in carrying out the design of the tool proposed in this paper.

Tabla 1 Values in dB for each of the 10 most enhanced frequencies in the fricative consonants [tʃ], [s], [x], [θ], and [f]. The most enhanced frequencies in [tʃ] and [s] are around 6500-7000 Hz, whereas in [x], [θ], and [f] they are in the lowest part of the spectrum. The fundamental difference between [θ] and [f] originates from the last 3 frequencies shown here: whereas in [θ], the seventh frequency showing enhancement is 3006 Hz, in [f] it is 11 534 Hz

tʃ	Freq, Hz	SD, dB/Hz
	7034.7184	***49.5939867***
	6721.24918	***47.4102637***
	6121.66661	***46.7913527***
	6014.04025	46.4455567
	6793.98022	46.1998035
	7129.76212	46.1967457
	7100.41158	46.1375997
	6591.87264	44.7069017
	7846.0364	44.4583153
s	Freq, Hz	SD, dB/Hz
	6704.75938	***55.5915904***
	6623.17329	***52.7830834***
	7005.89096	***52.4822078***
	7134.17278	52.2950957
	6547.17732	51.0174019
	7022.50095	50.574668
	6990.30341	50.4960023
	7100.02369	50.4922775
	6783.13793	50.2037553
x	Freq, Hz	SD, dB/Hz
	70.8179343	***66.5173011***
	137.942909	***55.5470978***
	113.134542	***54.9647791***
	1755.31159	51.2579729
	1627.88981	50.6611579
	211.476575	49.3627643
	1655.72318	49.1738561
	33.0424905	48.390016
	1690.64422	48.1926085
θ	Freq, Hz	SD, dB/Hz
	53.3065217	***57.9471076***
	40.2614916	***54.1348469***
	4.66598407	***51.5578407***
	14.2150504	50.9677095
	86.8198371	49.4484912
	113.831054	48.4791122
	3006.74241	48.2413807
	2952.19691	47.5413852
	5116.96267	46.6211333
f	Freq, Hz	SD, dB/Hz
	2***1.8909035***	***59.9376243***
	12.0531868	***54.5582861***
	53.4217708	***51.2787228***
	68.7385728	50.6013289
	39.3248854	50.2418703
	82.3634332	47.8169694
	11 534.6935	47.6140699
	14 392.2909	47.1712648
	11 621.3386	46.79903

Freq indicates frequency; SD, sound density.

Results

With the values obtained in the analysis, it is inferred that:

a) Among the fricative consonants, those known as *sibilants* (articulated with a horizontal opening narrower than the vertical opening, sometimes called "bulging"), which in Spanish are [s] and the second phase of [tʃ] (the "ch" in *ha*cha), present a considerable level of energy (46-55 dB) in the frequency range around 6500-7000 Hz. In contrast, low frequencies are not boosted.

b) Non-sibilant fricatives, [f] [x] (the "j" in "*baja*") and [θ] (the "z" in "*baza*" and the "c" in "*cena*"), do show boosting at low frequencies, in particular the [θ] (eg, 54 dB at 40 Hz) and [f] (eg, 51 dB at 53 Hz). In the latter case, the reinforcement of the sound energy is barely perceptible between 80 Hz and the range between

Tabla 2 Values in dB for each of the 10 most enhanced frequencies in the vowels [a], [e], [i], [o], [u]

a	Freq, Hz	SD, dB/Hz
	937.656286	***64.2417931***
	1125.55491	***62.4495366***
	749.586463	***62.0523815***
	186.967847	61.7017315
	1313.6144	60.6849474
	927.142618	60.4748342
	1114.85166	58.9390213
	739.977936	57.765972
	373.933471	57.6877286
e	Freq, Hz	SD, dB/Hz
	196.956074	***66.3890339***
	396.146016	***66.3802251***
	595.21857	***65.946384***
	188.012273	63.7817405
	383.928014	62.3257725
	581.736411	62.2029421
	573.176941	61.6391833
	566.180368	60.5731901
	376.776807	59.4434761
i	Freq, Hz	SD, dB/Hz
	205.609909	***74.7025744***
	196.712741	***69.0871327***
	191.456775	***66.0266579***
	411.995353	65.9668429
	186.325534	62.574912
	402.200341	62.078027
	395.208526	58.9933482
	180.797274	58.6331557
	387.992402	55.9892634
o	Freq, Hz	SD, dB/Hz
	573.248305	***66.1074428***
	381.501714	***64.7810525***
	188.791285	***63.9183019***
	561.237597	61.9475573
	552.720448	60.2597225
	371.109125	60.0962881
	546.849893	58.0828707
	957.093812	57.9745194
	765.287739	57.7010426
u	Freq, Hz	SD, dB/Hz
	200.14247	***71.7034853***
	401.777924	***70.7596058***
	190.751281	***66.752235***
	390.709272	66.6172194
	383.056985	63.8576324
	376.370384	63.3542002
	371.610492	62.6704557
	366.557911	61.4758289
	361.614772	59.260932

Freq indicates frequency; SD, sound density.

11 500 and 14 800, whereas in [θ] the energy appears more evenly distributed from 2800 Hz up to the highest frequency area: 41 dB at 11 300 Hz.

Therefore, as far as the fricatives are concerned, and starting from the initial assumption that higher frequency sounds require a greater hearing range, or if preferred, entail greater difficulty for patients with presbyacusis, a scale could be established from highest to lowest degree of perceptibility, as follows:

[x] [θ] [tʃ] [s] [f]
+ ________________ –

c) With regard to the vowels, in the first place, it should be noted that it is their first 2 formants (F1 and F2) which determine their timbre, ie, their identity as a sound: an [a] as a different sound from an [e], from an [i], etc, while the remaining higher formants (F3, F4, and F5) reflect the individual characteristics of each voice. The data obtained support the idea that all of them are easily perceptible sounds, since all their formants are between 190 and 6000 Hz, with the most critical zone for their identification located around the values listed in Table 2.

d) The analysis of the occlusives ([p], [t], [k]) shows, first of all, that they do not present high frequencies carrying a significant noise density, and, secondly, that, as has been pointed out repeatedly in the literature, they are inherently less perceptible sounds, since they have less sone in general: they are produced by a closure or occlusion in an area of the vocal tract, which is broken by a characteristic noise known as the explosion, which is not always present. Moreover, the maximum level of intensity that they reach will never be as marked as that of other sounds. From all this, it follows that their difficult perception by listeners, even those without a disease, is not given by the existence of high frequencies that may be outside the hearing range, but by their very own nature.

It is interesting that, in the tradition of phonetic studies of linguistics, scales have been proposed time and again for many years for "loudness" or "sone"[9,10] to rank the different types of sounds (Figure 1).

As can be seen, these scales coincide roughly with the gradient of difficulty of perceptibility inferable from what has been presented here so far, in other words the occlusives appear as the least perceptible sounds, as explained previously in point *d)* above; next, the fricatives are the least "sounded" (understanding "sounded" not as produced by the vibration of the vocal cords, but as having "sone") and, finally, all the segments presenting a certain harmonicity and defined formant structure, such as the nasals, laterals, vibrants, and vowels, are the most perceptible.

In fact, it has repeatedly been pointed out that sounds with a prominent formant structure and low attenuation are the most "resounding." Resonance in speech is always underscored by sonority, so that the resonant cavities making up the vocal tract respond in a more efficient manner when there is vibration in the vocal cords, in which case the concentration of acoustic energy in specific frequency bands is higher and more evident. In this sense, it could be argued that the level of perceptibility or sone does not depend solely on the intensity of the sound, ie, the amount of energy it presents, but how this energy is distributed over its spectrum.[11]

Proposal of the model sentence and its rationale

Based on the findings reached following the analyses described above, it was considered that a sentence which is valid as a tool for clinical practice is the Spanish phrase "Ana vio ese coche rojizo fino" ("Ana saw that fine reddish car"), the spectrogram of which is shown in Figure 2. The explanation of why this conclusion has been reached is the following:

a) The sentence proposed is a simple affirmative phrase, consisting of subject, verb, and complement, which from a suprasegmental point of view (ie, with respect to melody, accentuation, and rhythm with which it is emitted) does not present any special features that can help the listener in the process of perceiving sounds. For example, it is known that an interrogative sentence systematically favours the perception of high frequencies, which are reinforced in some positions; therefore a sentence was chosen with an assertive or neutral intonation curve.

b) The phrase is short and forms a single phonic group, ie, a statement fragment that can be emitted between pauses, for example breathing pauses, without internal interruptions. The number of syllables corresponds to the average for the Spanish language.

c) The meaning of the words in the sequence does not help to understand their meaning. The adjective "reddish" is not immediately associated with the noun "car," while it would be more predictable with, for example, "hair."

d) In the sentence constructed it is possible to distinguish 3 distinct parts in terms of the type of sounds associated with each. "Ana vio" ("Ana saw") is the first fragment, and in it all the consecutive elements (including the voiced labial fricative "v" [β]) are sounded, have a clear formant structure or, in other words, are "sounded" in the sense explained above; they contain energy in various parts of their spectrum and are therefore highly perceptible. The second segment is "ese coche" ("that car") and the degree of difficulty for audition is increased because although there are middle vowels ([e] and [o]), which in principle are quite perceptible, there are 2 sibilants, the [s] of "ese" ("that") and the second phase of the [tʃ], "*ch*" in "coche" ("car"), which are, as mentioned above, fricatives with a large amount of energy in high frequency areas. Finally, "rojizo y fino" (the adjectival phrase "reddish and fine") is the most complicated from a hearing perspective as it presents various acute vowels [i], and the 2 sounds which are most difficult to perceive, as shown above, the [θ], "z" and the [f] as well as the [x], "j".

If the clinical experimentation shows that the hearing and subsequent understanding of this sentence gradually deteriorates from the first to the last of its sections, then there will be sufficient grounds to assume that the ability to detect aurally the presence of enhanced energy in the high or very high frequencies is essential for the intelligibility

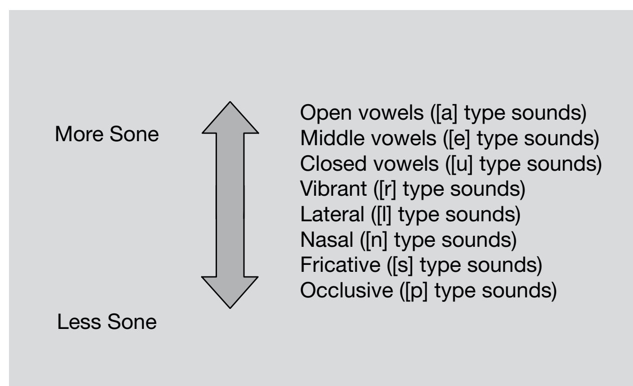

Figure 1 Scale of sonority ordering the different types of sounds.

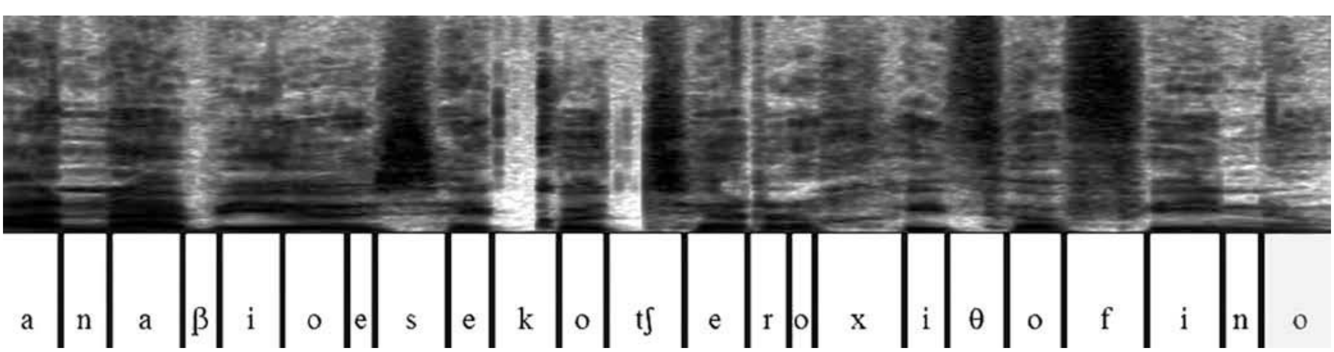

Figure 2 Narrow band spectrogram of the proposed sentence "Ana vio ese coche rojizo fino" ("Ana saw that fine reddish car") recorded with a sampling frequency of 44 100 Hz, obtained with Praat at a maximum frequency of 22 050 Hz, drawn in a 2 ms Hanning window. Note how in the sound [θ] the distribution of sound energy is more homogeneous across all frequencies, while in [f] the zone of greatest noise is concentrated in the highest zone, and in [s] and in the fricative phase of [tʃ] it is located in the middle of the spectrum. Harmonics are distinguished in all vowels up to frequencies around 12 kHz, although [i] also presents higher frequencies; nevertheless, the highest noise load in all of them is in the first formants.

of certain messages, and it will become more evident why some patients who no longer have this ability display special difficulties to achieve understanding.

Conclusions

The possibility of associating the ability of an ear affected by sensorineural hearing loss to perceive a sentence designed with phonetic criteria ("Ana vio ese coche rojizo fino") represents the development of a specific tool that may result in the design of amplification systems to improve speech intelligibility. The proposed sentence can be administered by the individual to the ear under examination through electronic devices contemplating gender, age, ambient noise, coincidence with other voices, so as to allow the evaluation of auditory perception in conditions close to real life. Additionally, its association with techniques already in place, such as logoaudiometry, could enable the neuroacoustic examination of associative areas versus primary auditory areas, which would be potentially useful in studying central auditory disease.

References

1. Moore BC, Stone MA, Alcántara JI. Comparison of the electroacoustic characteristics of five hearing aids. Br J Audiol. 2001;35:307-25.
2. Maddieson I, Precoda K. UCLA Phonological Segment Inventory Database. (Database with details of 451 languages available for download.) Available from: www.linguistics.ucla.edu/faciliti/sales/software.htm
3. Tato JM, Alfaro A. Audiometría del lenguaje. Discriminación por vía ósea: umbral de captación. Índice de utilidad social auditivo. Otorrinolaringológica. 1949;1:3.
4. Poch Viñals R. La exploración funcional auditiva. Madrid: Ed. Paz Montalvo; 1958.
5. Cárdenas MR, Marrero V. Cuadernos de Logoaudiometría. Cuadernos de la Universidad Nacional de Educación a Distancia. Valladolid: Simancas Ediciones; 1994.
6. Deliyski DD, Evans MK, Shaw HS. Influence of Data Acquisition environment on accuracy of acoustic voice quality Measurements. J Voice. 2005;19:176-86.
7. Titze IR. Toward standards in acoustic analysis of voice. J Voice. 1994;8:1-7.
8. Titze IR. Workshop on Acoustic Voice Analysis: Summary Statement. National Center for Voice and Speech; Denver, Colorado: 1994.
9. Clements GN. The role of sonority in core syllabification. In: Kingston J, Beckman ME, editors. Papers in laboratory phonology I. Between the grammar and physics of speech. Cambridge: CUP; 1990. p. 283-333.
10. Butt M. Sonority and the explanation of syllable structure. Linguistische Berichte. 1992;137:45-67.
11. Clements GN. Does sonority have a phonetic basis? Laboratoire de Phonétique et Phonologie. CNRS / Sorbonne-Nouvelle, Paris, France 2006. Available from: http://ed268.univ-paris3.fr/lpp/publications/2006_Clements_Does_Sonority.pdf